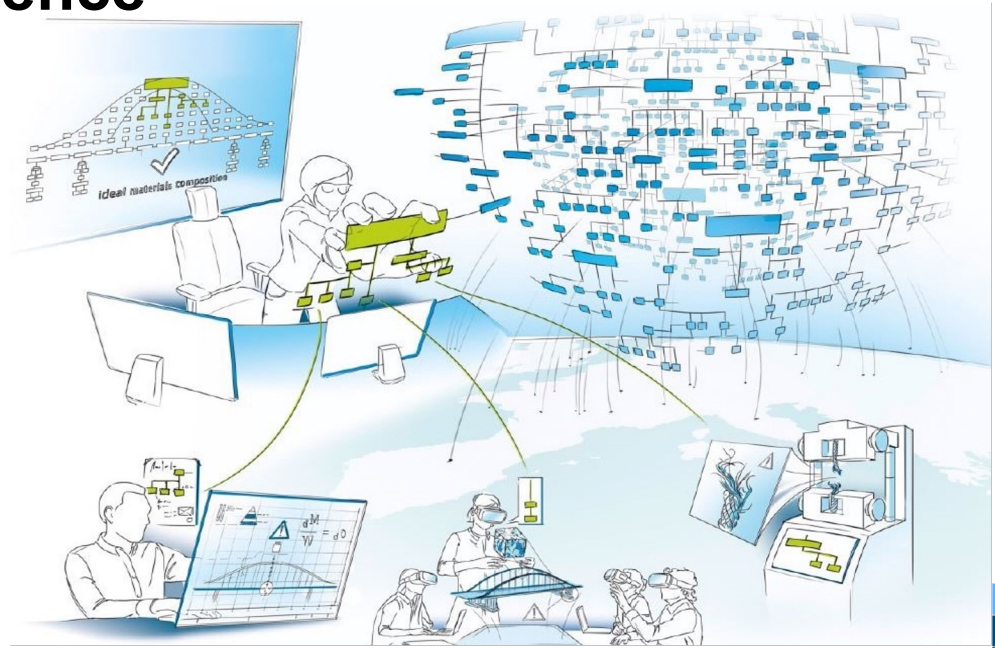
Linking Research Data

NFDI-MatWerk Conference

Siegburg, 27.-29.06.23



Workshop Team:

Ajay Kirar

Benedikt Heinrichs

Reetu Joseph

Rossella Aversa

Sirieam Hunke

Yusra Shakeel

Agenda

- Introduction
- Tutorial: NFDI-MatWerk Data and Metadata Repositories
  - Q/A session
- Presentation: Accessing Multiple Metadata Repositories
  - Q/A Session
- Tutorial: Recipe to create a FAIR DO
  - Q/A Session
- Wrap-up

Introduction

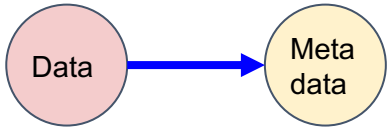
Aim of the workshop

- Introduce the FAIR Digital Object concept.
- Illustrate how researchers can easily handle their data according to the FAIR principles.

Introduction: FAIR Digital Object

Data object

- Link to the location
  e.g. storage location of the **data file**
- Link(s) to metadata object(s)
- PID
- Type
- ….



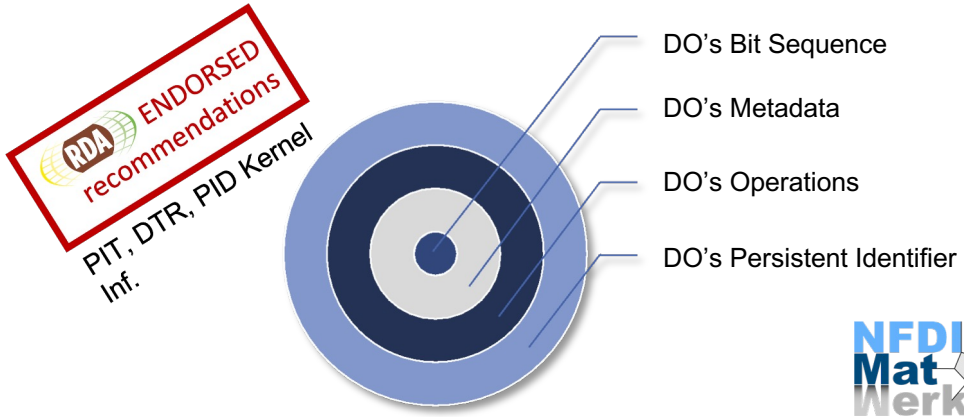
Metadata object

- Link to the location
  e.g. storage location of the **metadata document**
- PID
- Type
- ….

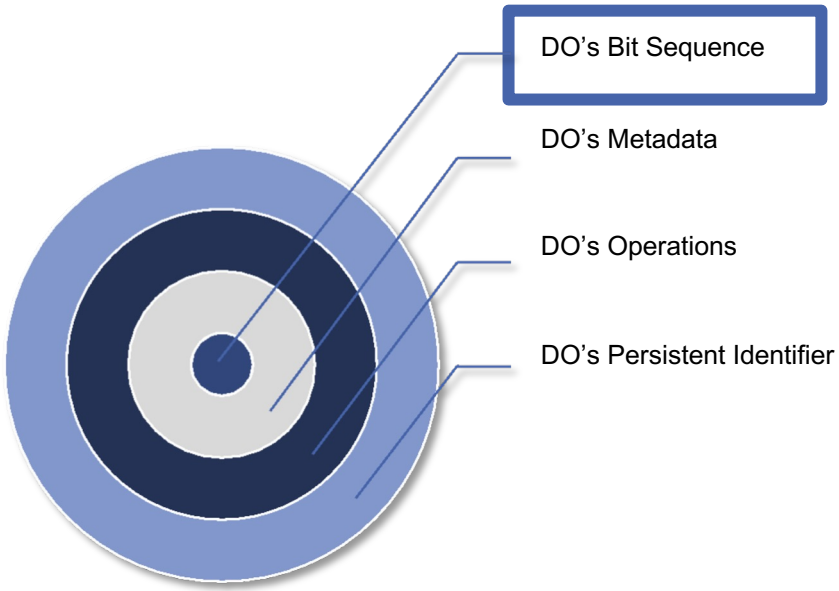
FAIR Digital Object:

- Representation of data
- Contains all information towards FAIR
- Conception: technology agnostic
- Bridges between data repositories, disciplines, etc.
- Implementation
  - Handle PID (Persistent Identifier)
  - Information Record = DO's Metadata

PID	PID Profile	Type	Location URL	. . .



Content of a FAIR DO

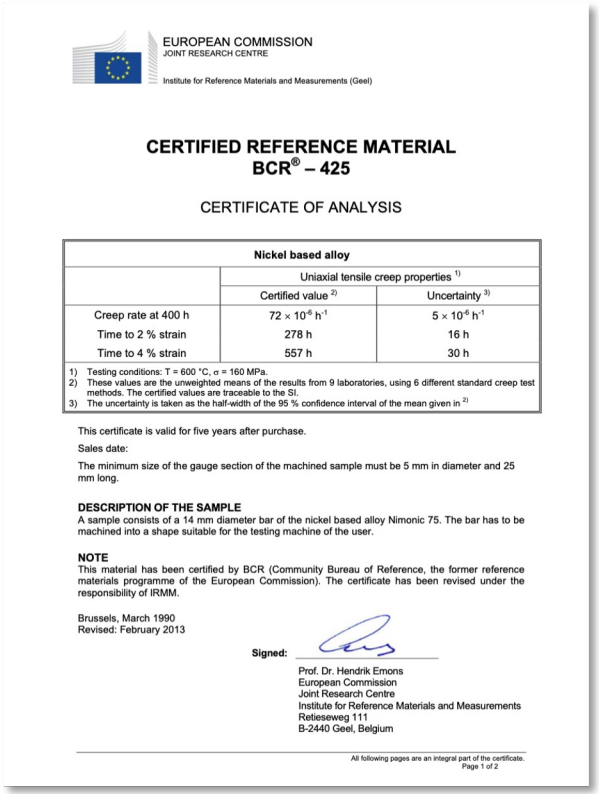


Data/Metadata in digital format

Example: PP18 Reference Dataset

- For creep tests on a reference material of Nimonic 75 nickel-base alloy
- Consists of different files - raw data (.lis), certificates, characterization and descriptions (.pdf)
- On Zenodo (exportable to JSON, JSON-LD)
- We demonstrate with 1 pdf file from the dataset: **BCR-425_cert.pdf**

BCR-425_cert.pdf and its Metadata



EUROPEAN COMMISSION
JOINT RESEARCH CENTRE
Institute for Reference Materials and Measurements (Geel)

CERTIFIED REFERENCE MATERIAL
BCR® – 425

CERTIFICATE OF ANALYSIS

Nickel based alloy		
	Uniaxial tensile creep properties [1]	
	Certified value [2]	Uncertainty [3]
Creep rate at 400 h	72×10^{-6} h^{-1}	5×10^{-6} h^{-1}
Time to 2 % strain	278 h	16 h
Time to 4 % strain	557 h	30 h

1) Testing conditions: T = 600 °C, σ = 160 MPa.
2) These values are the unweighted means of the results from 9 laboratories, using 6 different standard creep test methods. The certified values are traceable to the SI.
3) The uncertainty is taken as the half-width of the 95 % confidence interval of the mean given in [2]

This certificate is valid for five years after purchase.

Sales date:

The minimum size of the gauge section of the machined sample must be 5 mm in diameter and 25 mm long.

DESCRIPTION OF THE SAMPLE
A sample consists of a 14 mm diameter bar of the nickel based alloy Nimonic 75. The bar has to be machined into a shape suitable for the testing machine of the user.

NOTE
This material has been certified by BCR (Community Bureau of Reference, the former reference materials programme of the European Commission). The certificate has been revised under the responsibility of IRMM.

Brussels, March 1990
Revised: February 2013

Signed:

Prof. Dr. Hendrik Emons
European Commission
Joint Research Centre
Institute for Reference Materials and Measurements
Retieseweg 111
B-2440 Geel, Belgium

All following pages are an integral part of the certificate.
Page 1 of 2

Metadata in JSON format, exported from Zenodo


```
{
  "links": {
    "self": "https://zenodo.org/api/files/2e8ca55b-dfe8-45fd-93a1-d79d5500528f/BCR-425_cert.pdf"
  },
  "checksum": "md5:6353de153830c0a28b3ffafedf59ec23",
  "bucket": "2e8ca55b-dfe8-45fd-93a1-d79d5500528f",
  "key": "BCR-425_cert.pdf",
  "type": "pdf",
  "size": 37659
}
```

Research Data and Metadata Repositories

Data Management system used to store, manage and provide access to digital resources (Data/Metadata) following a policy or set of rules that define storage and access norms. -source: MDMC Glossary

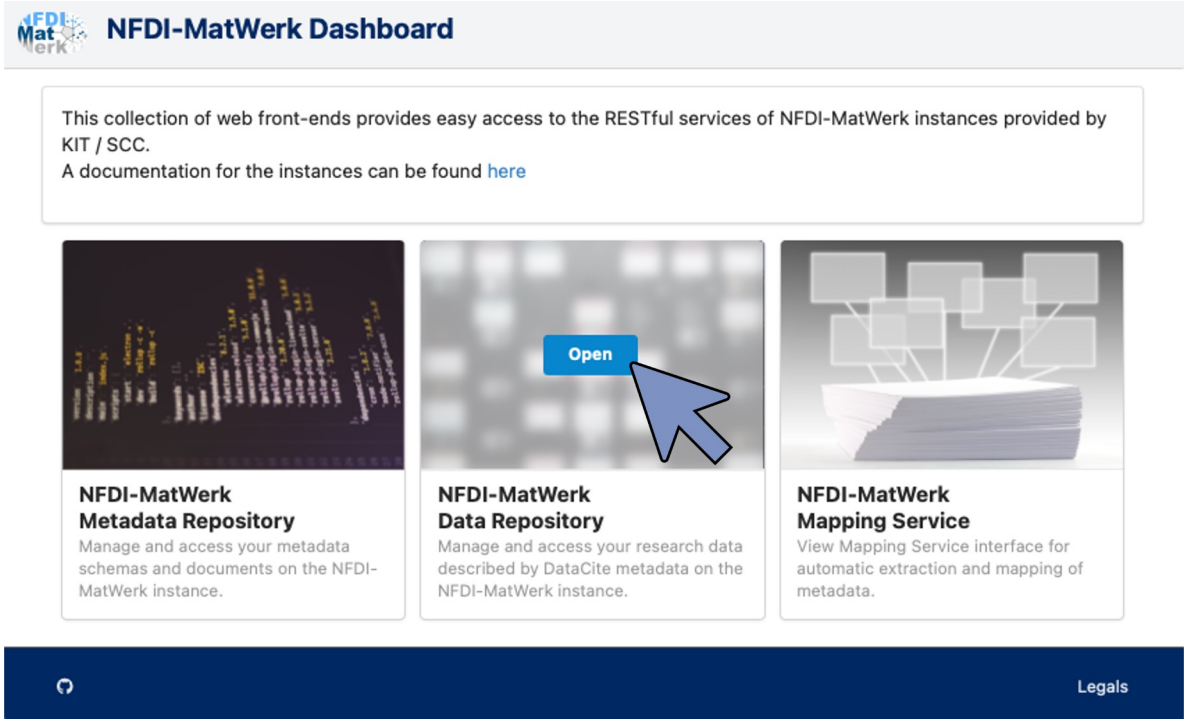
Spotlight Today:
NFDI- Matwerk Data and Metadata Repositories

Advantages of NFDI-MatWerk Repositories

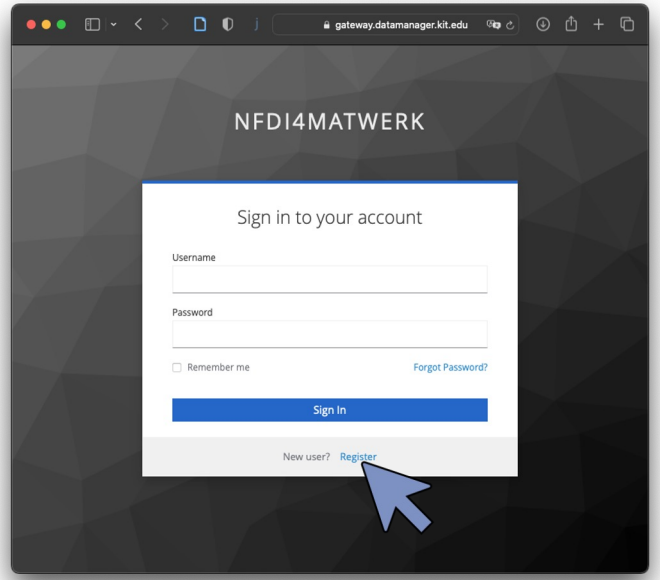
- Easy to use GUI, uses DataCite Metadata Schema for resources
- Integration to other scientific workflows and software using the (customisable) Application Programming Interface (API)
- Possibility of installing local instances for high performance storage, large amounts of data and privacy
- Data Repository allows live interaction with the contents
- Metadata Repository supports an arbitrary number of XML and JSON schemas including versioning and automatic validation
    - Multiple metadata documents can describe the same file from different perspectives

Accessing the NFDI-MatWerk Repositories

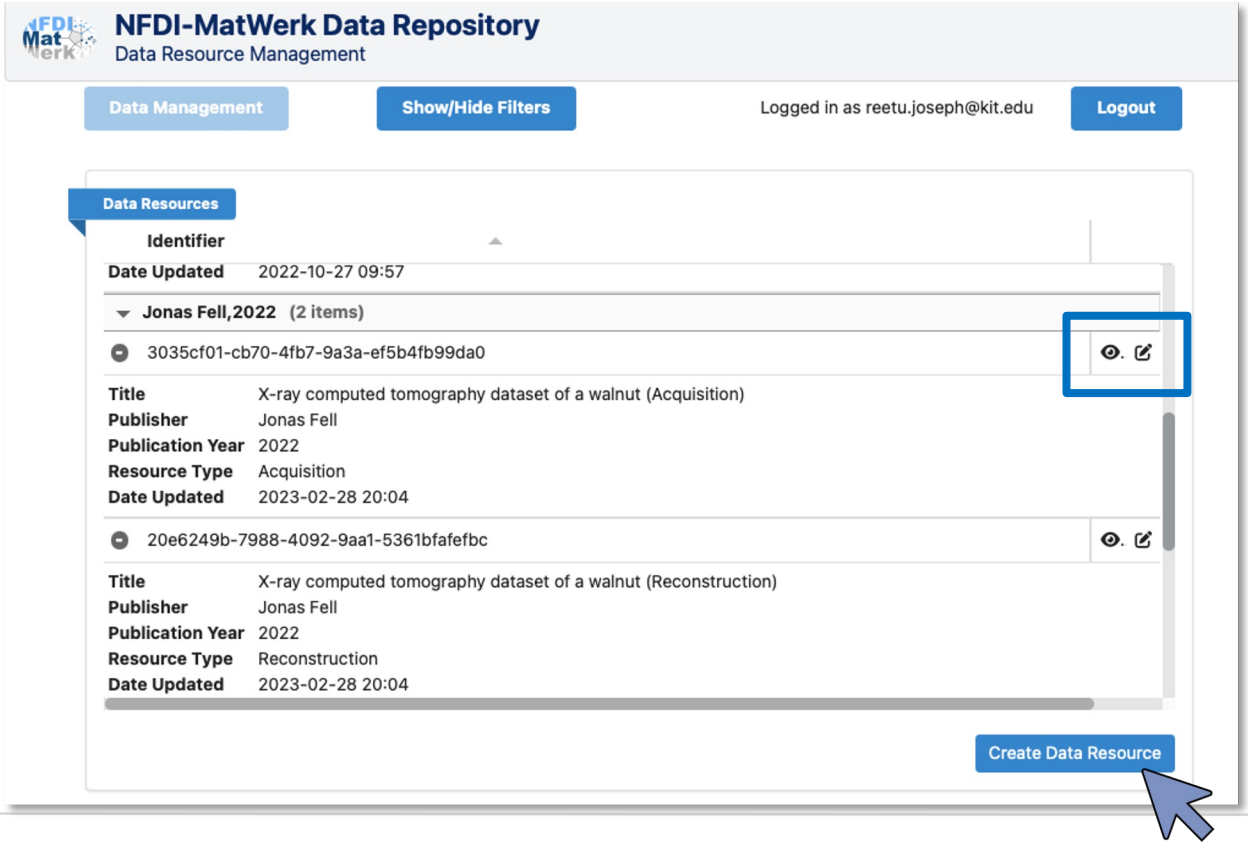
https://matwerk.datamanager.kit.edu

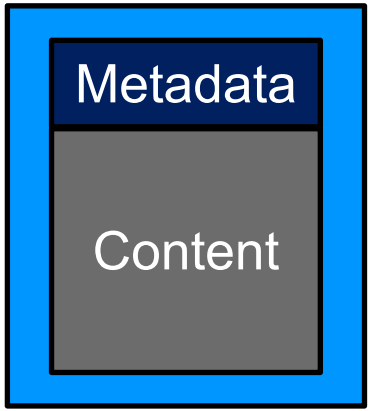


Create User Credentials



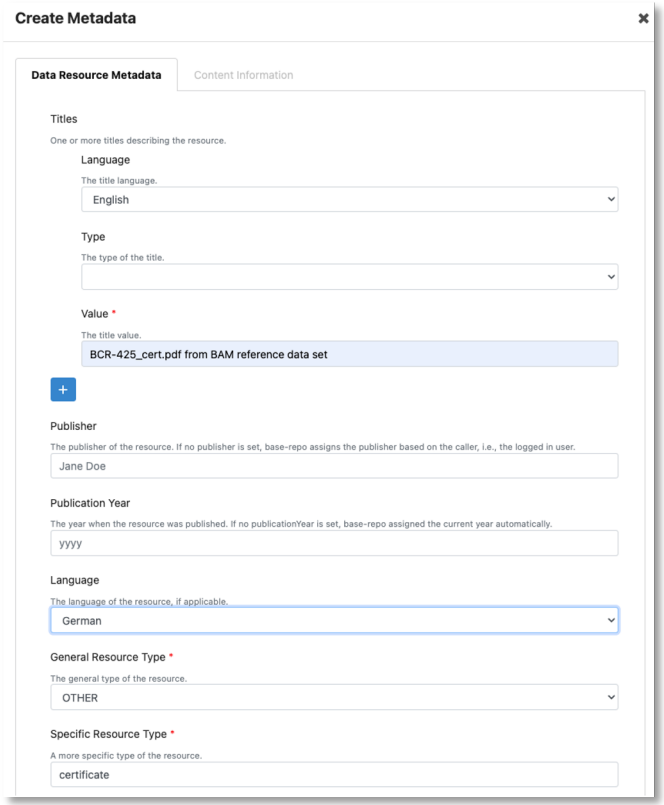
Upload a File on the NFDI-MatWerk Data Repository

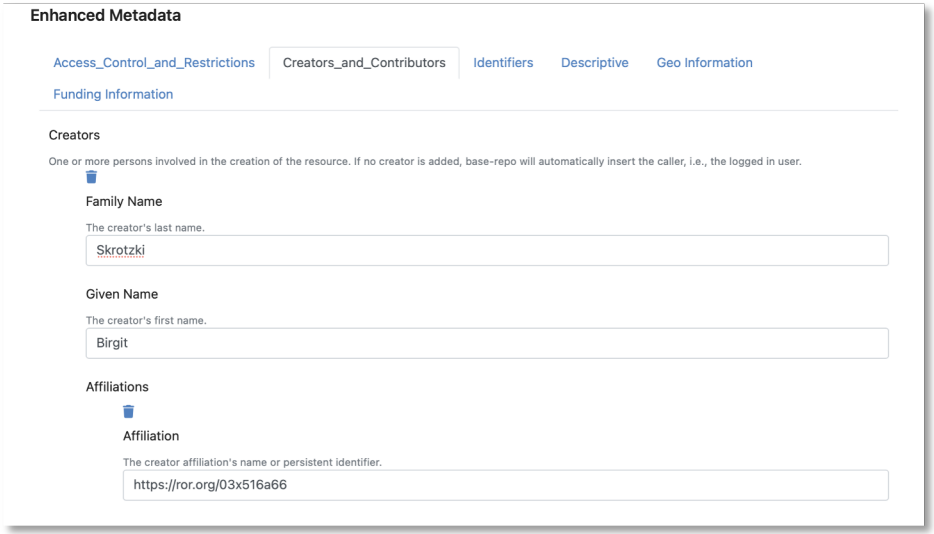




Data Resource

Enter Metadata for the Data Resource

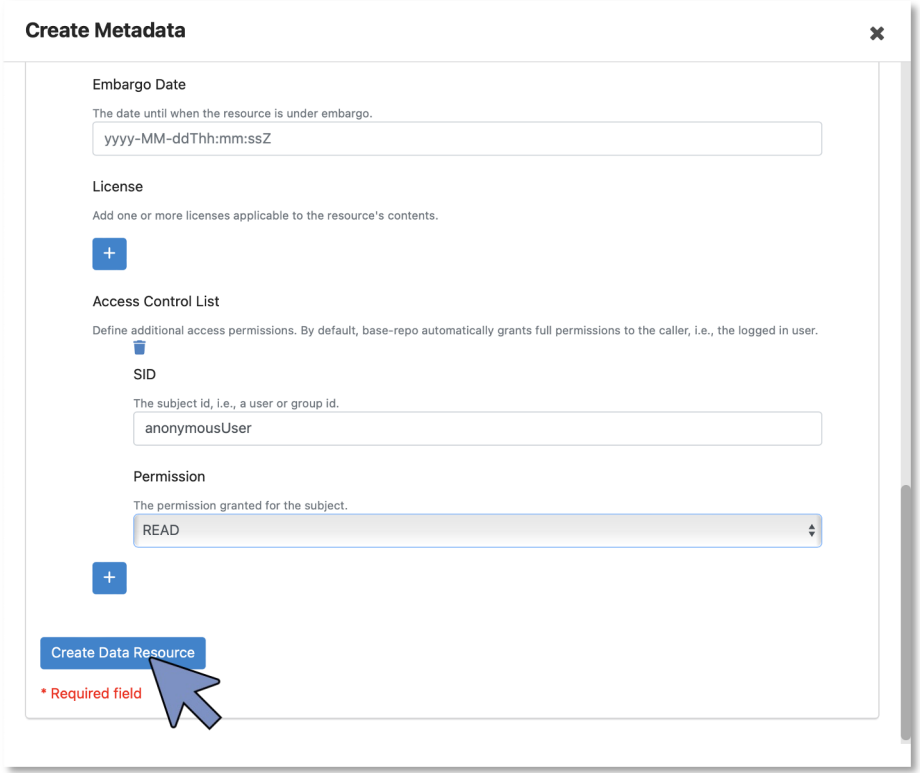




Metadata of the data resource of **BCR-425_cert.pdf** from the BAM reference dataset (PP18)

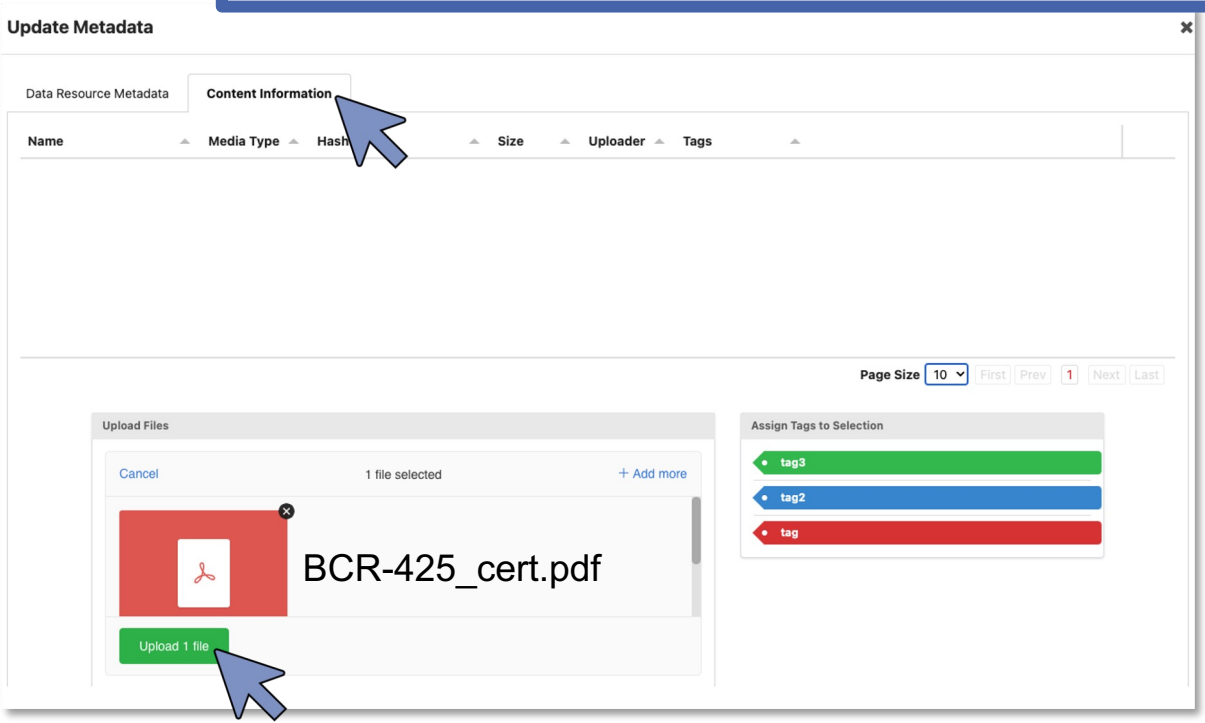
Access Control for the Data Resource

For public visibility,
“SID”(subject id): “anonymousUser”
Permission: “READ”



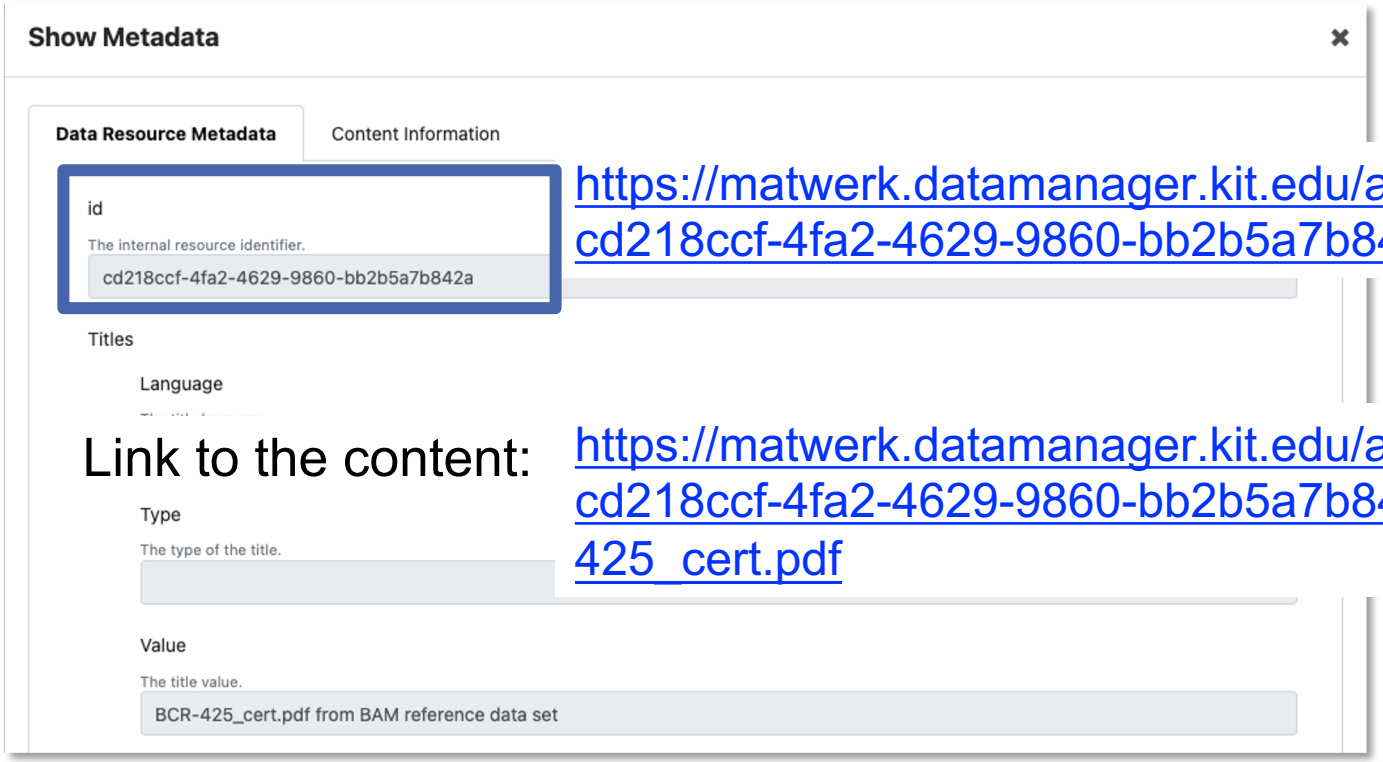
Drop or Browse to upload the Data

Select the “Content Information” tab to upload the data



Drag & drop or browse the file(s) to be uploaded

Persistent Identifier of the Resource



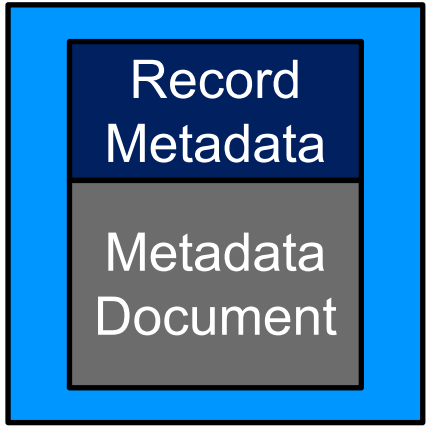
https://matwerk.datamanager.kit.edu/api/v1/dataresources/cd218ccf-4fa2-4629-9860-bb2b5a7b842a

Link to the content: https://matwerk.datamanager.kit.edu/api/v1/dataresources/cd218ccf-4fa2-4629-9860-bb2b5a7b842a/data/BCR-425_cert.pdf

Upload a Metadata Document in NFDI-MatWerk Metadata Repository

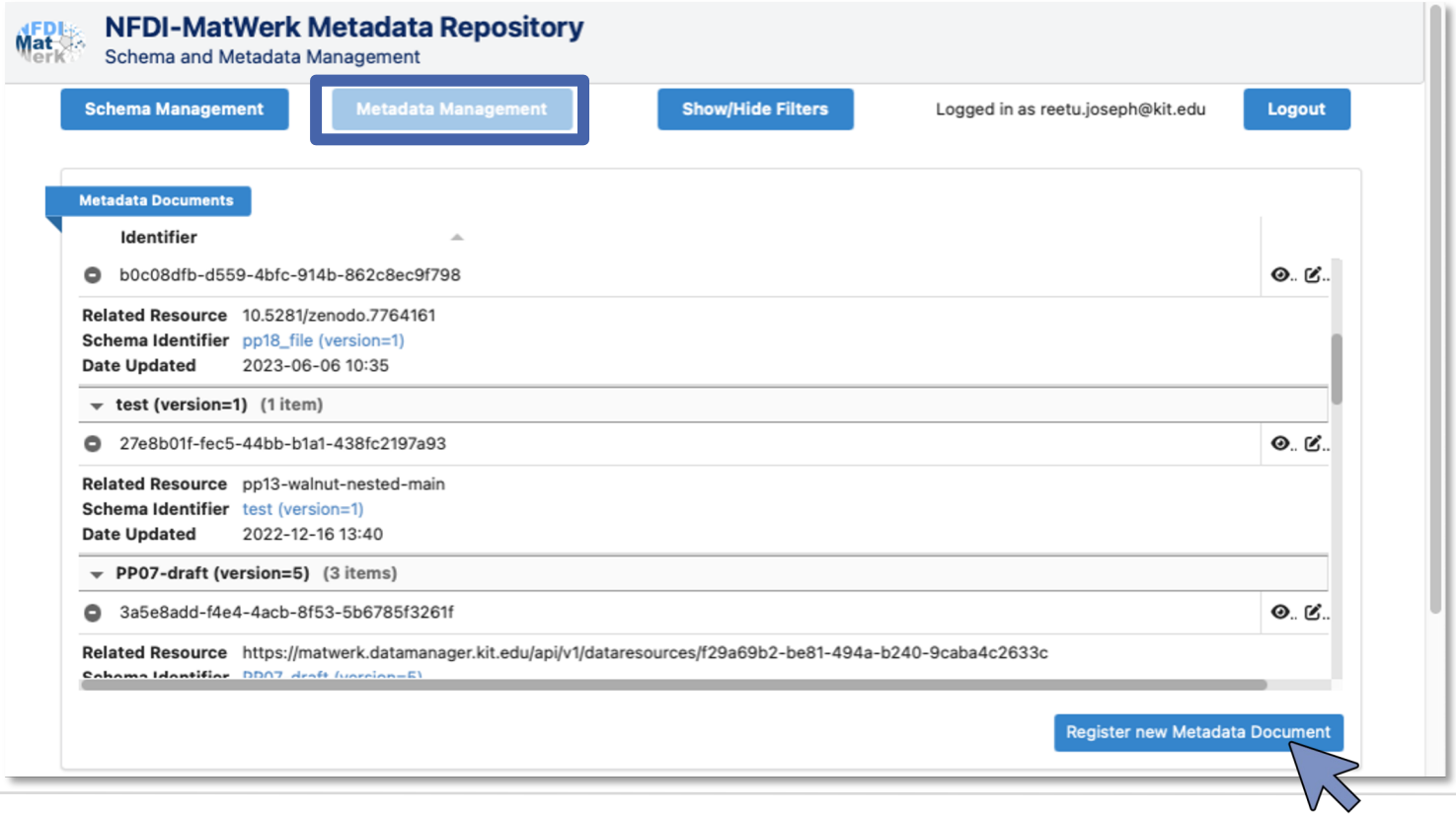
Metadata in JSON format, exported from Zenodo


```
{
  "links": {
     "self": "https://zenodo.org/api/files/2e8ca55b-dfe8-45fd-
     93a1-d79d5500528f/BCR-425_cert.pdf"
  },
  "checksum": "md5:6353de153830c0a28b3ffafedf59ec23",
  "bucket": "2e8ca55b-dfe8-45fd-93a1-d79d5500528f",
  "key": "BCR-425_cert.pdf",
  "type": "pdf",
  "size": 37659
}
```

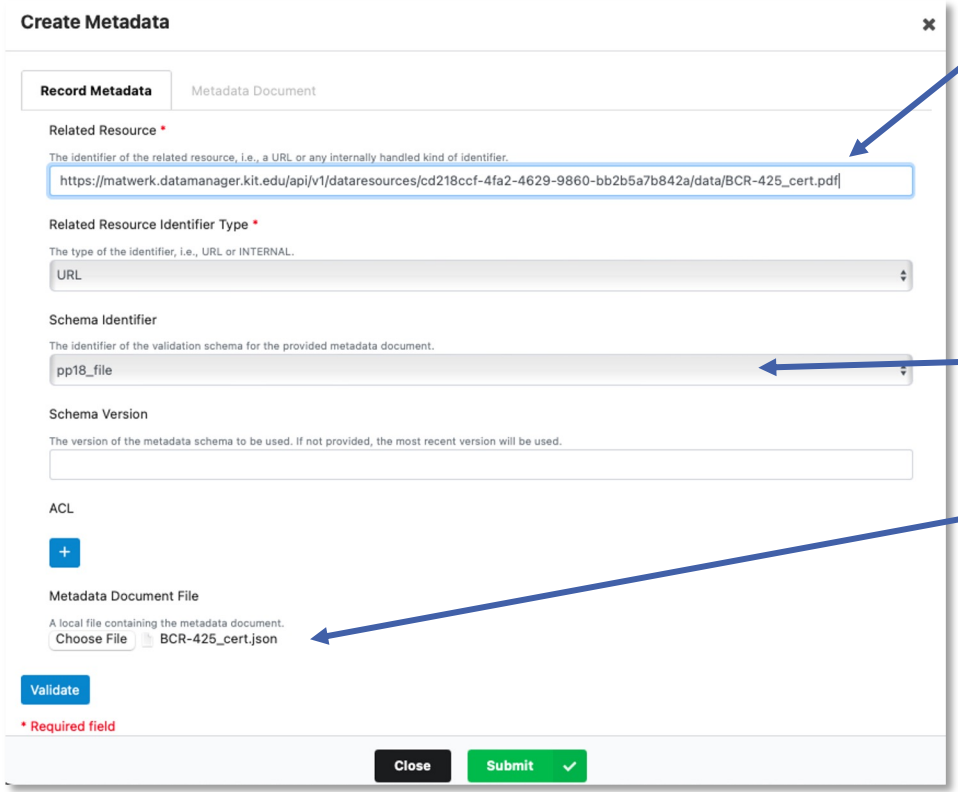



Metadata Document Record

Register the Metadata Document

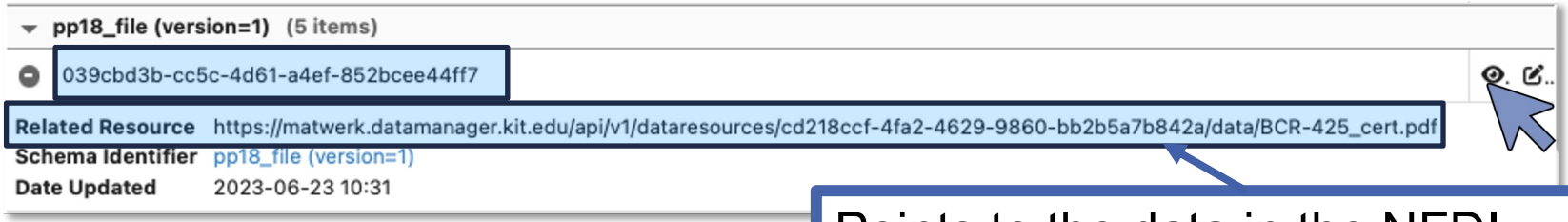


Register the Metadata Document

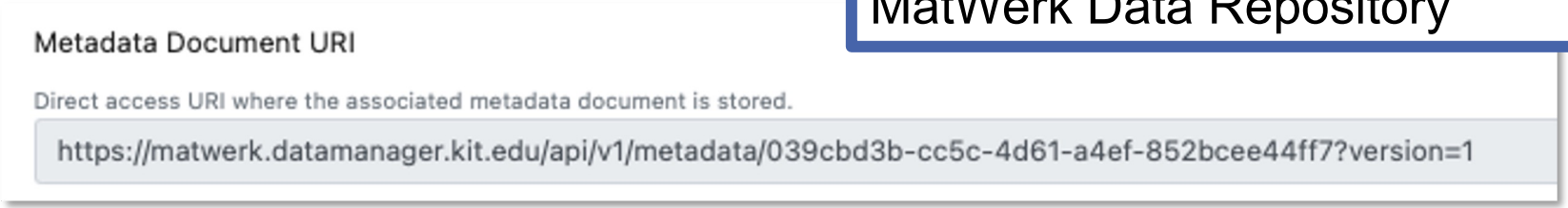


1. Link the metadata document of **BCR-425_cert.pdf** to the related data resource (https://matwerk.datamanager.kit.edu/api/v1/dataresources/cd218ccf-4fa2-4629-9860-bb2b5a7b842a/data/BCR-425_cert.pdf)
2. Choose the metadata schema by selecting the identifier of the schema from the drop-down list
3. Choose the metadata document "**BCR-425_cert.json"** to be uploaded
4. Click Validate and if valid, the uploaded document can now be viewed under the "Metadata Document" Tab, then click Submit

Persistent Identifier of Metadata Document



Points to the data in the NFDI-MatWerk Data Repository



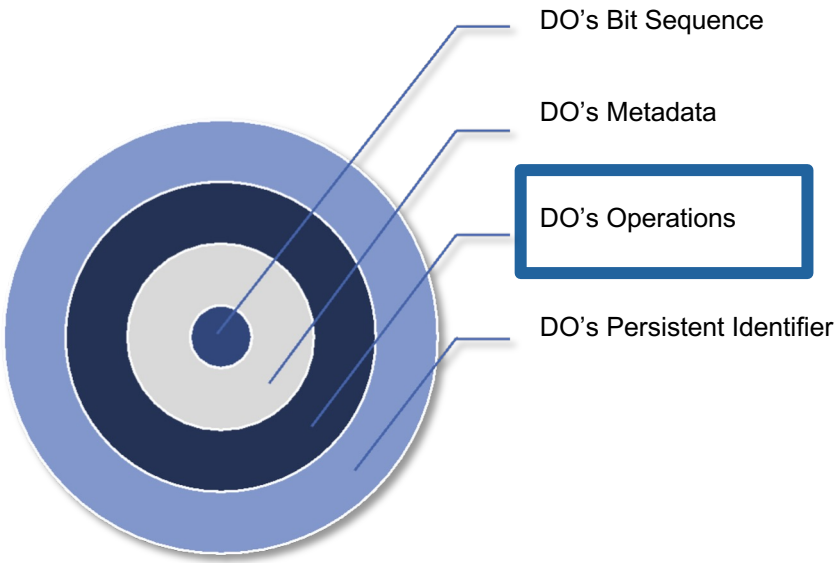
https://matwerk.datamanager.kit.edu/api/v1/metadata/039cbd3b-cc5c-4d61-a4ef-852bcee44ff7?version=1

Wrap up

- Using MatWerk Data and Metadata repositories, we can have well managed data and metadata with persistent identifiers,
- Metadata document is linked to the data it describes
- Saving validated metadata document along with the schema aids interoperability and reuse
- Both metadata document and data can be found, accessed and referenced from outside (e.g. when creating FAIR DOs…)

Q & A

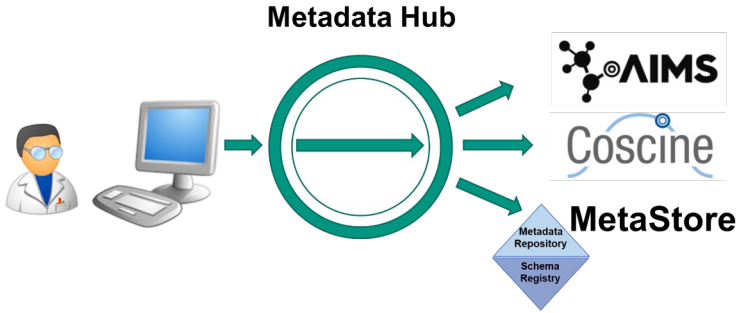
Accessing Multiple Metadata Repositories



Accessing Multiple Metadata Repositories

Metadata Hub

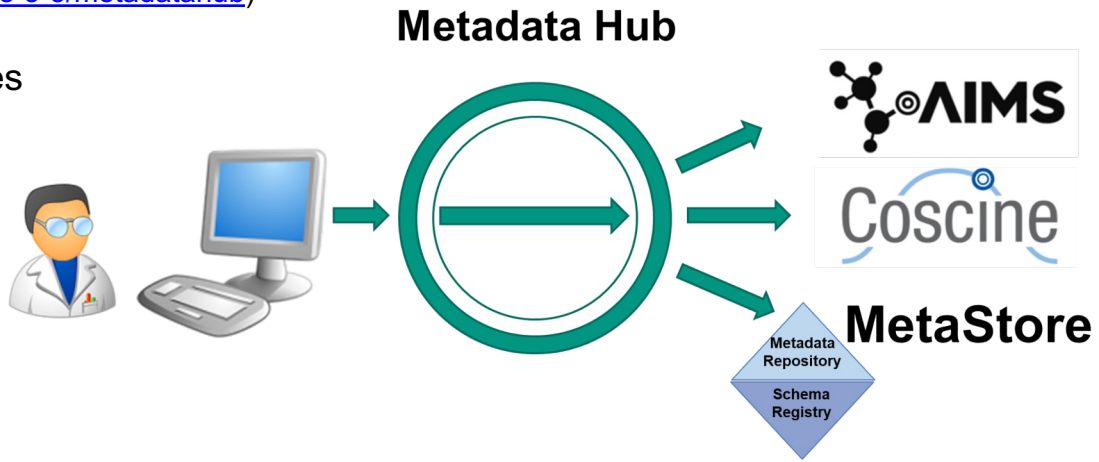
- Metadata repositories usually provide different interfaces
- Interoperability is wanted but cannot be achieved like this
- Goal:
  - Bring them together
- How:
  - A generic interface which combines different kinds of metadata repositories with one standard-based interface



Accessing Multiple Metadata Repositories

Metadata Hub

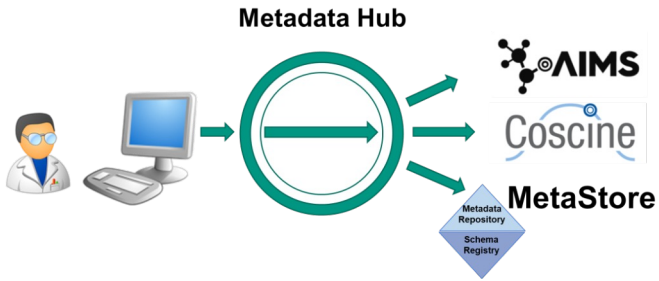
- Solution: The Metadata Hub
  (https://git.rwth-aachen.de/nfdi4ing/s-3/s-3-3/metadatahub)
- Brings together metadata stores
  - AIMS
  - Coscine
  - MetaStore
  - Your Solution?



Accessing Multiple Metadata Repositories

Metadata Hub

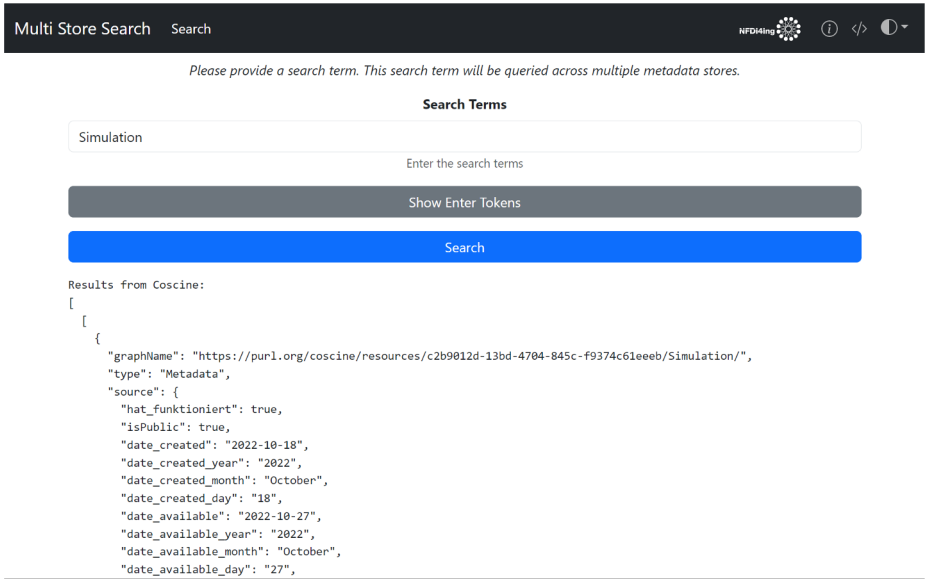
- For access → Turntable API (https://nfdi4ing.pages.rwth-aachen.de/s-3/s-3-3/turntable-interface/)
- Services are defined with a provided mapping file
- Pilots:
  - Coscine
  - MetaStore
- The difference in formats between the pilots was a big challenge



Accessing Multiple Metadata Repositories

Metadata Hub

- How to use it?
  (https://git.rwth-aachen.de/nfdi4ing/s-3/s-3-3/metadatahubdemonstrator)
  (https://git.rwth-aachen.de/nfdi4ing/s-3/s-3-3/multi-store-search)
  - Demonstrating UIs have been developed
  - Different metadata repositories can be selected
  - Requests like „Search“ can be sent
  - Multiple services can be queried

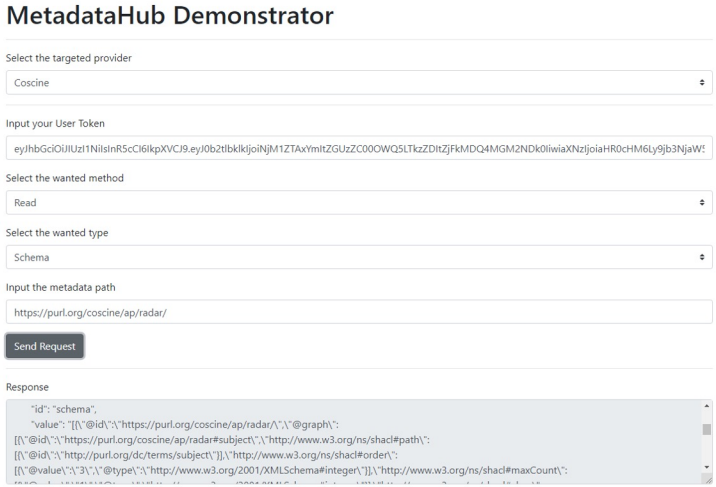


Accessing Multiple Metadata Repositories

Metadata Hub

- Example Use Case: Coscine (https://coscine.de/)

- Uses SHACL Application Profiles and RDF Metadata to annotate stored research data as FAIR Digital Objects (FDOs)
- By using the persistent identifier of a FDO, metadata of research data can be created, read or updated
- Example result:




```
dcterms:creator "Benedikt Heinrichs" .
dcterms:title "IC3K 2020 Poster" .
dcterms:created "2020-09-09"^^<http://www.w3.org/2001/XMLSchema#date> .
```

Q & A

Tutorial: Recipe to create a FAIR DO

Challenges

- There is a huge variety of **highly complex data** which is in **silos** in various repositories, cloud storage, local storage etc.
- Accessing data in silos can be quite challenging and for research data management there's no harmonized representation of data and metadata.

Way forward

- If we harmonize and have **technologically independent representation of data** we have various advantages like :
  - We can make representation of data into **actionable units**.
  - Build **scalable capabilities.** eg automation tools.
  - Easier **data linking**.

What is FAIR DO?

FAIR DO bind all critical information about an entity in one place and create a new kind of actionable, meaningful and technology independent object that pervades every aspect of life today :

A technical essence of a “thing” in cyberspace.

Source : https://fairdo.org/

What is FAIR DO?

FAIR DO bind all critical information about an entity in one place and create a new kind of actionable, meaningful and technology independent object that pervades every aspect of life today :

A technical essence of a “thing” in cyberspace.

→ Focus on machines

What is FAIR DO?

FAIR DO bind all critical information about an entity in one place and create a new kind of actionable, meaningful and technology independent object that pervades every aspect of life today :

A technical essence of a “thing” in cyberspace.

Relevant information extracted from something existing.

What is FAIR DO?

FAIR DO bind all critical information about an entity in one place and create a new kind of actionable, meaningful and technology independent object that pervades every aspect of life today :

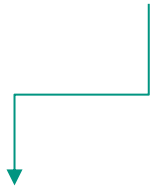
A technical essence of a “thing” in cyberspace.

Digital contents like repositories, softwares etc.

What is FAIR DO?

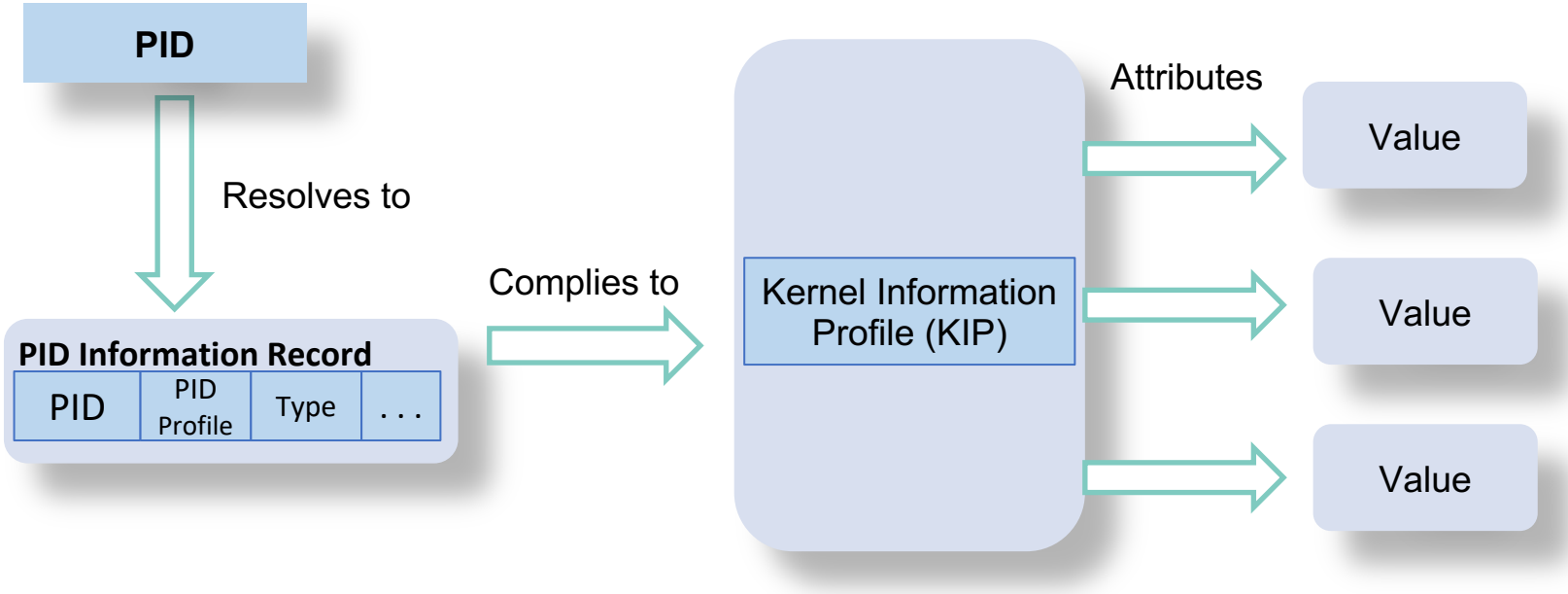
FAIR DO bind all critical information about an entity in one place and create a new kind of actionable, meaningful and technology independent object that pervades every aspect of life today :

A technical essence of a “thing” in cyberspace.

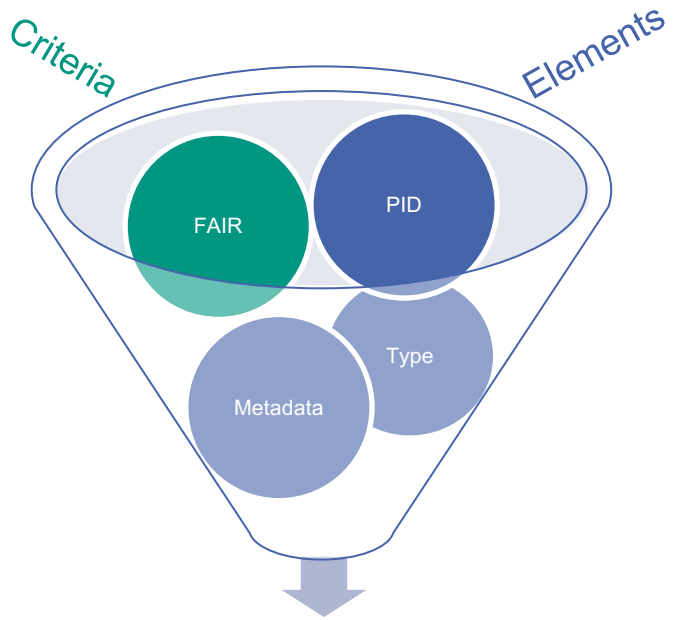


Digitaly present

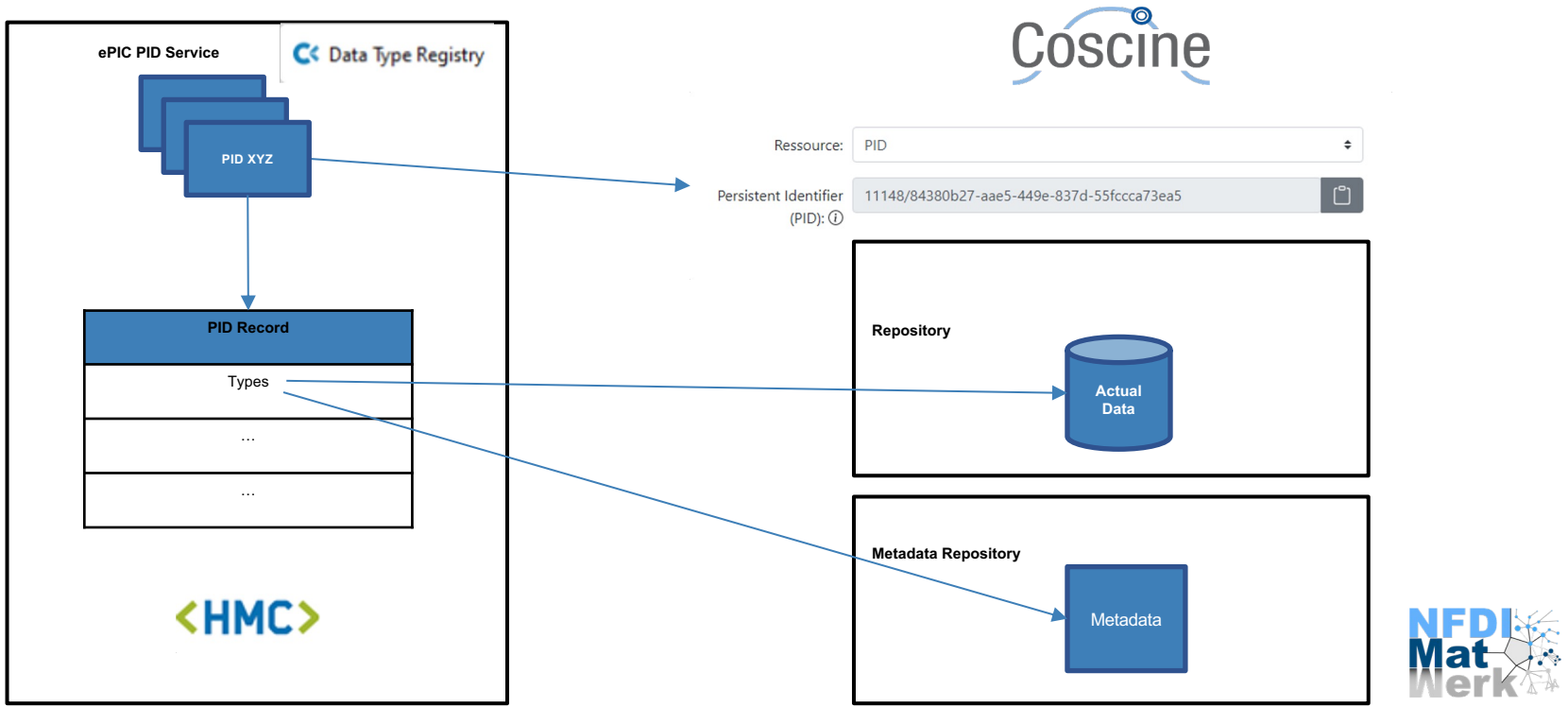
Key requirements to be a FAIR DO



What makes it a FAIR DO?

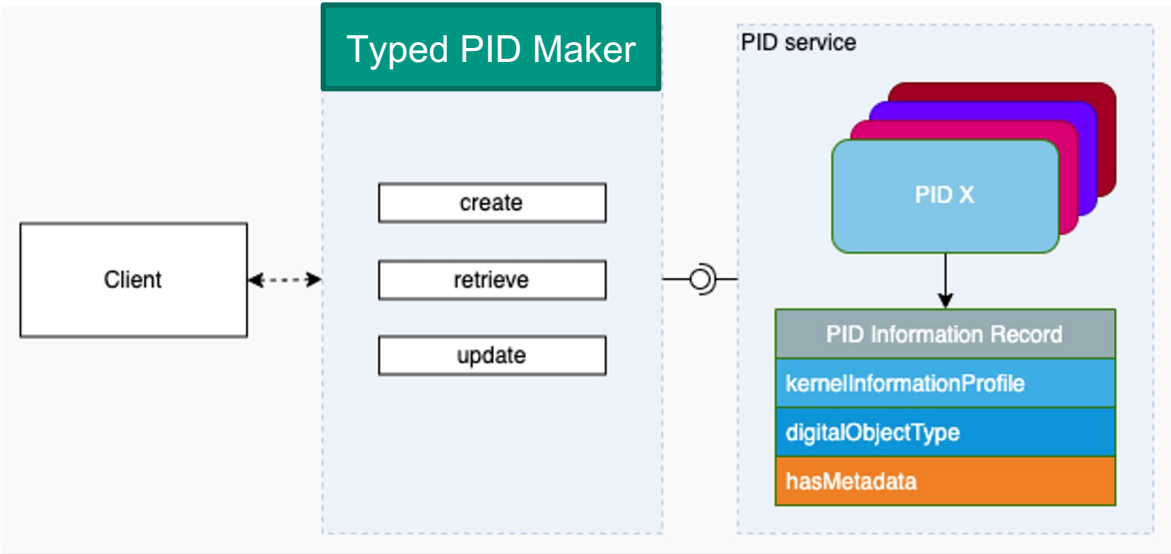


FAIR DO Creation – in action

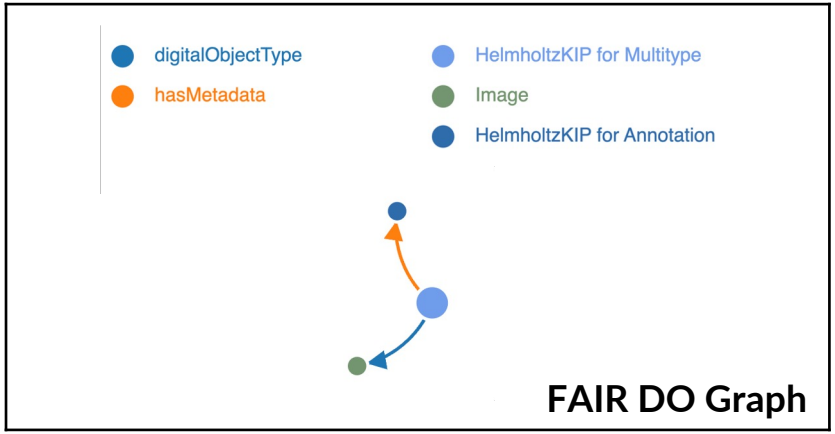


Typed PID Maker

- create, retrieve, update PIDs
- https://github.com/kit-data-manager/pit-service



Metadata



PID Information Record

Type	Value
kernelInformationProfile	21.T11148/863d938d632b53d62d52
digitalObjectLocation	https://b2share.eudat.eu/api/files/5fc88ad5-2f13-483c-8b80-a5862c91dbbb/Biological.tar#L7_dc3e2161576ff12aa04a2f6a4f7bb69a.jpg
digitalObjectType	21.T11148/1a1e620666cb1713acde
hasMetadata	21.T11981/73bfcca4-9f2b-4cfc-a003-30f5a51aab84

FAIR DO Creation – Updates

Finished:

Test for creating a **Coscine PID Record** with fixed types and values.

Used services : ePIC, Coscine, KIP, FAIR-DOscope

Ongoing:

Planning KIP for Coscine.

FAIR DO – Conclusions

FAIR DO can change the way we approach datasets, offering a clear and structured method of accessing and interpreting data using Persistent Identifiers (PIDs), ultimately enhancing the usability and interoperability of data by machines.

Q & A