

University of Stuttgart
Universitätsbibliothek

DaRUS

Dataverse at the University of
Stuttgart

Dataverse at the University of Stuttgart

<https://darus.uni-stuttgart.de>

DaRUS: DAta Repository of the University of Stuttgart

- Currently in Test Operation (since Oct 2018)
- 75 (test) users, 88 Dataverses

The screenshot shows the DaRUS website interface. At the top, there is a blue header with the University of Stuttgart logo and name. Below the header, the DaRUS logo is visible on the left, and navigation links for Search, User Guide, Support, and a user profile (Dorothea Iglezakis) are on the right. The main content area features a 'Metrics' section showing 00 Downloads, and buttons for Contact, Share, and Edit. A descriptive sentence reads: 'This is the data Repository of the university of Stuttgart.' Below this, four institutional logos are displayed: SimTech (Stuttgart Center for Simulation Science), SFB 1313, itt (Institute of Thermodynamics and Thermal Process Engineering), and IAG (Institute of Aerodynamics and Gas Dynamics). A search bar is present with a 'Find' button and a link to 'Advanced Search'. On the right of the search bar is a '+ Add Data' button. Below the search bar, a filter sidebar on the left shows 'Dataverses (88)', 'Datasets (96)', and 'Files (550)'. The main results area shows '1 to 10 of 184 Results' and a single result entry: 'Matlab code for estimation of permeability through image processing of microCT images' with a 'Draft' status and 'Unpublished' label, dated 'May 28, 2019 - C03'. A 'Sort' button is located at the bottom right of the results area.

Requirements of our Users

Technical University, main focus on :

- Engineering
- Natural sciences
- Digital humanities
- Social sciences

Publishing data is not anchored in professional culture

Requirements:

- Handling large amounts of data (TB, single file up to 150 GB)
- Describing data with discipline specific metadata
- Overview over (a lot of) hot data with expanded parameter space

Aim

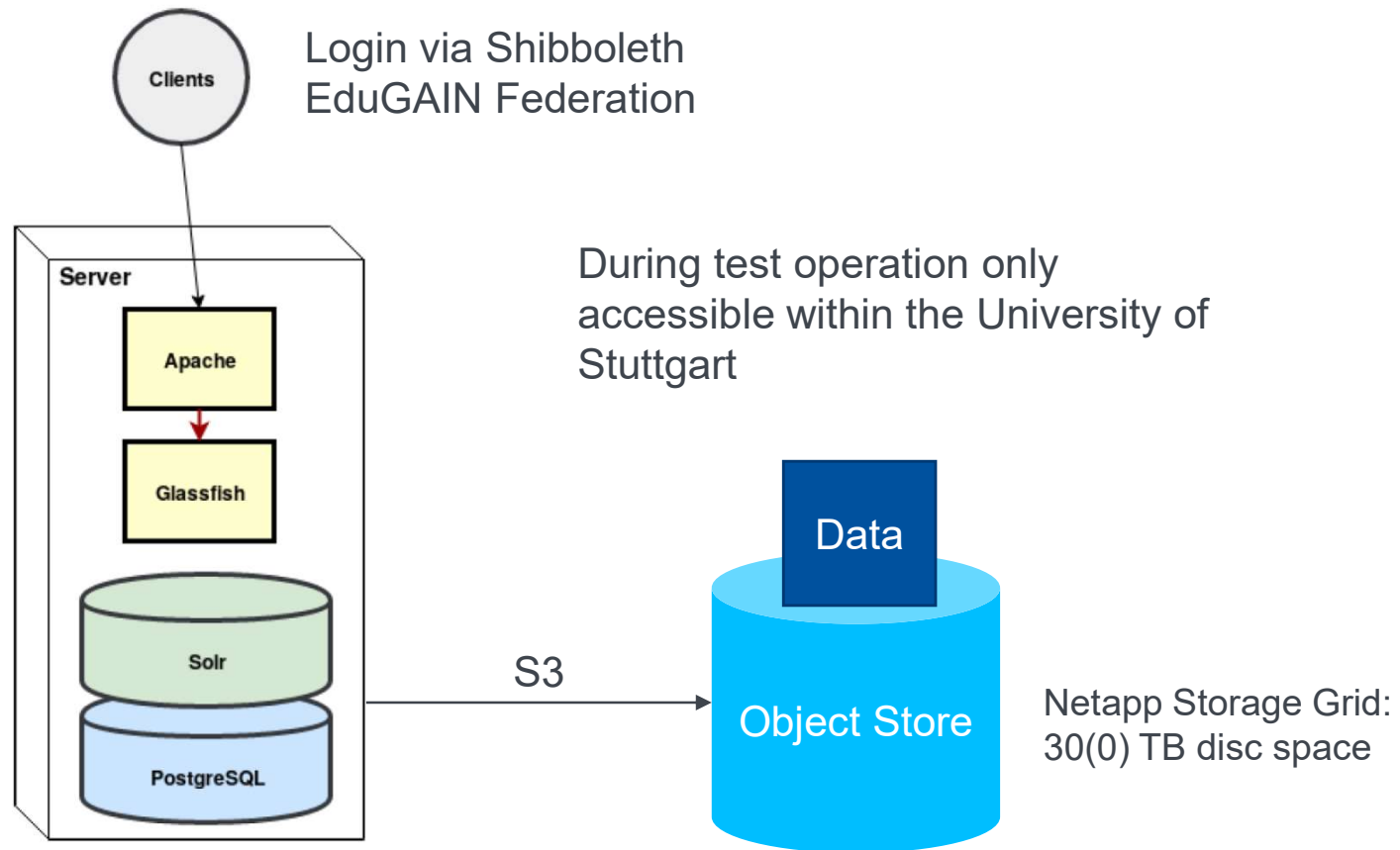
Why Dataverse?

- Data centered
- User and rights management
- Specific metadata per dataverse

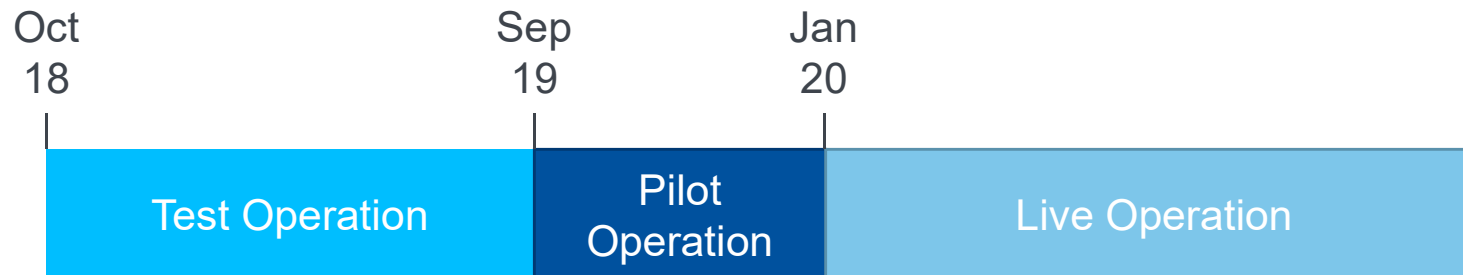
Application

- Managing and Sharing (Hot) Data
- Publication of Datasets

Architecture

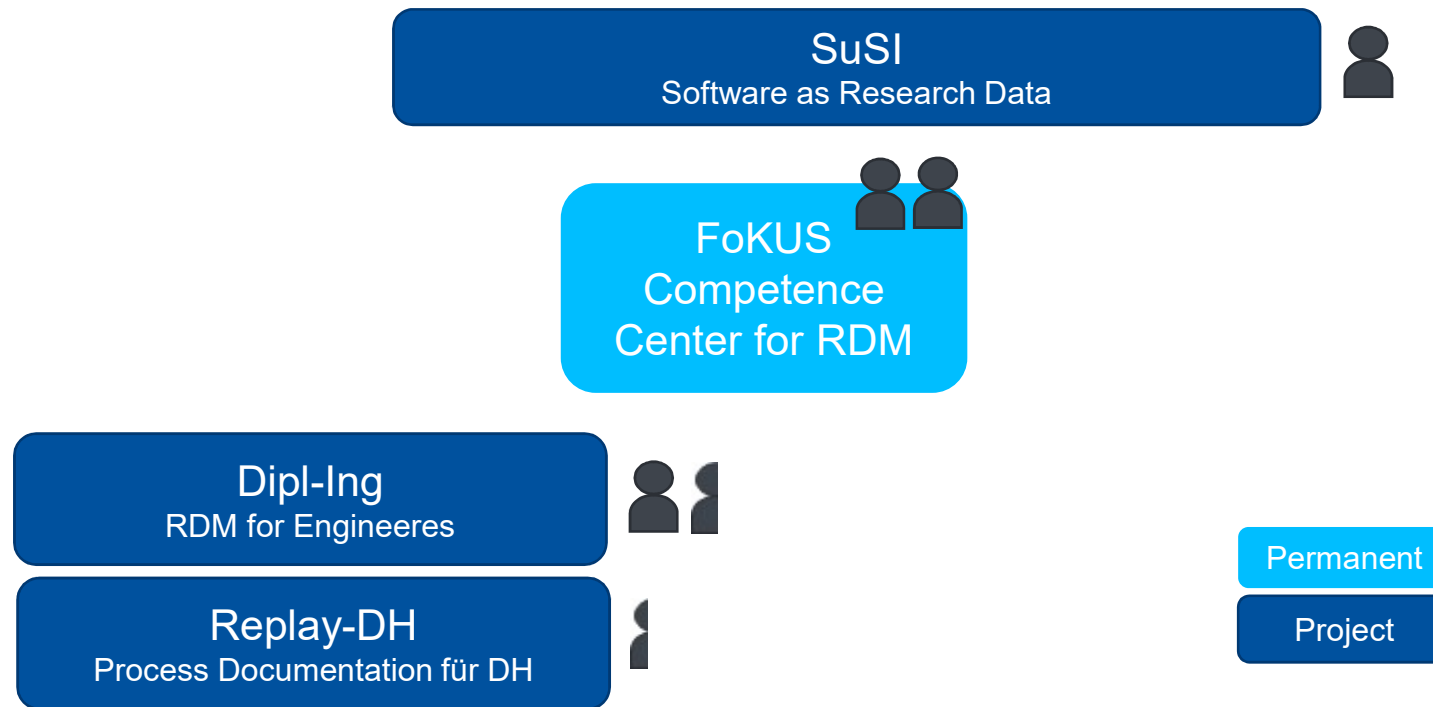


Timetable

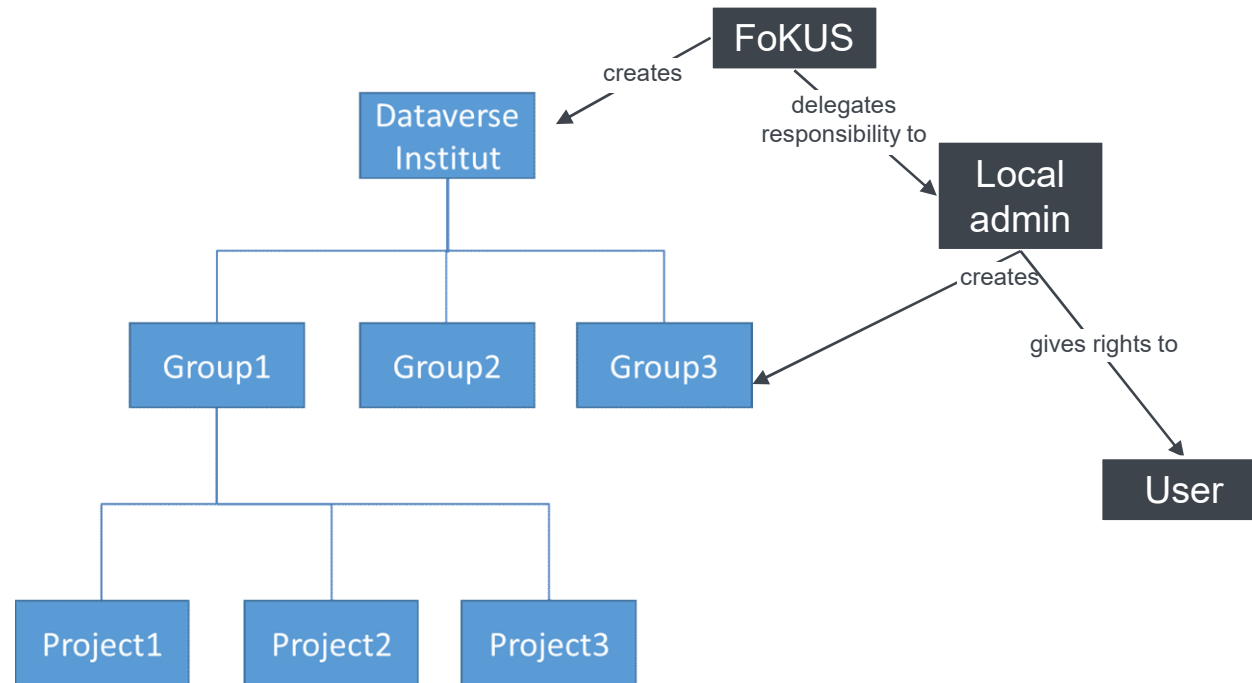


- Dummy DOIs
 - Test Users (SFB, EXC, 5 Institutes)
 - Regular meetings with test users
 - Test data sets
 - No data files > 100 GB (30 TB Storage)
- Real DOIs
 - Real datasets
 - Regular meetings with pilot users
 - Direct upload of large data files to Object Store
- Operation by computing center (not by project staff)
 - Open for the whole university
 - Half-yearly meetings with local admins

Organisation and Staff



Organisation of Dataverses



do not think in folders (Dataverses), but build structure over metadata and then filter
→ Dataverses important only for rights

New Metadata Blocks

- Process Metadata
- Engineering Metadata

Metadata Fields

 Choose the metadata fields to use in dataset templates and when adding a dataset to this dataverse.

Use metadata fields from DaRUS

Citation Metadata (Required) [\[+\] View fields + set as hidden, required, or optional](#)

Geospatial Metadata [\[+\] View fields + set as hidden, required, or optional](#)

Social Science and Humanities Metadata [\[+\] View fields + set as hidden, required, or optional](#)

Astronomy and Astrophysics Metadata [\[+\] View fields + set as hidden, required, or optional](#)

Life Sciences Metadata [\[+\] View fields + set as hidden, required, or optional](#)

Journal Metadata [\[+\] View fields + set as hidden, required, or optional](#)

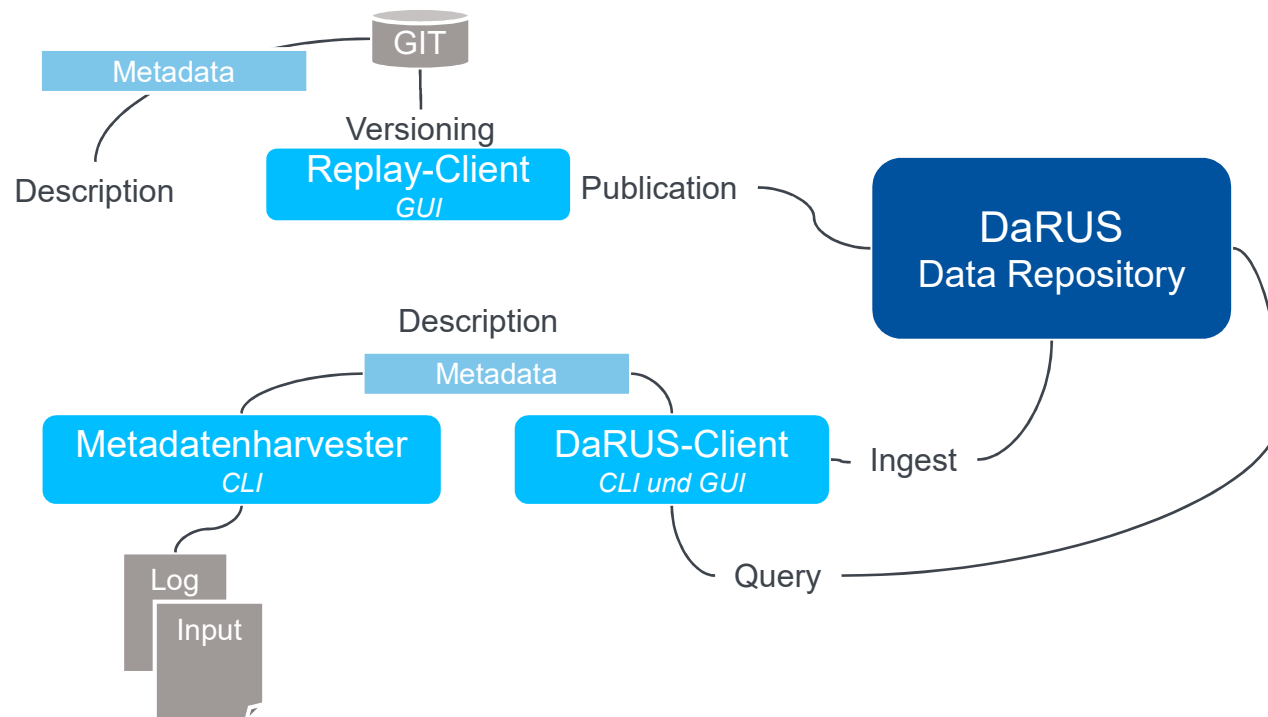
Process Metadata [\[+\] View fields + set as hidden, required, or optional](#)

Engineering Metadata [\[+\] View fields + set as hidden, required, or optional](#)

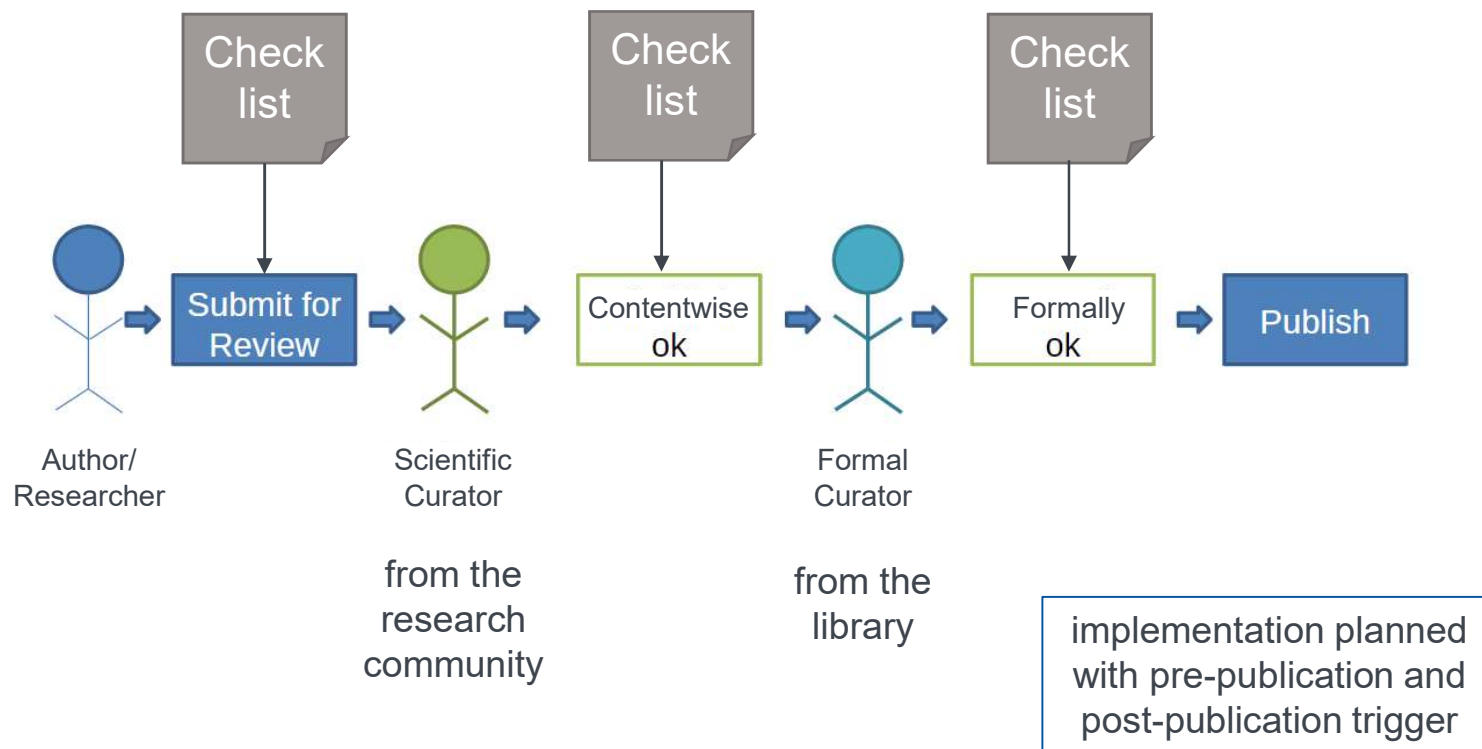
Experiences and actual Development

Ingest of the (meta)data - Automation

Start early in the process



Publishing Workflow



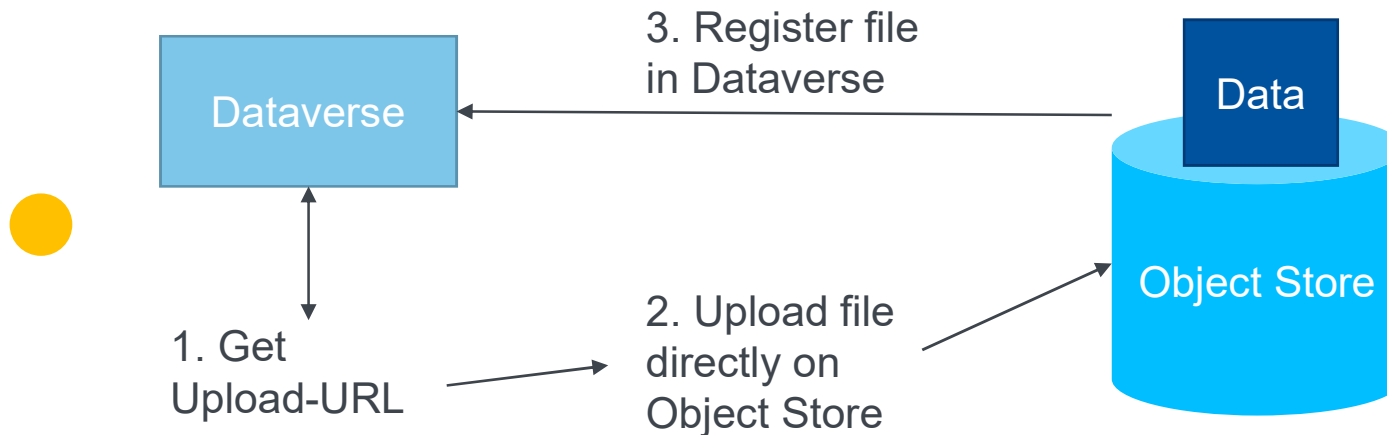
Handling of Large Amounts of Data

Separate Glassfish threads:

- Default time-outs for web interface
- Long time-outs for API interface



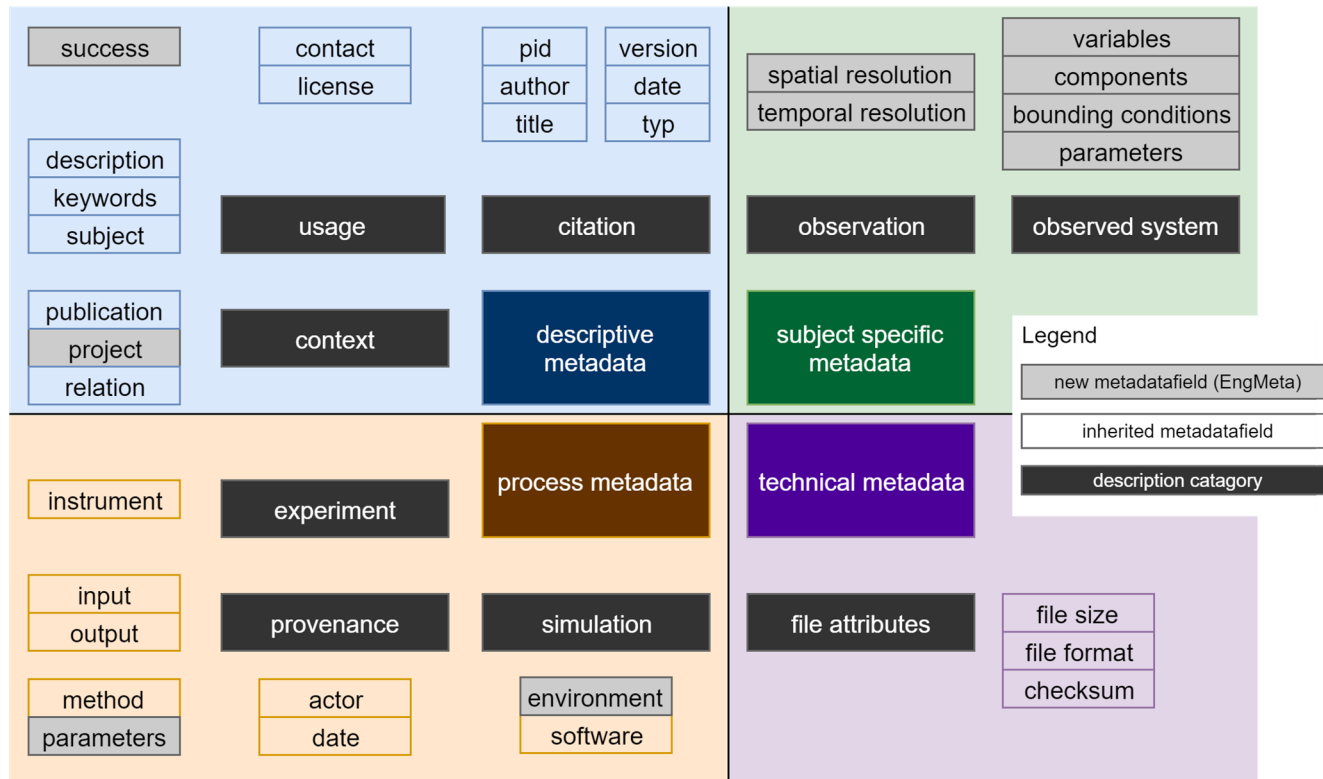
Upload up to 100 GB per file



- Connection of Object Store to tape library (IBM Spectre Protect)

Discipline Specific Metadata

EngMeta – Metadata for Engineering



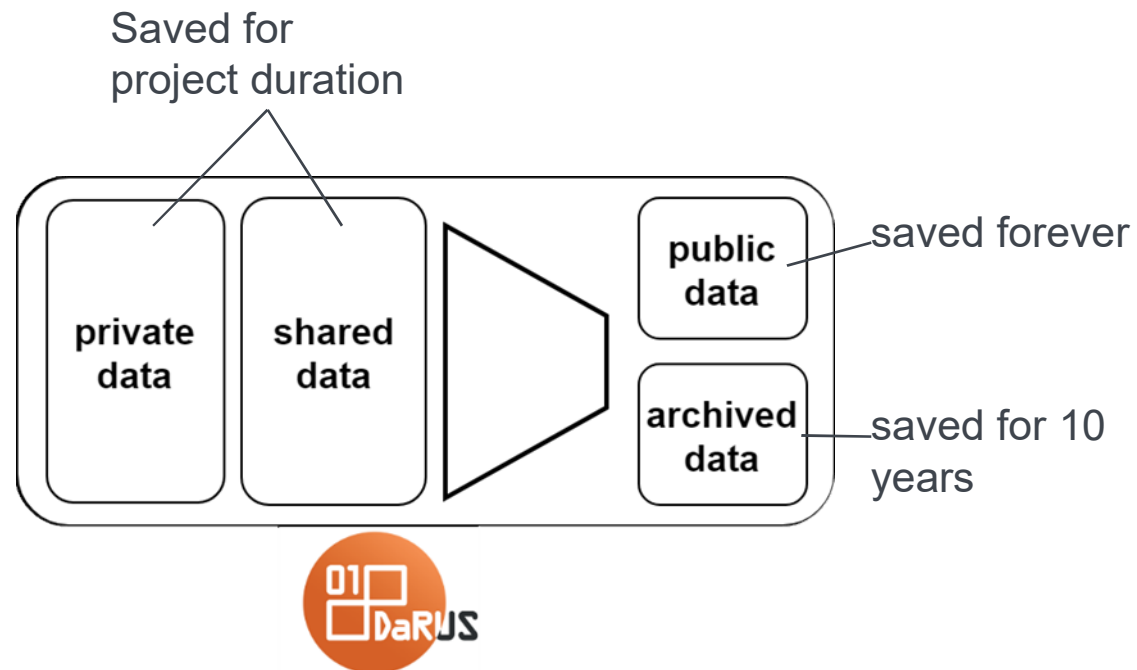
Overview Over Data

- Tabular overview over expanded parameter space
- Filter and search options for numerical metadata

Title	Date	Mach Number	Reynolds Number
Title linked to Dataset			

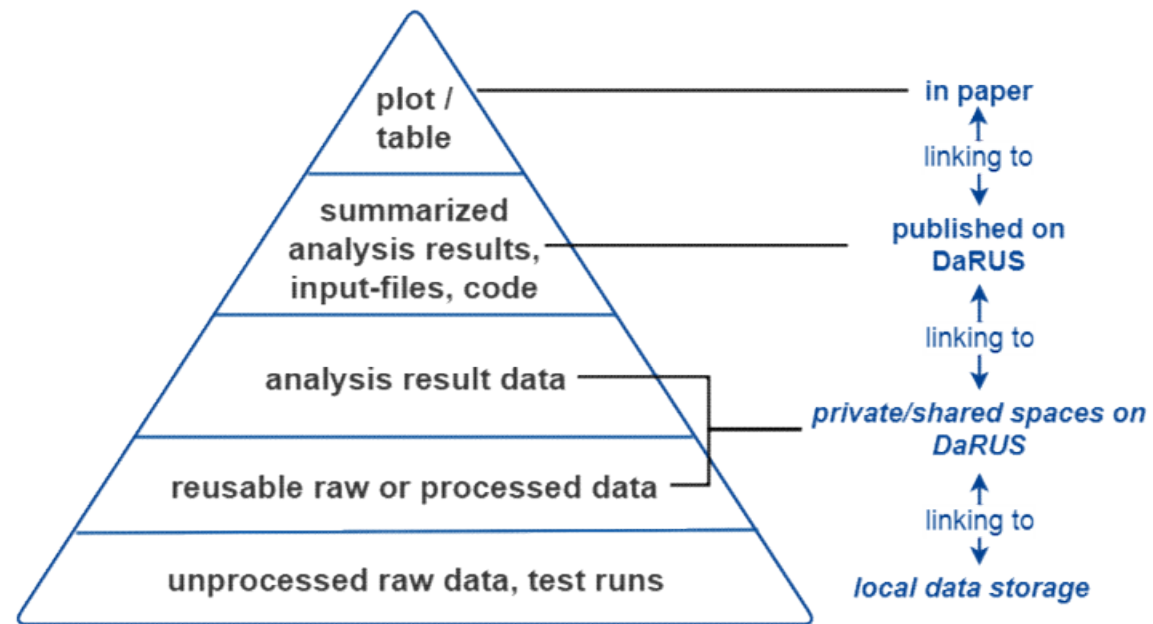
Deletion Management

First Ideas



Data Linking

Data pyramid



Datenpyramide für die Simulationwissenschaften, adaptiert nach (Reilly, et al. 2011)

Collaboration

- S3 interface to object storage
- Curation and publication workflows
- Software as research data
- Quality management and deletion policies
- Management of unmovable data
- Interfaces to different (discipline specific) repositories and databases