

Linking Research Data

NFDI-MatWerk Conference

Siegburg, 27.-29.06.23

Workshop Team:

Ajay Kirar

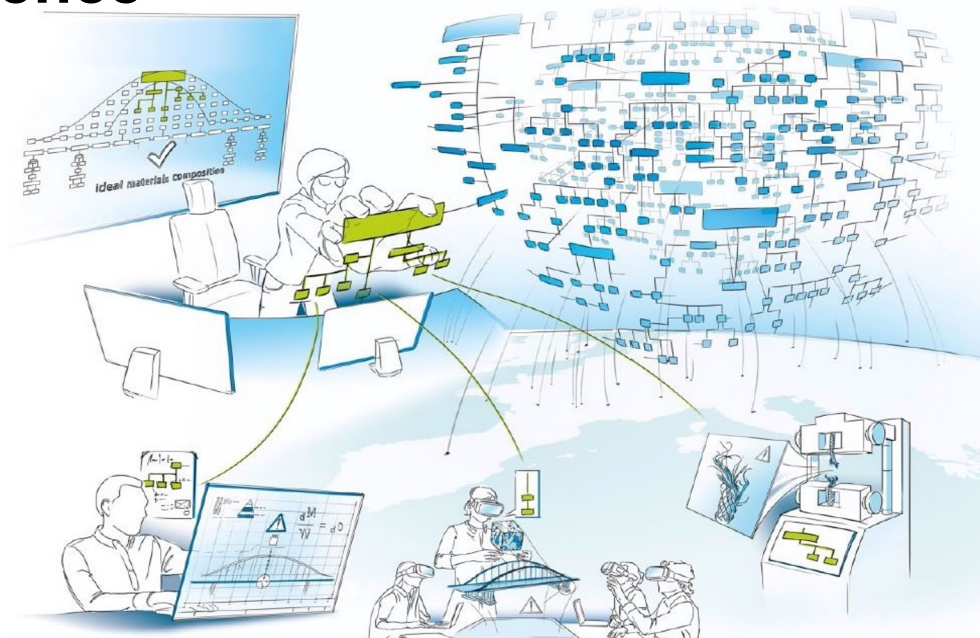
Benedikt Heinrichs

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

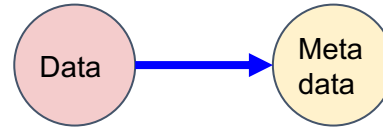
Aim of the workshop

- 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
-



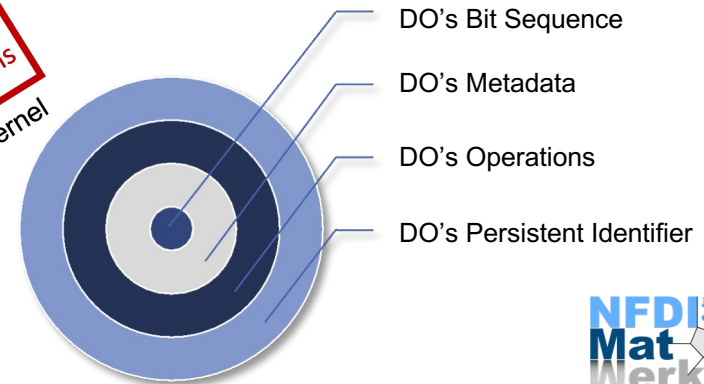
Metadata object

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

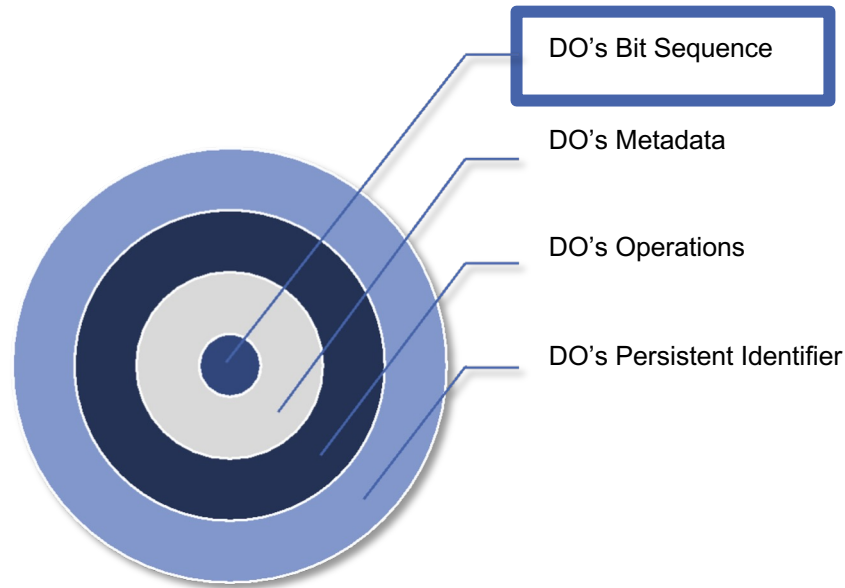
FAIR Digital Object:

- 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



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	$72 \times 10^{-6} \text{ h}^{-1}$	$5 \times 10^{-6} \text{ h}^{-1}$
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](https://zenodo.org)

```
{  
  "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

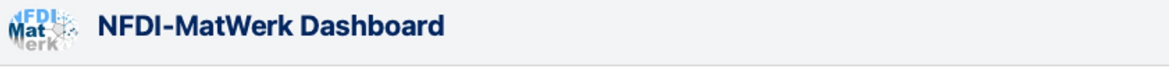
Advantages of NFDI-MatWerk Repositories

- 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.data
manager.kit.edu](https://matwerk.data.manager.kit.edu)

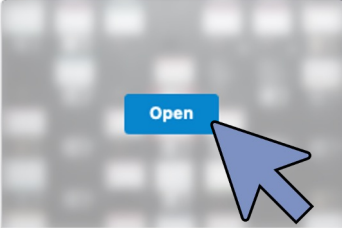


NFDI-MatWerk Dashboard


This collection of web front-ends provides easy access to the RESTful services of NFDI-MatWerk instances provided by KIT / SCC.
A documentation for the instances can be found [here](#)




NFDI-MatWerk Metadata Repository
Manage and access your metadata schemas and documents on the NFDI-MatWerk instance.



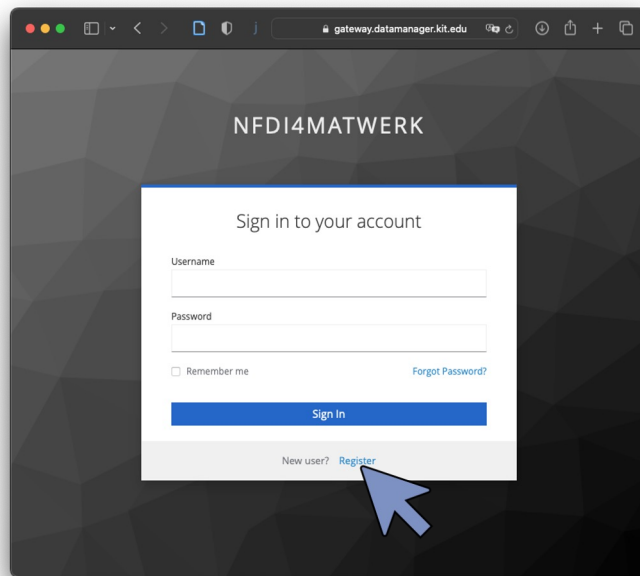
NFDI-MatWerk Data Repository
Manage and access your research data described by DataCite metadata on the NFDI-MatWerk instance.



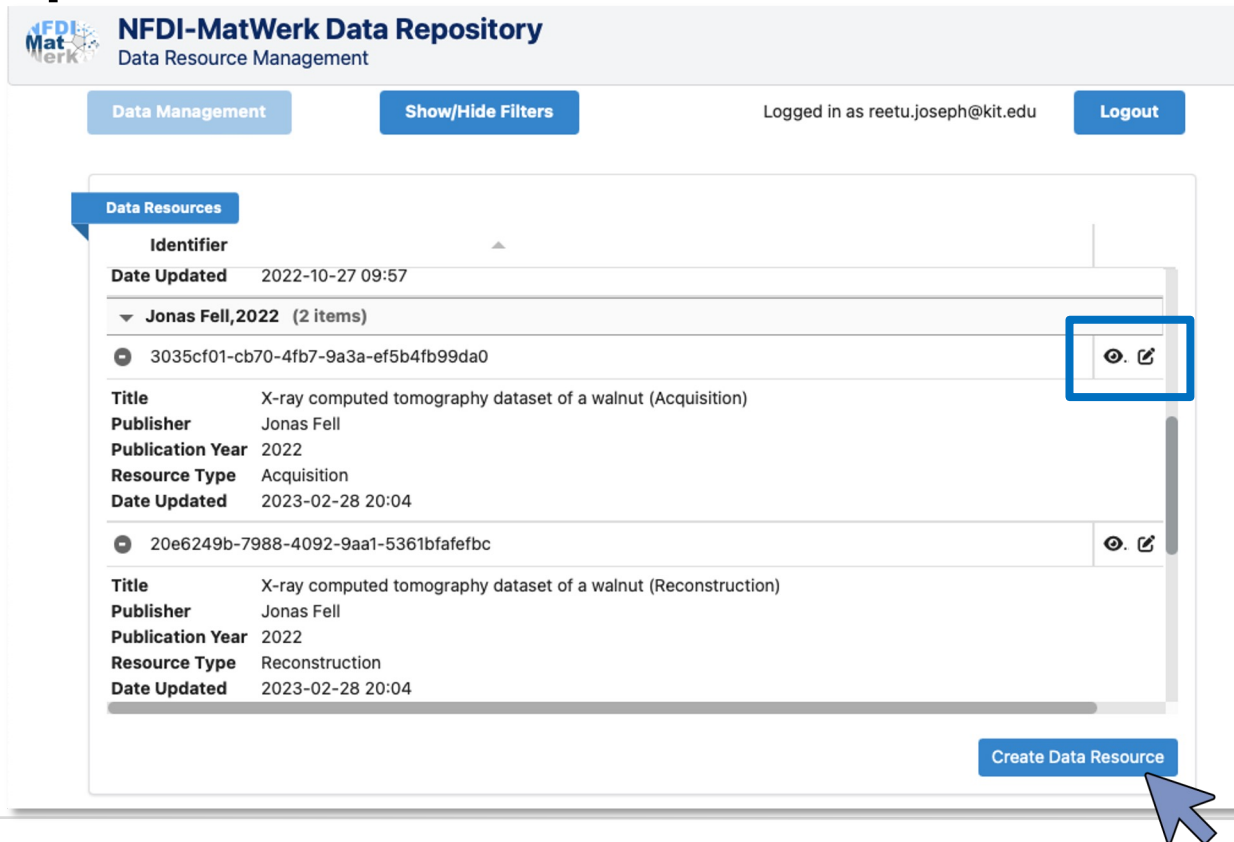
NFDI-MatWerk Mapping Service
View Mapping Service interface for automatic extraction and mapping of metadata.

Legals

Create User Credentials



Upload a File on the NFDI-MatWerk Data Repository



The screenshot shows the NFDI-MatWerk Data Repository interface. At the top, there is a header with the logo and the text "NFDI-MatWerk Data Repository Data Resource Management". Below the header, there are navigation buttons: "Data Management", "Show/Hide Filters", and "Logout". The user is logged in as "reetu.joseph@kit.edu". The main content area displays a list of "Data Resources". The first resource is identified by the identifier "3035cf01-cb70-4fb7-9a3a-ef5b4fb99da0" and is titled "X-ray computed tomography dataset of a walnut (Acquisition)". It was published by "Jonas Fell" in 2022 and was updated on "2023-02-28 20:04". A blue box highlights the edit and delete icons for this resource. The second resource is identified by "20e6249b-7988-4092-9aa1-5361bfafefbc" and is titled "X-ray computed tomography dataset of a walnut (Reconstruction)". It was also published by "Jonas Fell" in 2022 and updated on "2023-02-28 20:04". At the bottom right of the interface, there is a blue button labeled "Create Data Resource" with a mouse cursor pointing to it.

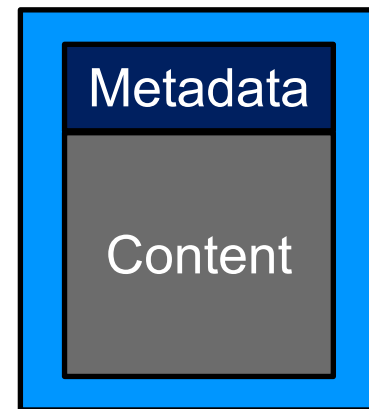
NFDI-MatWerk Data Repository
Data Resource Management

Data Management Show/Hide Filters Logged in as reetu.joseph@kit.edu Logout

Data Resources

Identifier	Date Updated
Jonas Fell, 2022 (2 items)	2022-10-27 09:57
3035cf01-cb70-4fb7-9a3a-ef5b4fb99da0	
Title	X-ray computed tomography dataset of a walnut (Acquisition)
Publisher	Jonas Fell
Publication Year	2022
Resource Type	Acquisition
Date Updated	2023-02-28 20:04
20e6249b-7988-4092-9aa1-5361bfafefbc	
Title	X-ray computed tomography dataset of a walnut (Reconstruction)
Publisher	Jonas Fell
Publication Year	2022
Resource Type	Reconstruction
Date Updated	2023-02-28 20:04

Create Data Resource



Data Resource

Enter Metadata for the Data Resource

Create Metadata ✕

Data Resource Metadata Content Information

Titles
One or more titles describing the resource.

Language
The title language.
English

Type
The type of the title.

Value *
The title value.
BCR-425_cert.pdf from BAM reference data set

Publisher
The publisher of the resource. If no publisher is set, base-repo assigns the publisher based on the caller, i.e., the logged in user.
Jane Doe

Publication Year
The year when the resource was published. If no publicationYear is set, base-repo assigned the current year automatically.
yyyy

Language
The language of the resource, if applicable.
German

General Resource Type *
The general type of the resource.
OTHER

Specific Resource Type *
A more specific type of the resource.
certificate

Enhanced Metadata

[Access_Control_and_Restrictions](#) [Creators_and_Contributors](#) [Identifiers](#) [Descriptive](#) [Geo Information](#)

[Funding Information](#)

Creators
One or more persons involved in the creation of the resource. If no creator is added, base-repo will automatically insert the caller, i.e., the logged in user.

Family Name
The creator's last name.
Skrotzki

Given Name
The creator's first name.
Birgit

Affiliations

Affiliation
The creator affiliation's name or persistent identifier.
<https://ror.org/03x516a66>

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”

Create Metadata

Embargo Date

The date until when the resource is under embargo.

License

Add one or more licenses applicable to the resource's contents.

[+](#)

Access Control List

Define additional access permissions. By default, base-repo automatically grants full permissions to the caller, i.e., the logged in user.

[-](#)

SID

The subject id, i.e., a user or group id.

Permission

The permission granted for the subject.

[+](#)

[Create Data Resource](#)

* Required field

Drop or Browse to upload the Data

Select the “Content Information” tab to upload the data

The screenshot displays the 'Update Metadata' interface. At the top, there are two tabs: 'Data Resource Metadata' and 'Content Information'. A blue mouse cursor points to the 'Content Information' tab. Below the tabs is a table with columns: Name, Media Type, Hash, Size, Uploader, and Tags. The table is currently empty. At the bottom of the interface, there is a 'Page Size' dropdown set to '10' and navigation buttons for 'First', 'Prev', '1', 'Next', and 'Last'. An 'Upload Files' dialog box is open in the foreground, showing '1 file selected' and a file named 'BCR-425_cert.pdf' with a PDF icon. A green 'Upload 1 file' button is visible at the bottom of the dialog, with a blue mouse cursor pointing to it. To the right of the dialog, there is an 'Assign Tags to Selection' section with three tags: 'tag3' (green), 'tag2' (blue), and 'tag' (red).

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

Persistent Identifier of the Resource

Show Metadata ×

Data Resource Metadata Content Information

id
The internal resource identifier.
cd218ccf-4fa2-4629-9860-bb2b5a7b842a

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


Titles

Language

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

Type
The type of the title.
[Empty field]

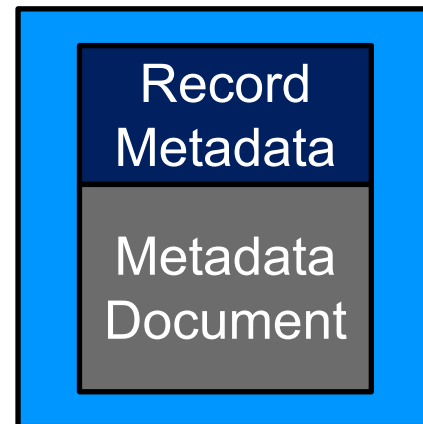
Value
The title value.
BCR-425_cert.pdf from BAM reference data set



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

 **NFDI-MatWerk Metadata Repository**
Schema and Metadata Management

[Schema Management](#) [Metadata Management](#) [Show/Hide Filters](#) Logged in as reetu.joseph@kit.edu [Logout](#)

Metadata Documents

Identifier	
 b0c08dfb-d559-4bfc-914b-862c8ec9f798	 
Related Resource 10.5281/zenodo.7764161	
Schema Identifier pp18_file (version=1)	
Date Updated 2023-06-06 10:35	
▼ test (version=1) (1 item)	
 27e8b01f-fec5-44bb-b1a1-438fc2197a93	 
Related Resource pp13-walnut-nested-main	
Schema Identifier test (version=1)	
Date Updated 2022-12-16 13:40	
▼ PP07-draft (version=5) (3 items)	
 3a5e8add-f4e4-4acb-8f53-5b6785f3261f	 
Related Resource https://matwerk.datamanager.kit.edu/api/v1/dataresources/f29a69b2-be81-494a-b240-9caba4c2633c	
Schema Identifier PP07_draft (version=5)	

[Register new Metadata Document](#)

Register the Metadata Document

Create Metadata ×

Record Metadata Metadata Document

Related Resource *
The identifier of the related resource, i.e., a URL or any internally handled kind of identifier.

Related Resource Identifier Type *
The type of the identifier, i.e., URL or INTERNAL.

Schema Identifier
The identifier of the validation schema for the provided metadata document.

Schema Version
The version of the metadata schema to be used. If not provided, the most recent version will be used.

ACL

Metadata Document File
A local file containing the metadata document.
 BCR-425_cert.json

* Required field

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

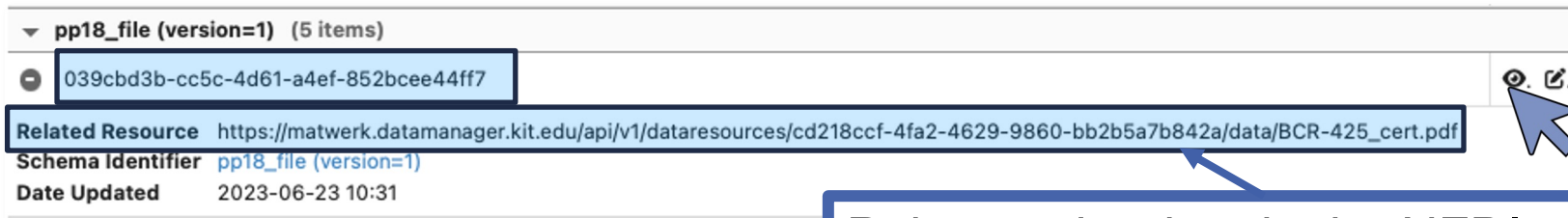
▼ pp18_file (version=1) (5 items)

● 039cbd3b-cc5c-4d61-a4ef-852bcee44ff7

Related Resource https://matwerk.datamanager.kit.edu/api/v1/dataresources/cd218ccf-4fa2-4629-9860-bb2b5a7b842a/data/BCR-425_cert.pdf

Schema Identifier pp18_file (version=1)

Date Updated 2023-06-23 10:31



Points to the data in the NFDI-MatWerk Data Repository

Metadata Document URI

Direct access URI where the associated metadata document is stored.

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

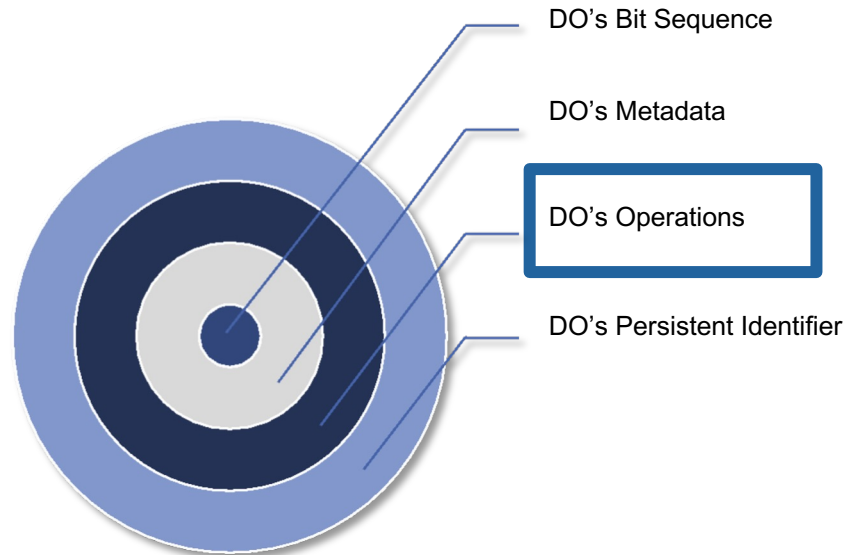
<https://matwerk.datamanager.kit.edu/api/v1/metadata/039cbd3b-cc5c-4d61-a4ef-852bcee44ff7?version=1>

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

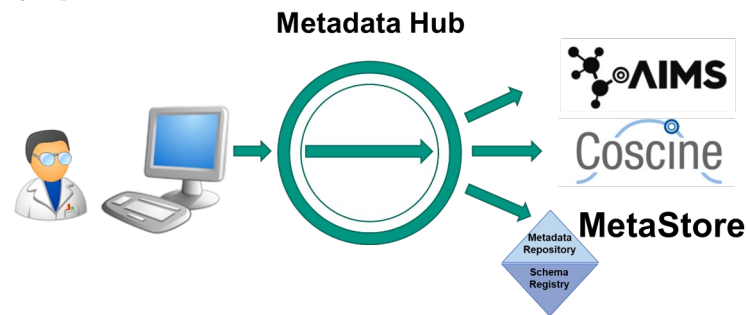
Accessing Multiple Metadata Repositories



Accessing Multiple Metadata Repositories

Metadata Hub

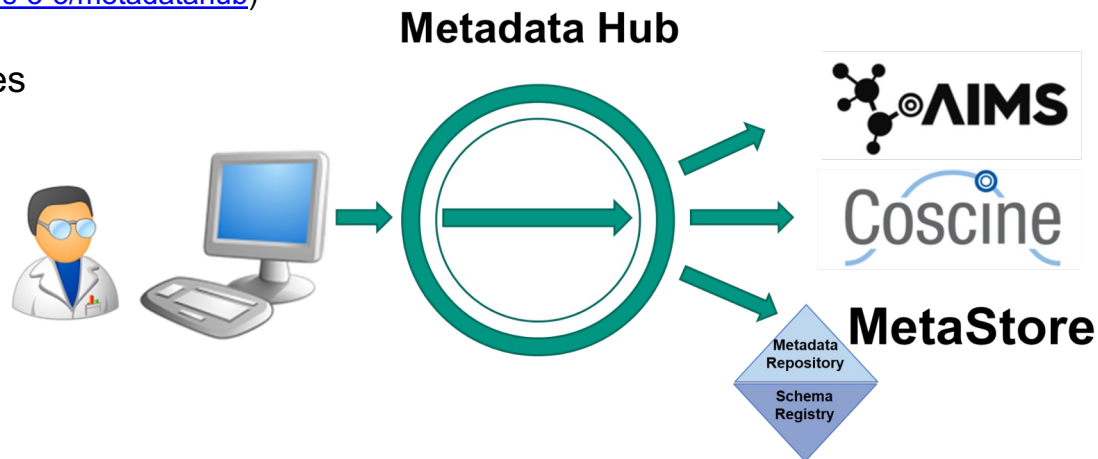
- Metadata repositories usually provide different interfaces
- Interoperability is wanted but cannot be achieved like this
- Goal:
 - Bring them together
- How:
 - A generic interface which combines different kinds of metadata repositories with one standard-based interface



Accessing Multiple Metadata Repositories

Metadata Hub

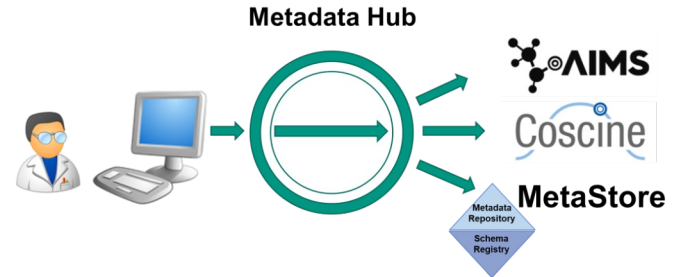
- Solution: The Metadata Hub
(<https://git.rwth-aachen.de/nfdi4ing/s-3/s-3-3/metadatahub>)
- Brings together metadata stores
 - AIMS
 - Coscine
 - MetaStore
 - Your Solution?



Accessing Multiple Metadata Repositories

Metadata Hub

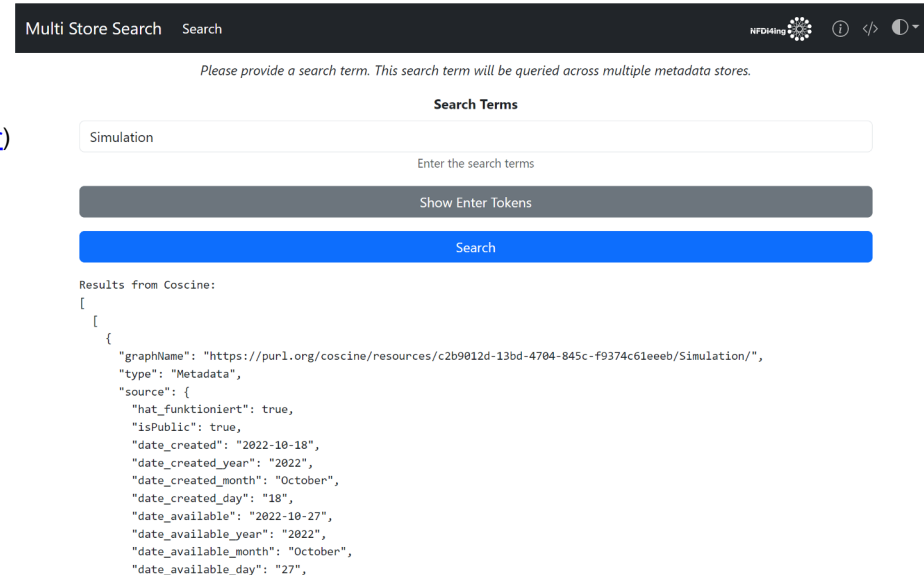
- For access → Turntable API
(<https://nfdi4ing.pages.rwth-aachen.de/s-3/s-3-3/turntable-interface/>)
- Services are defined with a provided mapping file
- Pilots:
 - Coscine
 - MetaStore
- The difference in formats between the pilots was a big challenge



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



Multi Store Search Search

Please provide a search term. This search term will be queried across multiple metadata stores.

Search Terms

Simulation

Enter the search terms

Show Enter Tokens

Search

Results from Coscine:

```
[
  [
    {
      "graphName": "https://purl.org/coscine/resources/c2b9012d-13bd-4704-845c-f9374c61eeeb/Simulation/",
      "type": "Metadata",
      "source": {
        "hat_funktioniert": true,
        "isPublic": true,
        "date_created": "2022-10-18",
        "date_created_year": "2022",
        "date_created_month": "October",
        "date_created_day": "18",
        "date_available": "2022-10-27",
        "date_available_year": "2022",
        "date_available_month": "October",
        "date_available_day": "27",
      }
    }
  ]
]
```

Accessing Multiple Metadata Repositories

Metadata Hub



- Example Use Case: Coscline (<https://coscline.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> .
```

MetadataHub Demonstrator

Select the targeted provider

Coscline

Input your User Token

eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJ0b2t2bklkajoiInM1ZTAxYm1zGUzZC00OWQ5LTkzZDItZjFkMDQ4MGM2NDk0IiwiaXNzIjoiaHR0cHM6Ly9jb3J3aWw!

Select the wanted method

Read

Select the wanted type

Schema

Input the metadata path

<https://purl.org/coscline/ap/radar/>

Send Request

Response

```
{  
  "id": "schema",  
  "value": [{"@id": "https://purl.org/coscline/ap/radar/": {"@graph":  
    [{"@id": "https://purl.org/coscline/ap/radar#subject": {"http://www.w3.org/ns/shacl#path":  
      [{"@id": "https://purl.org/dc/terms/subject": {"http://www.w3.org/ns/shacl#order":  
        [{"@value": "3": {"@type": "http://www.w3.org/2001/XMLSchema#integer": {"http://www.w3.org/ns/shacl#maxCount":
```

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?

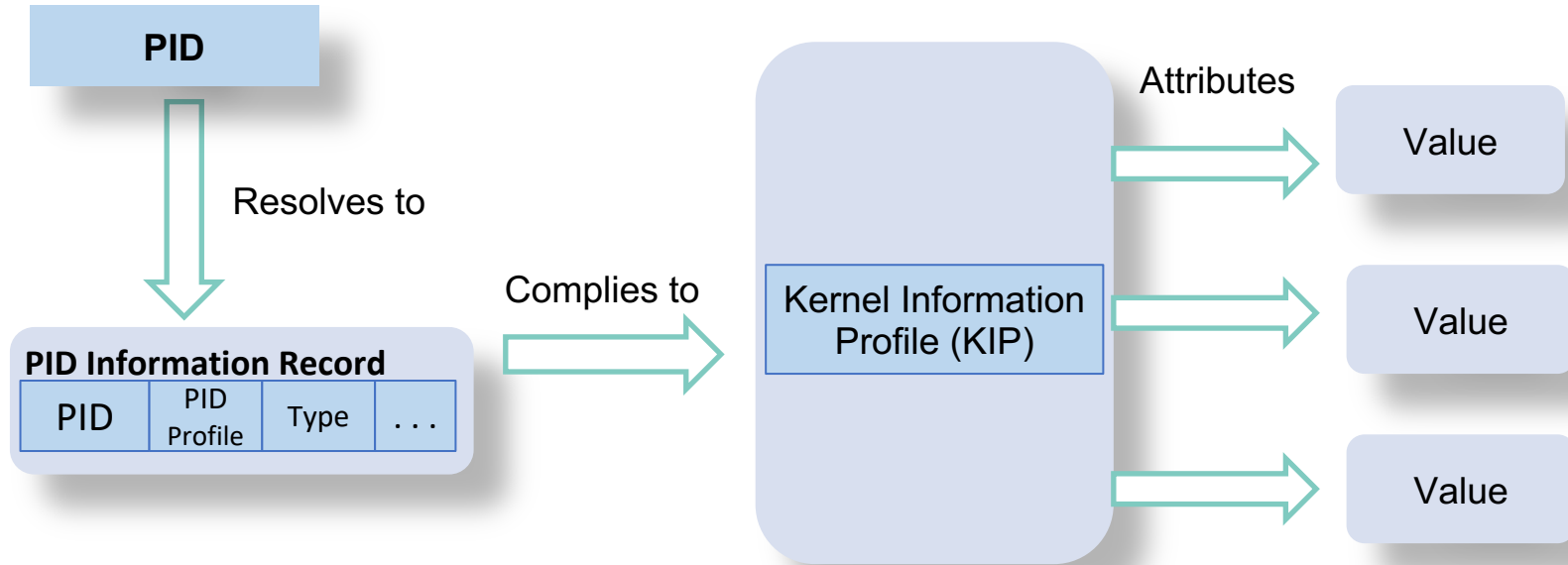
FAIR DO bind all critical information about an entity in one place and create a new kind of actionable, meaningful and technology independent object that pervades every aspect of life today :

A technical essence of a “thing” in cyberspace.

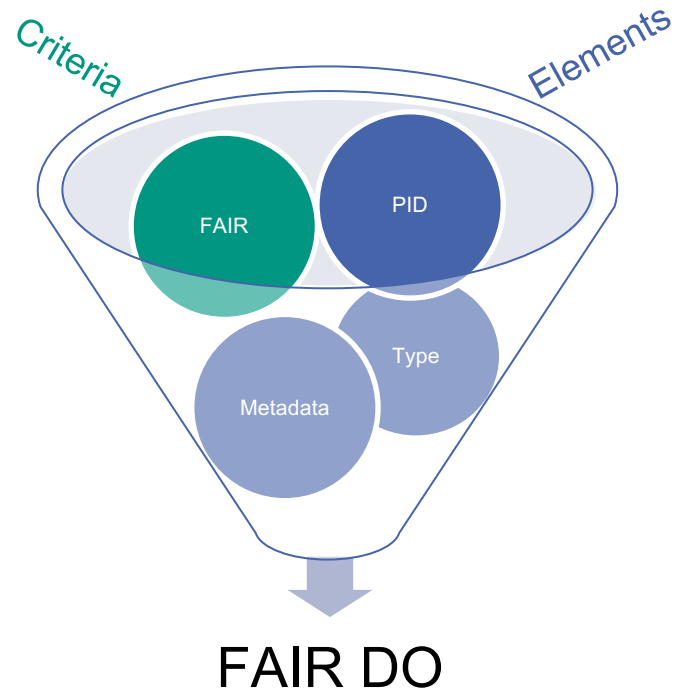


Digitally present

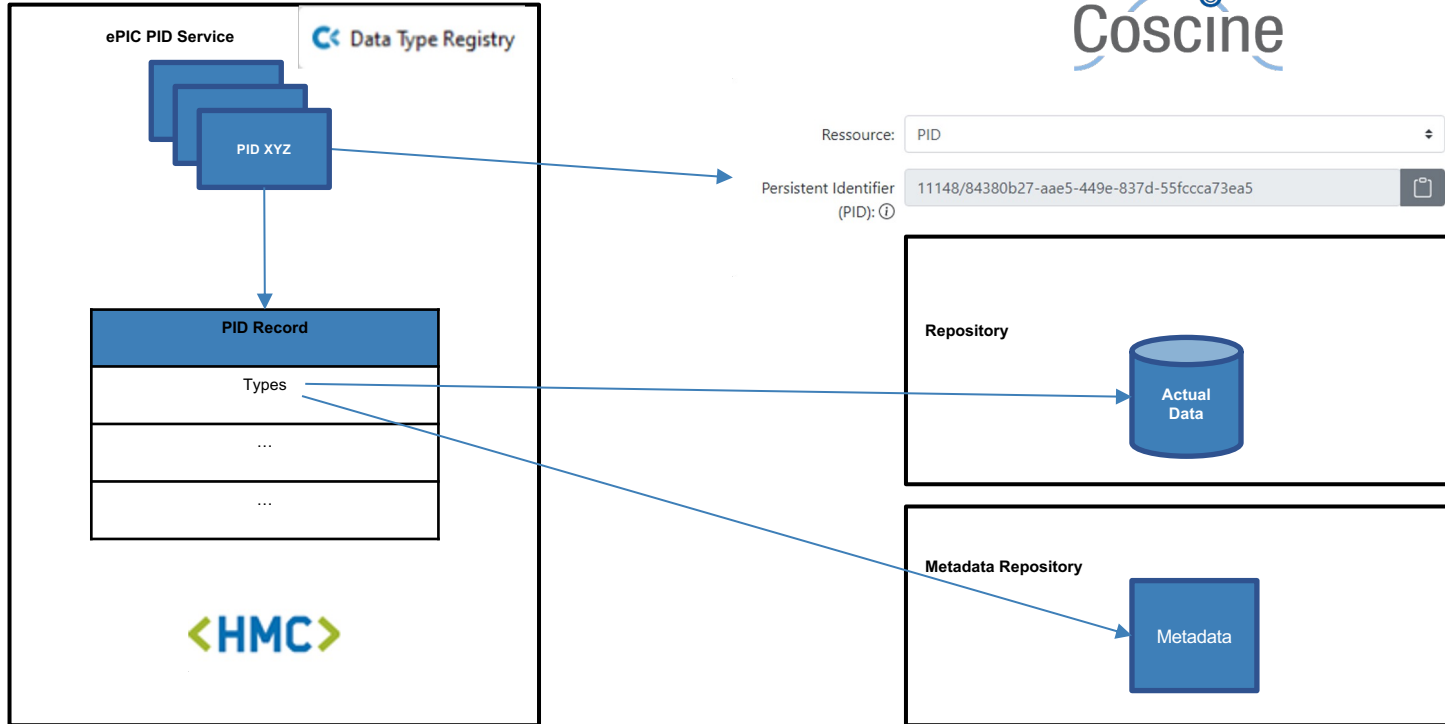
Key requirements to be a FAIR DO



What makes it a FAIR DO?

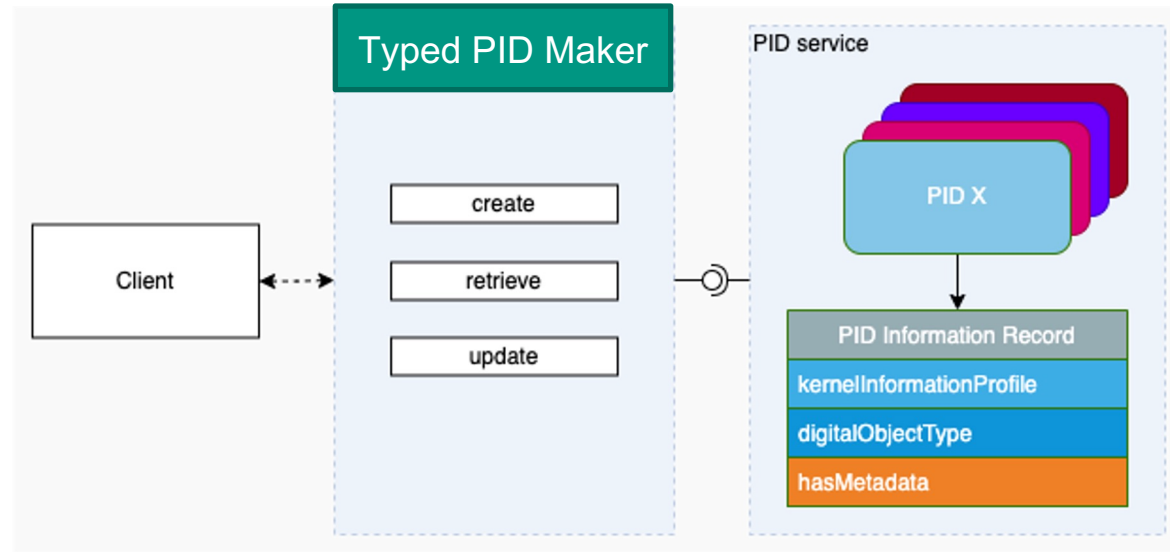


FAIR DO Creation – in action

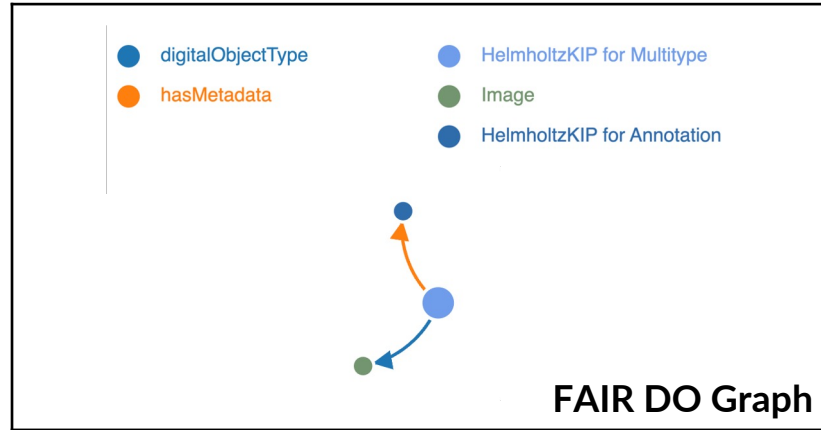


Typed PID Maker

- 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