



**Forschungsdatenmanagement
am LRZ im Kontext von GeRDI**

Dieter Kranzlmüller, Tobias Weber, Hai Nguyen
13.03.2018 | ZKI Frühjahrstagung 2018, Konstanz

ZKI-Frühjahrstagung 2018 13.03.2018 | Dieter Kranzlmüller (LRZ) 2

Reproduzierbarkeit und Nachnutzung



Problem:
Reproduzierbarkeit
und Nachnutzung
von Forschungs-
ergebnissen



- Fehlende Dokumentation und Beschreibung (Metadaten)
- Nachprüfbarkeit wissenschaftlicher Ergebnisse
- Heterogenität und fehlende Standards
- Repetition von Forschung
- Kosten für Zeit und Infrastruktur

Herausforderungen



- Technische Herausforderungen
 - Schnittstellen (Interoperabilität)
 - Formate und Metadatenstandards
- Aufwand und Kosten
- Unklarheit bei Verantwortlichkeiten und Strukturen
- Speicherung von sensiblen Daten
- Umgang mit geistigen Eigentum

Forschungsdatenmanagement (FDM) und Infrastrukturen



Forschungsdatenmanagement

„Alle Aktivitäten, die mit der Aufbereitung, Speicherung, Archivierung und Veröffentlichung von Forschungsdaten verbunden sind“¹



Aufbau und Vernetzung von Forschungsdateninfrastrukturen

Personal, Dienste und Werkzeuge zur Unterstützung der FDM-Aktivitäten



¹ Simukovic, E., Kindling, M., & Schirmbacher, P. (2013); Umfrage zum Umgang mit digitalen Forschungsdaten an der Humboldt-Universität zu Berlin
ZKI-Frühjahrstagung 2018 13.03.2018 | Dieter Kranzlmüller (LRZ)

Forschungsdaten-Lebenszyklus



Creating Data

- Design Research
- Plan Data Management (formats, storage, etc.)
- Plan consent for sharing
- Locate existing data
- Collect data (experiment, observe, measure, simulate)
- Capture and create metadata



UK Data archive
ZKI-Frühjahrstagung 2018

Forschungsdaten-Lebenszyklus



Processing Data

- Enter data, digitise, transcribe, translate
- Check, validate, clean data
- Anonymise data where necessary
- Describe data
- Manage and store data



Forschungsdaten-Lebenszyklus



Analysing Data

- Interpret data
- Derive data
- Produce research outputs
- Author publications
- Prepare data for preservation



Forschungsdaten-Lebenszyklus



Preserving Data

- Migrate data to best format
- Migrate data to suitable medium
- Backup and store data
- Create metadata and documentation
- Archive data



Forschungsdaten-Lebenszyklus



Giving access to data

- Distribute data
- Share data
- Control access
- Establish copyright
- Promote data



Forschungsdaten-Lebenszyklus



Re-using data

- Follow-up research
- New research
- Undertake research reviews
- Scrutinise findings
- Teach and learn

```

    graph TD
      Re-using[Re-using Data] --> Creating[Creating Data]
      Creating --> Processing[Processing Data]
      Processing --> Analysing[Analysing Data]
      Analysing --> Preserving[Preserving Data]
      Preserving --> Access[Giving Access to Data]
      Access --> Re-using
      style Re-using fill:#f96
    
```

UK Data archive
ZKI-Frühjahrstagung 2018
13.03.2018 | Dieter Kranzlmüller (LRZ)
11

Das Leibniz-Rechenzentrum



Optimale Unterstützung der Wissenschaftler und Studenten
 mit einer hochmodernen IT-Infrastruktur
 bei ihrer täglichen Arbeit mit Forschungsdaten

Kompetenter Partner
für zuverlässige und qualitativ
hochwertige Dienste



Forschungsbereiche

- Umweltwissenschaften
- Angewandte Wissenschaften
- Digitale Geisteswissenschaften
- ...



ZKI-Frühjahrstagung 2018
13.03.2018 | Dieter Kranzlmüller (LRZ)
12

LRZ Dienste





HPC



Netz



Speicherlösungen



Virtuelle Realität und Visualisierung



Training



Know-How



Beratung



Support

ZKI-Frühjahrstagung 2018 13.03.2018 | Dieter Kranzlmüller (LRZ) 13

Höchstleistungsrechenzentrum und Big Data Kompetenzzentrum



SuperMUC Phase 1

- 3.2 Pflops Spitzenleistung
- 147.456 Sandy Bridge CPU Kerne

SuperMUC Phase 2

- 3.6 Pflops Spitzenleistung
- 86.016 Haswell CPU Kerne

Bayerisches Big Data Kompetenzzentrum

- Compute Cloud
- DGX-1
- GPU-Cloud



ZKI-Frühjahrstagung 2018 13.03.2018 | Dieter Kranzlmüller (LRZ) 14

Speicher und Archiv System



Insgesamt 200 Petabyte Speicher

- HPC Cluster Dateisysteme
- Data Science Storage
- Archiv- und Backupsystem
- Bayern Cloud



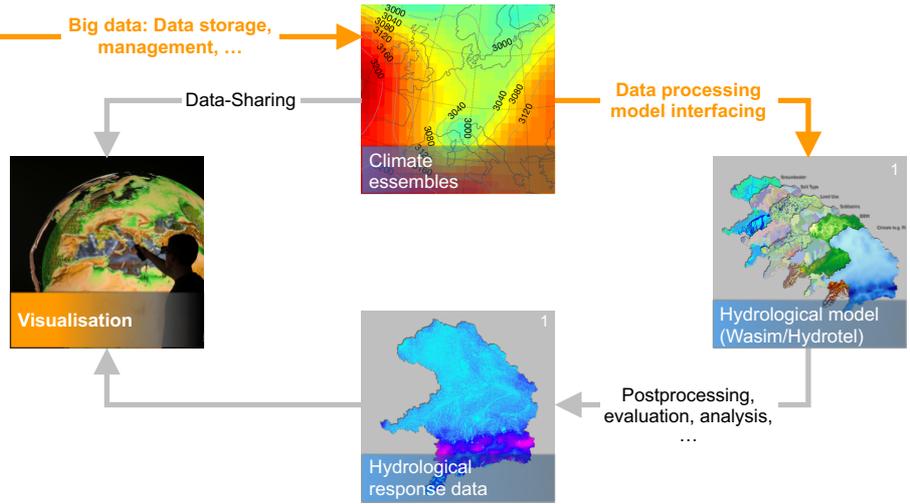
ZKI-Frühjahrstagung 2018 13.03.2018 | Dieter Kranzlmüller (LRZ) 15

Beispiel: Projekt ClimEx



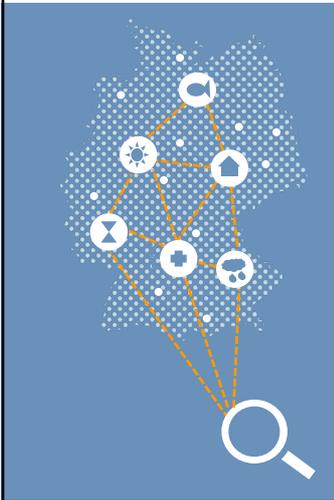
High Performance Computing:
Regional climate model (CRCM 5)

- Simulation auf Super-MUC mit 80 Millionen CPU Stunden
- 15.000 Jahre Klimaensembles
- Datenauswertung mit Machine Learning
- >500 TB Daten



1 Copyright LMU/Geo
ZKI-Frühjahrstagung 2018 13.03.2018 | Dieter Kranzlmüller (LRZ) 16

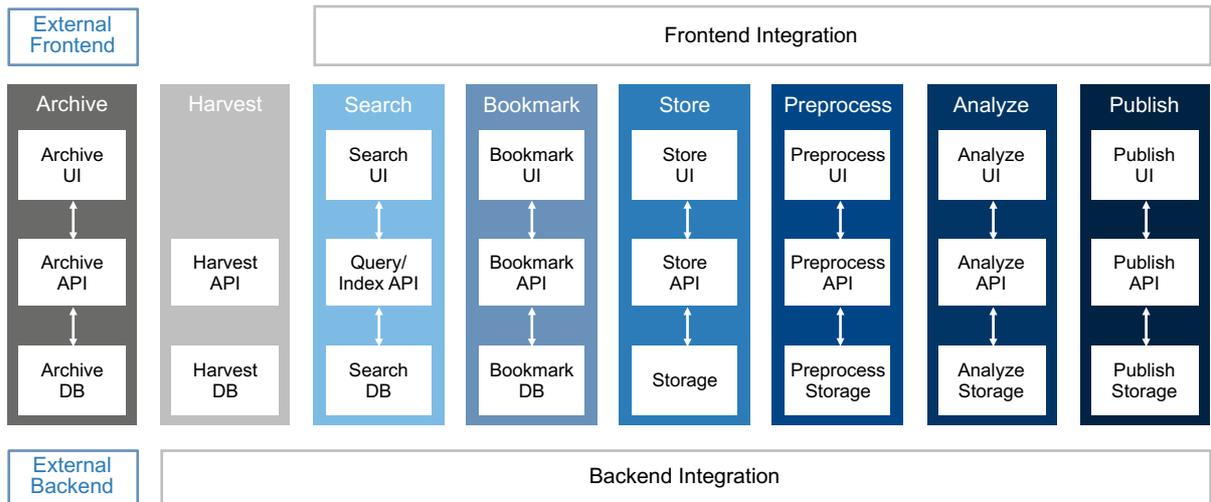
Generic Research Data Infrastructure

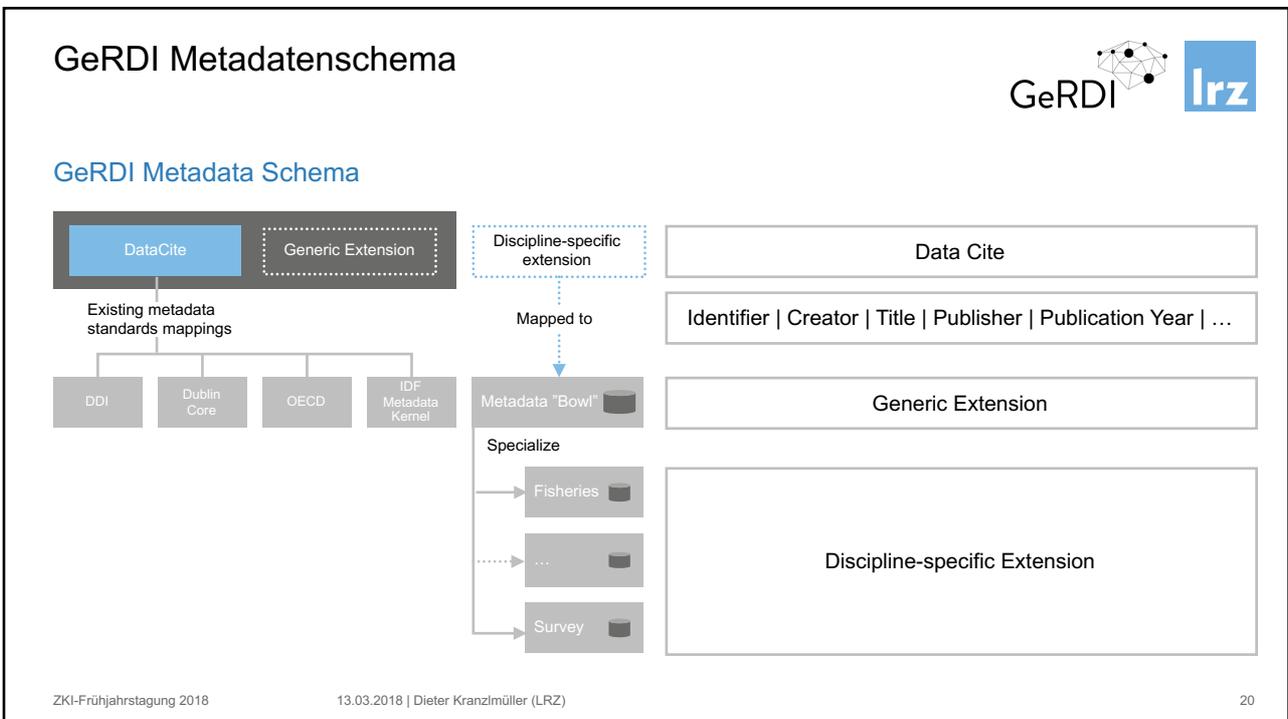
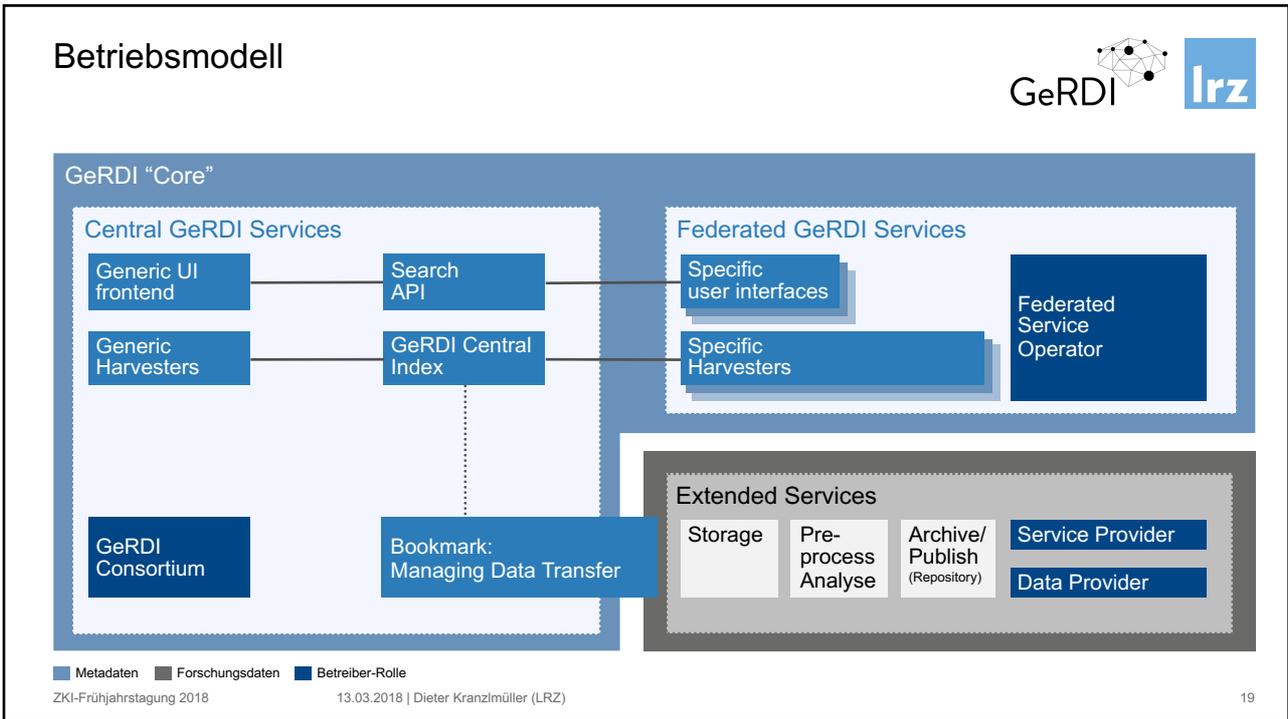


- Ziel: Aufbau einer Forschungsdateninfrastruktur in Deutschland
- Erste Phase: 2016 - 2019
- Grundlage: FAIR Prinzipien
Findable | **A**ccessible | **I**nteroperable | **R**eusable
- Einbindung neun Communities in die Entwicklung
- Modellhafte Struktur und Anschlussfähigkeit an nationale und europäische Forschungsdateninfrastrukturen
- Projektleitung: Klaus Tochtermann (ZBW)



Architektur

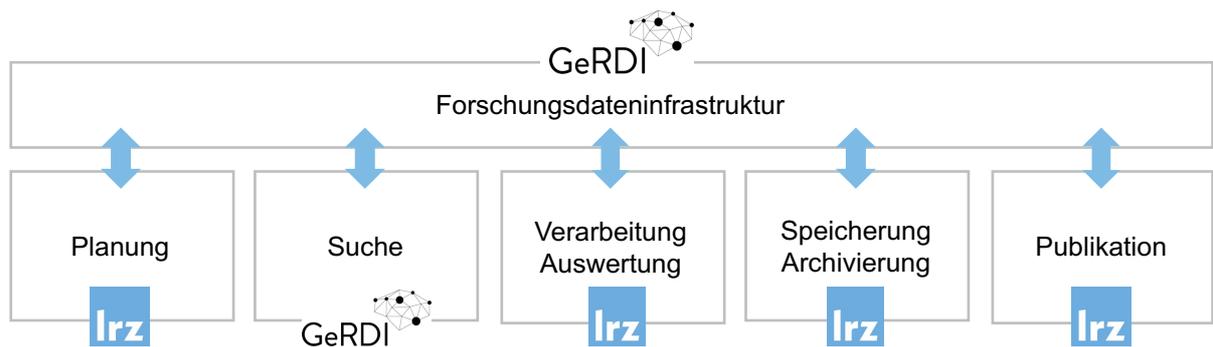




FDM mit LRZ und GeRDI



- Integration von Forschungsdatenrepositorien mit LRZ Diensten
- Verknüpfung bestehender LRZ Dienste
- Unterstützen von FDM Workflows



ZKI-Frühjahrstagung 2018

13.03.2018 | Dieter Kranzlmüller (LRZ)

21

Fazit



- FDM ist ein wesentlicher Aspekt für nachhaltige Forschung
- Daten- und Rechenzentren als wichtiger Partner bei FDM
- FDM am LRZ als Kombination aus LRZ und GeRDI Diensten
- Schnittstellen ermöglichen förderiertes FDM



ZKI-Frühjahrstagung 2018

13.03.2018 | Dieter Kranzlmüller (LRZ)

22

Ausblick

- GeRDI Phase 2 und Ausweiten auf ganz Deutschland
- Beratung und Schulungen im Umgang mit Forschungsdaten
- Forschungsdatenrepositorium für eine FAIRe Datenspeicherung



Forschungsdatenmanagement am LRZ im Kontext von GeRDI

Prof. Dr. Dieter Kranzlmüller

**Leibniz-Rechenzentrum
der Bayerischen Akademie
der Wissenschaften**

Boltzmannstr. 1, 85748 Garching

Email: Dieter.KranzlmueLLer@lrz.de

