

LRZ Research Data Management & NFDI Neuroscience

Leibniz Supercomputing Centre

07/2021 | Online | S. Hachinger (LRZ RDM team)

Leibniz Supercomputing Centre

of the Bavarian Academy of Sciences and Humanities (BAdW)



- ✓ IT Service Provider (“Rechenzentrum”) for the **Munich** Universities
- ✓ Regional Computing Centre for Research Institutions in **Bavaria**
- ✓ **German** National Supercomputing Centre (in GCS, one of the three strongest machines in D); **serves EU users** (PRACE)

*all services, easy access,
local customer base*

|
*„premium“ services, access by
proposal, EU customer base*

- ✓ Classical **High-Performance-Computing** clusters
- ✓ **Cloud-Computing Resources** (OpenStack, VMWare) - IaaS Virtual Machines
- ✓ **Data-analysis platform service** solutions in making (e.g. RStudio server, Jupyter)

RDM support at LRZ: metadata & „FAIR“



Setting up a simple metadata store for existing data stores

- LRZ Research Data Management (“Let The Data Sing”) for supercomputing

OAI-PMH server for metadata catalogue

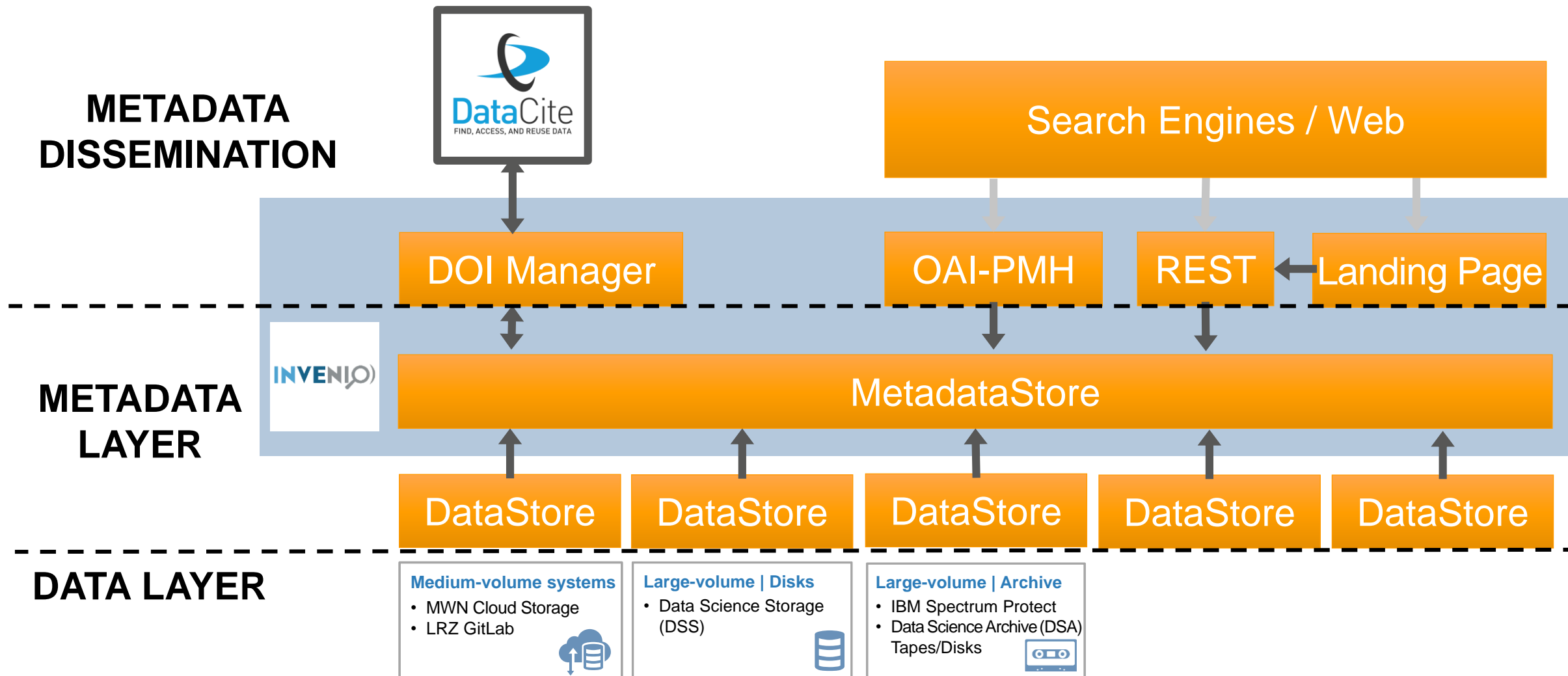
- Help with server set-up, automatization

Having catalogues integrated in

- EUDAT
- BASE
- ... (other search engines)

data store(s), community repositories

“FAIR Data for High Performance Computing at LRZ” Let the Data Sing: A lightweight RDM layer



Where are we going with this system? Metadata standard & „crawling“

- DataCite metadata in .metadata.yaml
 - discipline-agnostic, minimal (for DOI)
 - .yaml file creation via web interface
 - links to domain-specific metadata possible
- Metadata „defaults“ or instructions to system to get it:
 - from files with metadata (e.g. HDF5, netcdf)
 - from directory names
- Harmonisation of methods at different computing centres (in InHPC-DE and other projects, from 2022 on) as aim

```
/dss/dssf~fig.yaml [-M--] C: 1 L:[ 49/327 49+ 0] *(1531/11681b)= 32 0x20
- creatorName:
  default: "Stephan Hachinger"
nameType:
  default: "Personal"
givenName:
  default: "Stephan"
familyName:
  default: "Hachinger"
nameIdentifiers:
  - nameIdentifier:
    default: "0000-0001-8341-1478"
    nameIdentifierScheme:
      default: "ORCID"
    schemeURI:
      default: "http://orcid.org/"
affiliations:
  - affiliation:
    default: "Bavarian Academy of Sciences and Humanities"
    affiliationIdentifier:
      default: "https://ror.org/001rdaz60"
    affiliationIdentifierScheme:
      default: "ROR"
    schemeURI:
      default: "https://ror.org/"
titles:
  title:
    default: "ViWA Global Simulation Data"
publisher:
  default: "Leibniz Supercomputing Centre"
PublicationYear:
  default: "2020"
subjects:
  - subject:
    default: "Hydrology"
    subjectScheme:
      default: "Library of Congress"
    schemeURI:
      default: "https://id.loc.gov/authorities/subjects.html"
    valueURI:
      default: "https://id.loc.gov/authorities/subjects/sh85063458.html"
  - subject:
    default: "Water efficiency"
    subjectScheme:
      default: "Library of Congress"
```


Possible tasks for LRZ in NFDI Neuroscience



NFDI Neuroscience – LRZ concept

- *Data Store Interoperability LTDS – NFDI Neuroscience*
 - Converters e.g. DataLad ↔ LTDS metadata
 - integration of HPC data at LRZ with NFDI Neuroscience
 - Given the standardisation efforts among supercomputing centres, this should evolve to a general „HPC data“ approach for NFDI Neuroscience
- *NFDI-Neuroscience tools at LRZ*
 - NFDI Neuroscience analysis tools on LRZ Systems
 - Access to data at LRZ via APIs (give PID – get data)
 - NFDI Neuroscience data management tools on LRZ systems
- *Access control*
 - Restricted access to data via APIs (giving, e.g. a PID)
- *Consulting, co-development*

LRZ RDM Team:
rdm@lists.lrz.de

Stephan Hachinger:
hachinger@lrz.de
+49-89-35831-7830

- LRZ NFDI Support, Lead RDM Team
Stephan Hachinger
- LRZ BioMedICT Task Force
Megi Sharikadze
- LRZ RDM Team
Rubén García-Hernandéz, Stephan Hachinger,
Mohamad Hayek, Johannes Munke, and associated colleagues

Mail address:
Leibniz-Rechenzentrum
Boltzmannstr. 1
85748 Garching bei München