

# Successful stories in NFFA-Europe Pilot and beyond

29.11.2023 – FAIR Data Day, Bologna (Italy)

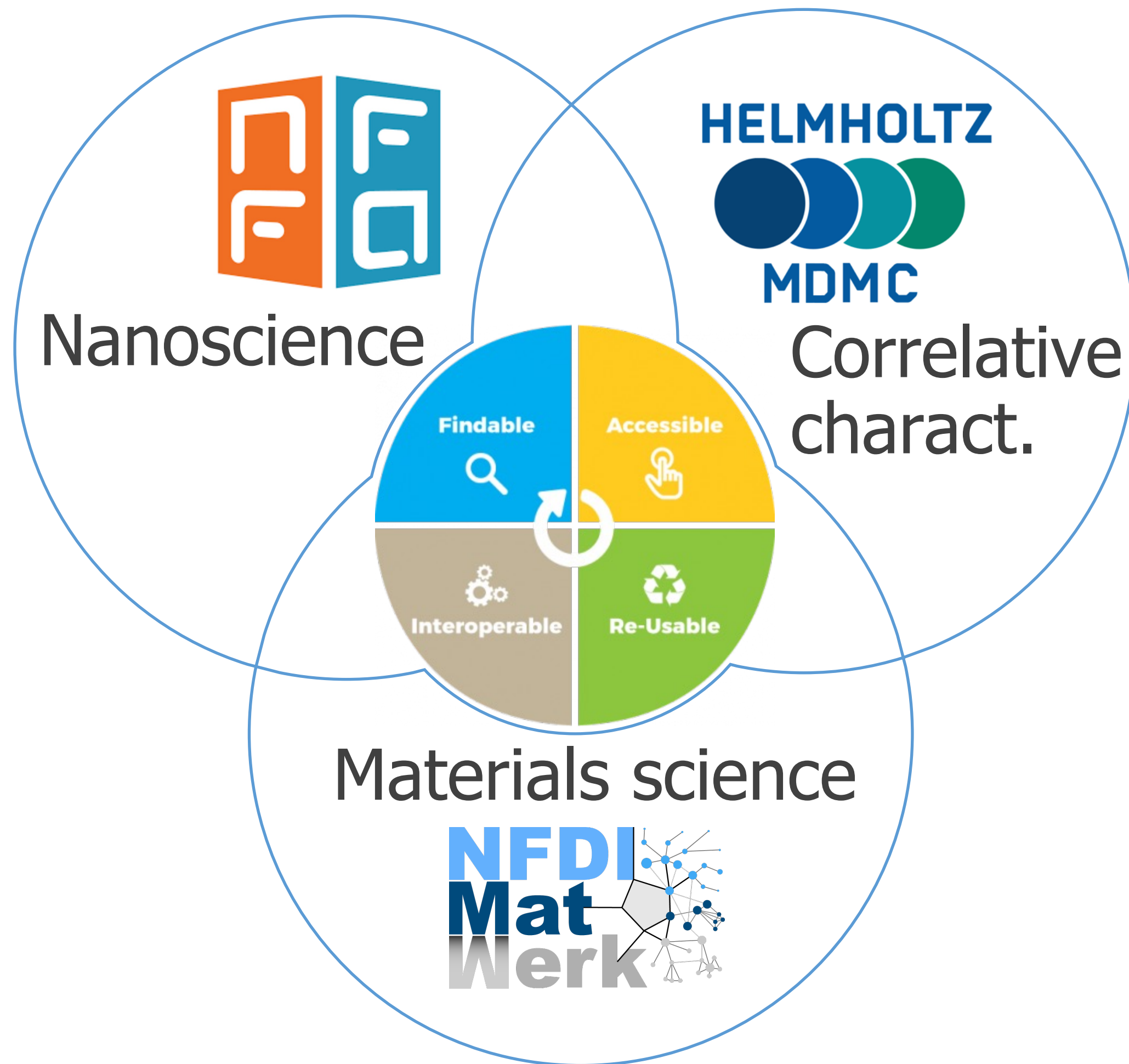
---

Dr. Rossella Aversa

Karlsruhe Institute of Technology  
NFFA-Europe Pilot – WP9 Leader



# It's all about Data Management



## Common aim:

- Data management practices following the FAIR principles
- Tools and infrastructure solutions

## Outline:

- Ongoing activities and achievements
- Reached in collaboration
- Development guided by requirements

# Are we talking about the same thing?

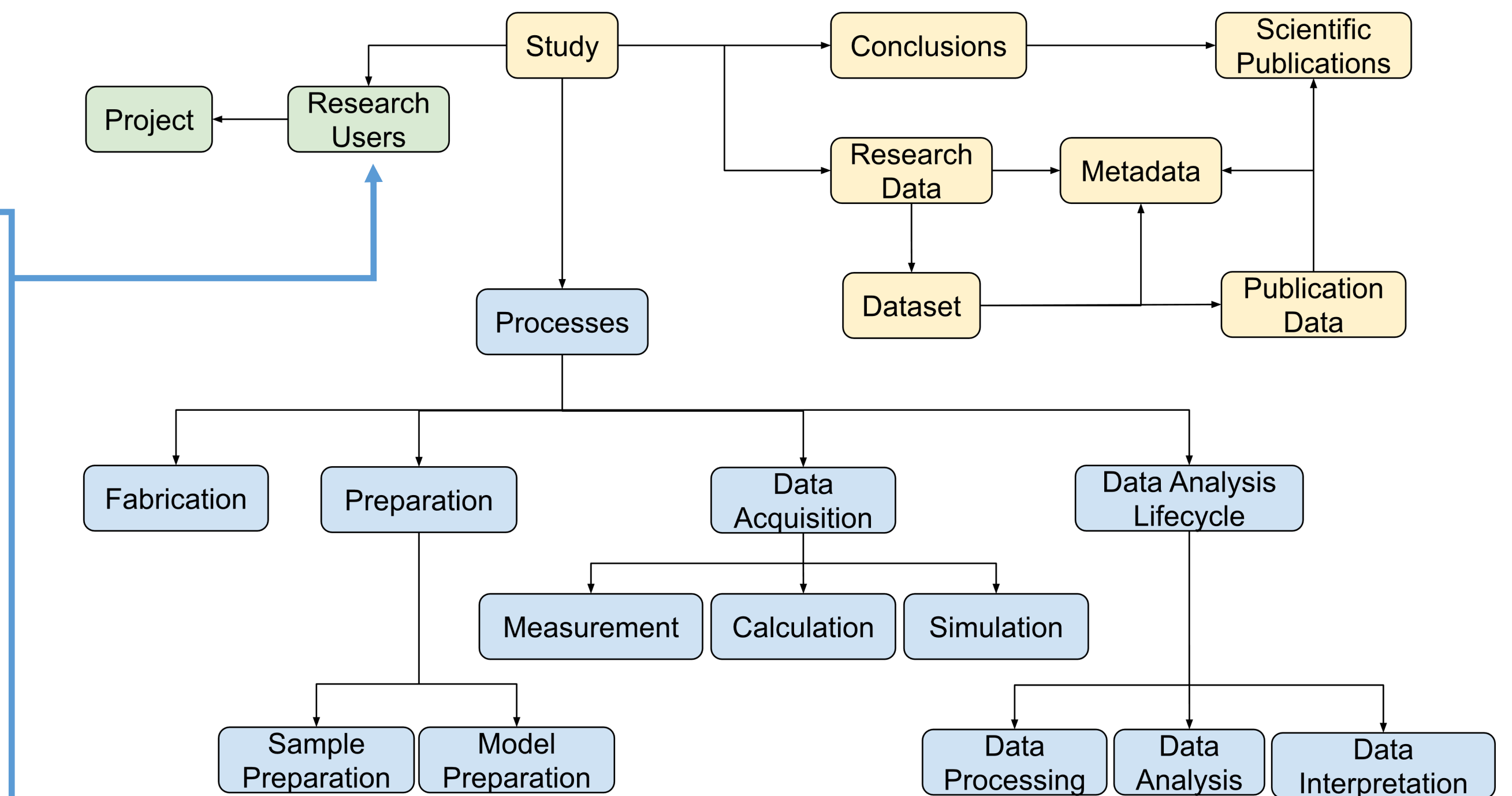
- MDMC-NEP Glossary of Terms



## Research User


Person, usually member of a **Project**, who conducts any part of the **Study**, in order to collect and/or analyse **Research Data**, or is interested in reusing **Research Data** by a third party (e.g. **Reference Data**) with the final aim to extract insights that support the answer to some specific research question (i.e., **Conclusions**).

**Research Users** may be assigned with a role (data curator, instrument scientist, team leader, team member).



<https://jl-mdmc-helmholtz.de/mdmc-activities/metadata-working-group/metadata-wg-topics/semantics/glossary/>

# Are we talking about the same thing?

- MDMC-NEP Glossary of Terms
- EVOKS Vocabulary service 
  - Persistent identifier to each term
  - Websites, ELNs, automatic processes...
  - Centrally maintained

**Research User**

User Name

User Role

- Data Curator
- Instrument Scientist
- Team Leader
- Team Member**



PREFERRED TERM

**User Role** 

NARROWER CONCEPTS

Data Curator  
 Instrument Scientist  
 Team Leader  
 Team Member

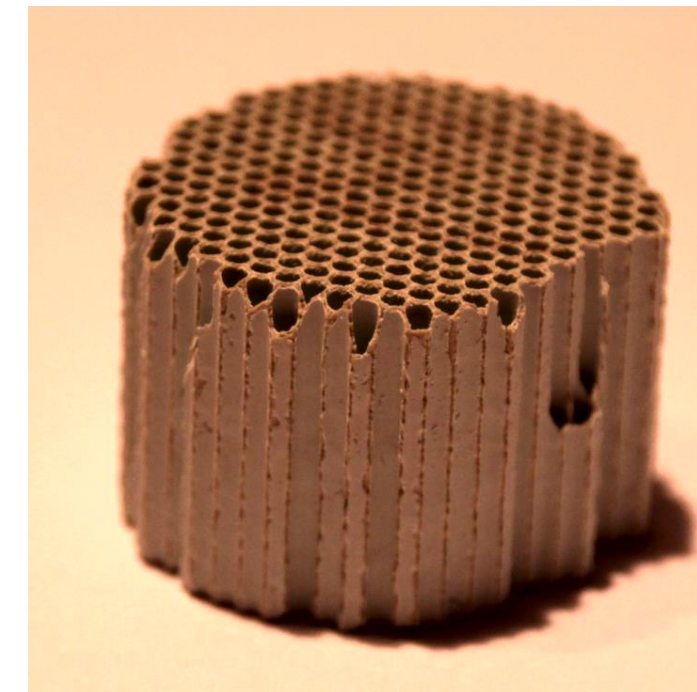
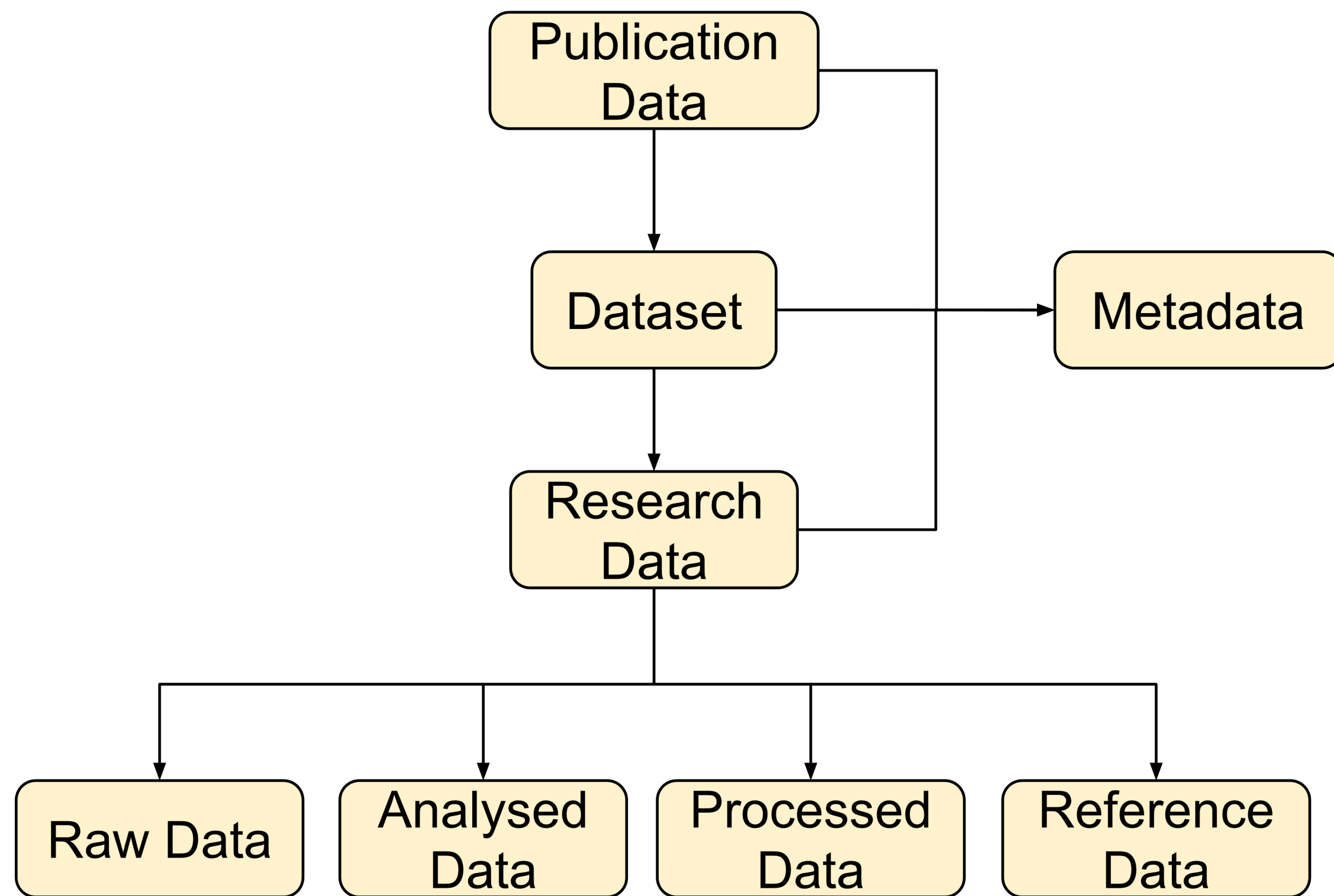
URI

<http://matwerk.datamanager.kit.edu:8001/DemoTerms-1/en/page/userrole> 

DOWNLOAD THIS CONCEPT:

[RDF/XML](#) [TURTLE](#) [JSON-LD](#)

# What can I describe?



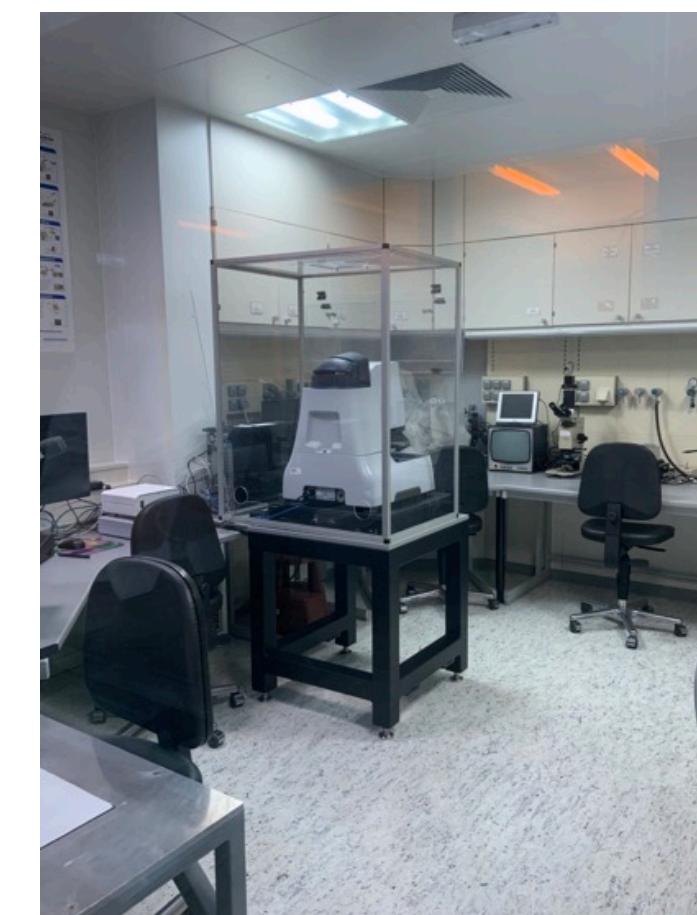
Sample, courtesy of M. Mail



Sample holder, courtesy of R. Thelen



Vertical Light Interferometer, courtesy of R. Thelen



Atomic Force Microscope, courtesy of R. Thelen



Cofocal Scanning Microscope, courtesy of R. Thelen

# How do I describe my data?

**Metadata:** descriptive information about data

**Metadata Document:** structured information about a data resource

**Metadata Schema:** outline of the overall structure of the metadata (elements, value types, rules, ...)

Example: JSON metadata schema for Sample

```

"properties": {
  "sampleName": {
    "type": "string"
  },
  "sampleComponents": {
    "type": "array",
    "items": {
      "type": "object",
      "properties": {
        "sampleComponentName": {
          "type": "string"
        },
        "sampleComponentChemicalFormula": {
          "type": "string"
        }
      }
    },
    "required": ["sampleComponentName"]
  }
},
"samplePreparation": { ...
  },

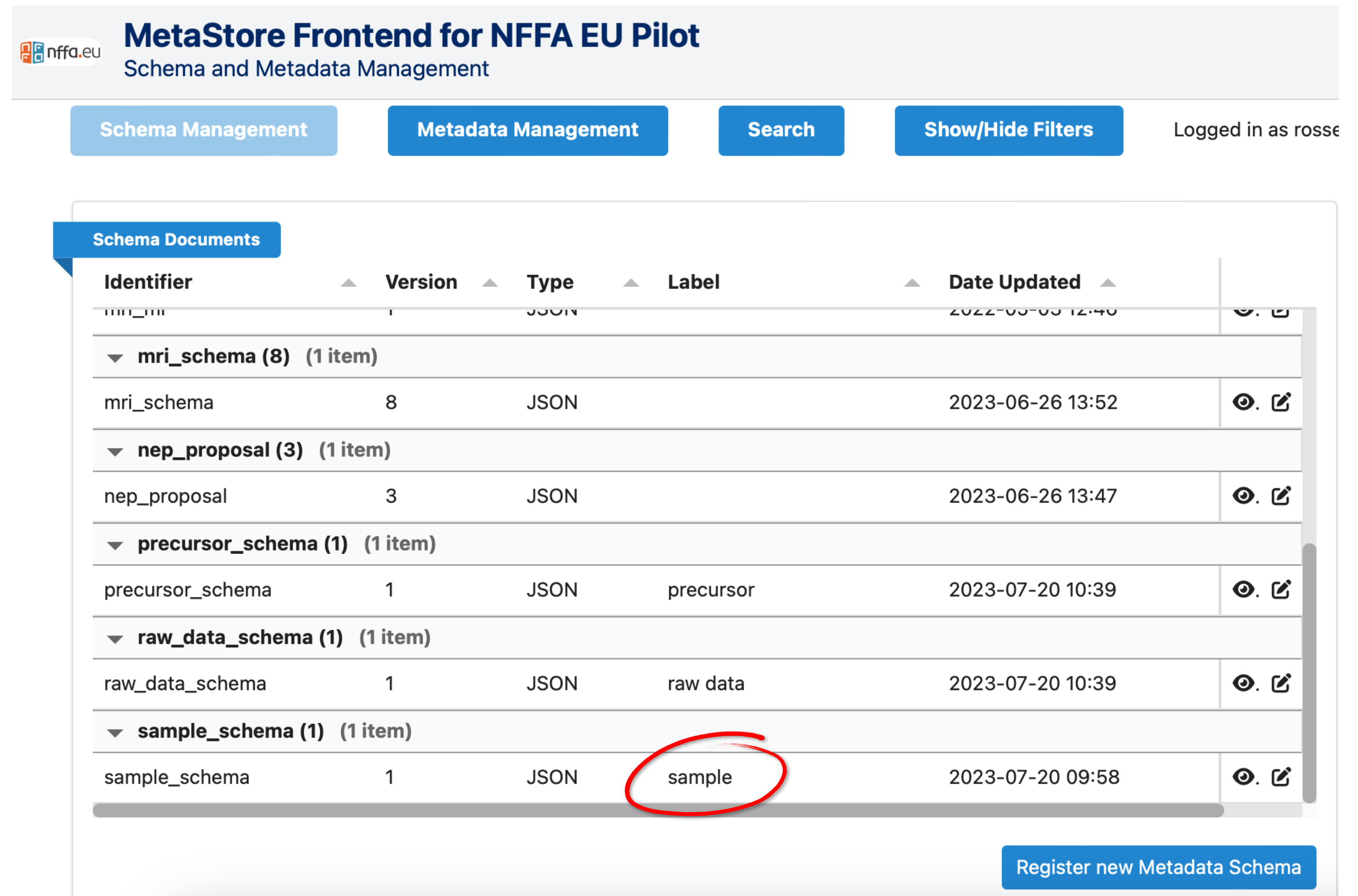
```

# Where do I find the metadata schemas?

## MetaRepo:

- Metadata schema registry
- Metadata repository
- API and GUI (customizable)
- Automatic validation
- Versioning
- Access control management
- Arbitrary number of XML, JSON schemas

<https://metarepo.nffa.eu>



**MetaStore Frontend for NFFA EU Pilot**  
Schema and Metadata Management

Schema Management Metadata Management Search Show/Hide Filters Logged in as rosse

**Schema Documents**

Identifier	Version	Type	Label	Date Updated
▼ mri_schema (8) (1 item)				
mri_schema	8	JSON		2023-06-26 13:52
▼ nep_proposal (3) (1 item)				
nep_proposal	3	JSON		2023-06-26 13:47
▼ precursor_schema (1) (1 item)				
precursor_schema	1	JSON	precursor	2023-07-20 10:39
▼ raw_data_schema (1) (1 item)				
raw_data_schema	1	JSON	raw data	2023-07-20 10:39
▼ sample_schema (1) (1 item)				
sample_schema	1	JSON	sample	2023-07-20 09:58

Register new Metadata Schema

# Which metadata schemas are available?

- proposal
- (minimal) input
- (minimal) precursor
- (minimal) sample
- (minimal) raw data
- (minimal) analysed data

Work in Progress

- SEM
- SEM/FIB tomography
- Nano CT/micro CT
- TEM
- MRI
- STM → see talk by M. Panighel

Ongoing collaboration on Materials Deposition (L. Ferrario, D. De Angelis, M. Panighel)



# How do I write my metadata?

## MetaRepo:

- Register metadata documents according to existing schemas
- Validate them

Schema Identifier

The identifier of the validation schema for the provided metadata document

- ✓ sample\_schema
- raw\_data\_schema
- precursor\_schema
- nep\_proposal
- mri\_schema
- mri\_ml
- mldata\_basic\_schema
- input\_schema
- datacite
- analysed\_data\_schema

Metadata Document File

A local file containing the metadata document.

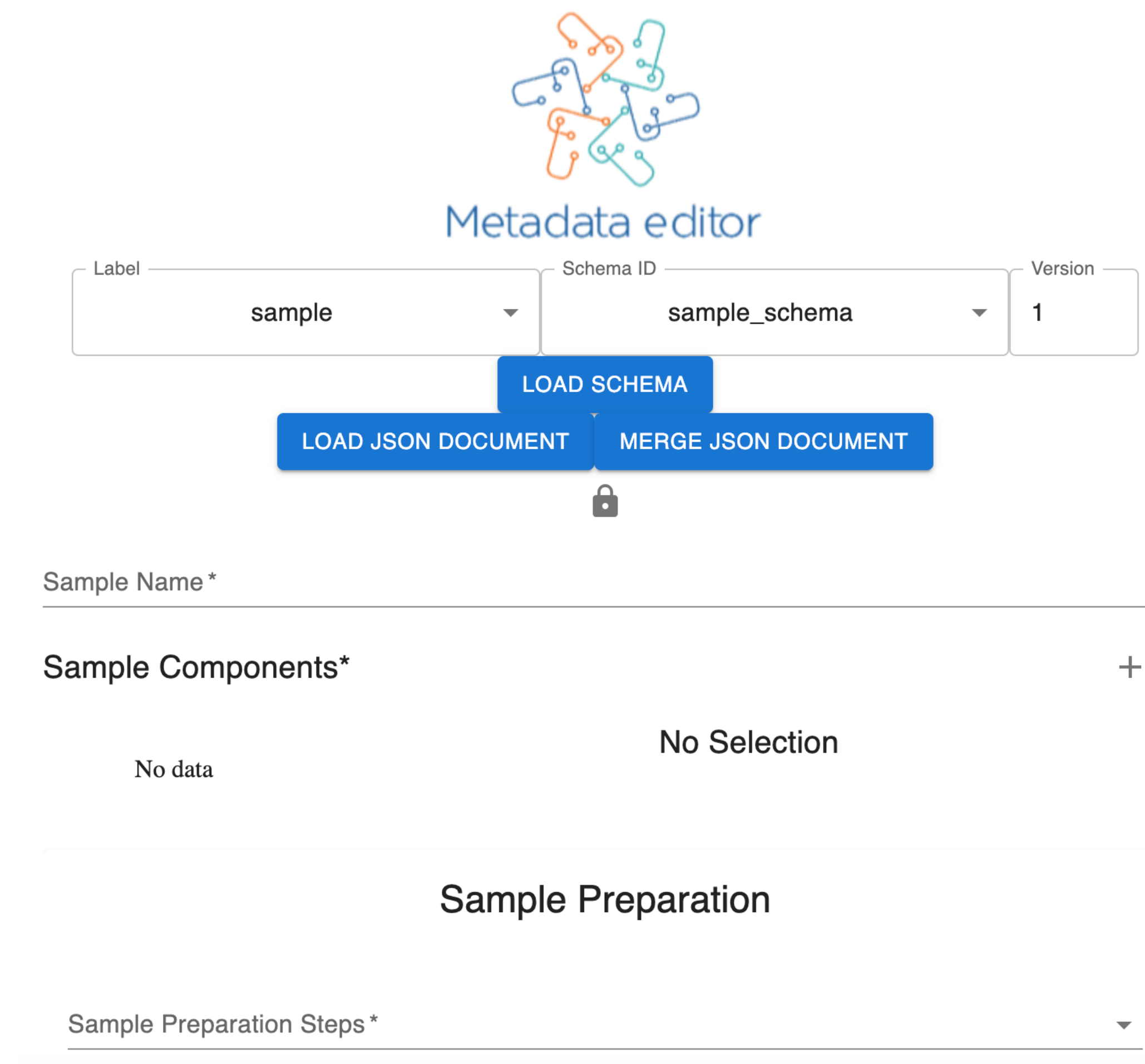
Choose File no file selected

Validate

# What if I'm not familiar with JSON?

## Metadata editor:

- Edit metadata documents according to schemas
- Register documents via the GUI
- Local application (exposes an API for instruments/ELNs)



The screenshot shows the 'Metadata editor' interface. At the top, there is a logo consisting of several interlocking loops in orange and blue. Below the logo, the title 'Metadata editor' is displayed. The interface includes three dropdown menus: 'Label' with the value 'sample', 'Schema ID' with the value 'sample\_schema', and 'Version' with the value '1'. Below these are three blue buttons: 'LOAD SCHEMA', 'LOAD JSON DOCUMENT', and 'MERGE JSON DOCUMENT'. A small lock icon is positioned below the buttons. The main content area is currently empty, showing 'Sample Name \*' with a text input field, 'Sample Components\*' with a plus sign, and 'No data' and 'No Selection' text. Below this is a section titled 'Sample Preparation' with a 'Sample Preparation Steps \*' dropdown menu.

<https://metadata-editor.gitlab.io/documentation/>

# Should I fill all metadata by hand?

## Metadata editor (local)

## Automatic Metadata Extraction and Mapping tool

- Input: image, file, ...
- Output: metadata document
- New plugins continuously under development



### NFDI-MatWerk Mapping Service

Automatic extraction and mapping of metadata.

#### Mapping Component

Choose a suitable mapping from available options

##### SEM/FIB Tomography Acquisition to TXT

Creates a summary of all metadata extracted from images in a comma delimited txt file. LU: 23.08.2023

Select

##### SEM/FIB Tomography Acquisition to JSON

Extracts metadata from a SEM/FIB Tomography Acquisition zip file and maps it to the appropriate schema to create a JSON metadata document. LU: 24.08.2023

Select

##### SEM Zip to txt

Takes a zip folder of arbitrary SEM tiff images and returns a comma separated txt file with a summary of all of the embedded metadata. LU: 05.09.2023

Select

##### SEM to JSON

This plugin is able to handle a variety of SEM images and processes them using the Hyperspy library. A resulting metadata document in .JSON format is then

Select

Drag & Drop your files or [Browse](#)

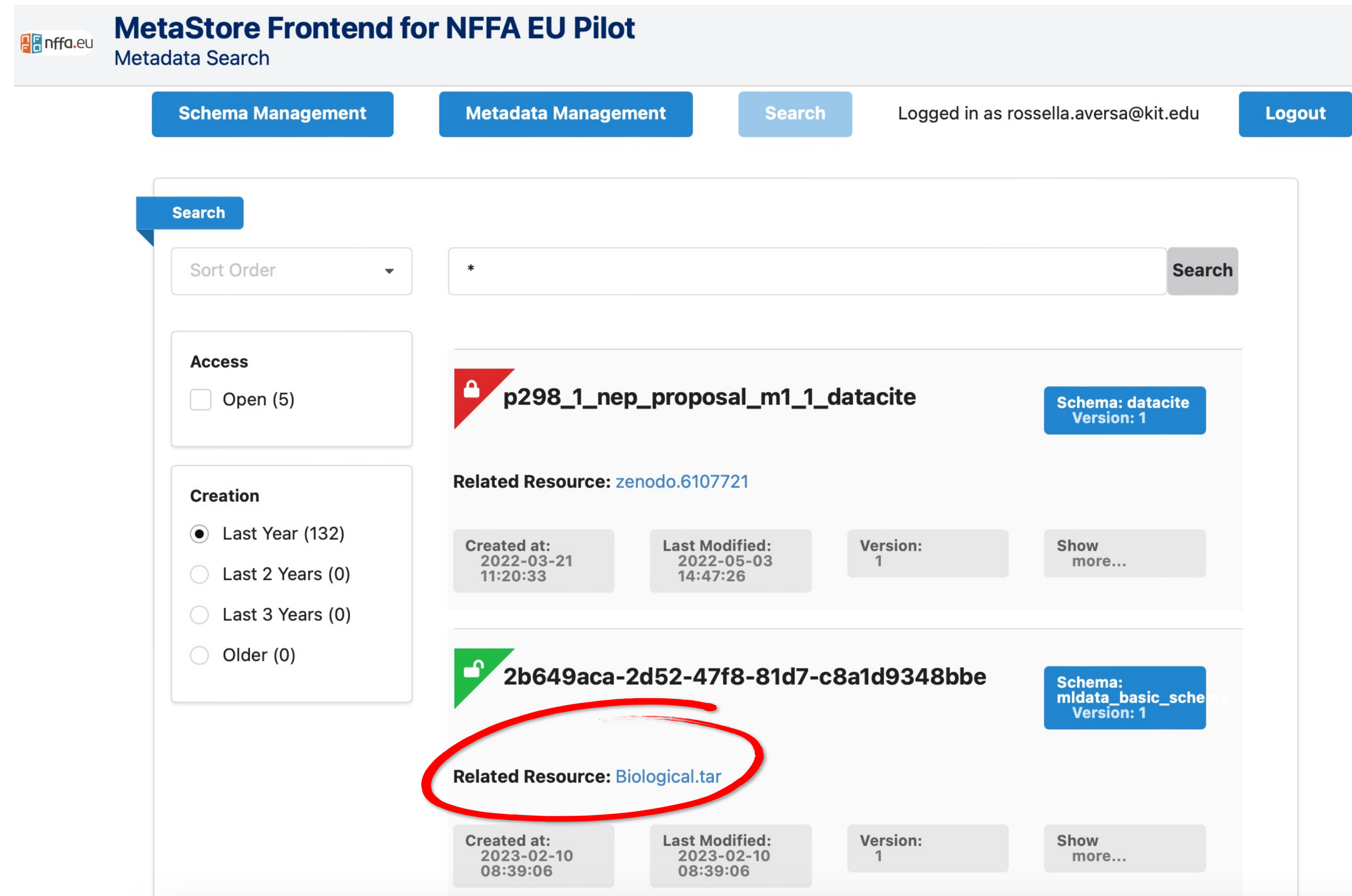
<https://matwerk.datamanager.kit.edu/mapping-service-ui.html>

Map document

# How do I find data from metadata?

## MetaRepo:

- Use content of metadata documents to search for relevant data
- What is the data about?  
Is it useful for my needs?
- Full-text search
- (basic, customizable) faceted search
- Private vs Public resources



**MetaStore Frontend for NFFA EU Pilot**  
Metadata Search

Schema Management Metadata Management Search Logged in as rossella.aversa@kit.edu Logout

Search

Sort Order \*

Access  
 Open (5)

Creation  
 Last Year (132)  
 Last 2 Years (0)  
 Last 3 Years (0)  
 Older (0)

**p298\_1\_nep\_proposal\_m1\_1\_datacite** Schema: datacite Version: 1

Related Resource: [zenodo.6107721](#)

Created at: 2022-03-21 11:20:33 Last Modified: 2022-05-03 14:47:26 Version: 1 Show more...

**2b649aca-2d52-47f8-81d7-c8a1d9348bbe** Schema: mldata\_basic\_sche Version: 1

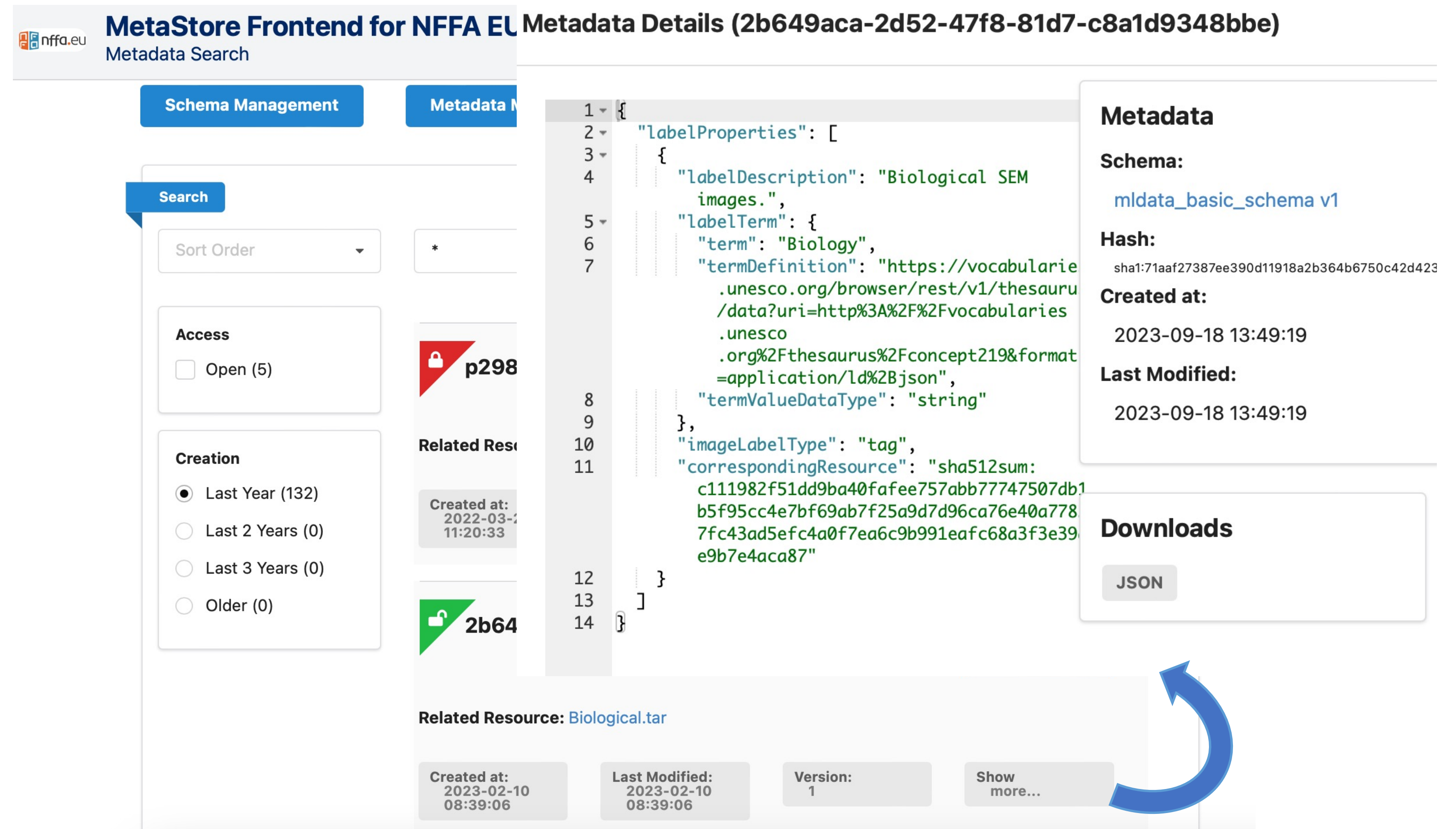
Related Resource: [Biological.tar](#)

Created at: 2023-02-10 08:39:06 Last Modified: 2023-02-10 08:39:06 Version: 1 Show more...

# How do I find data from metadata?

## MetaRepo:

- Use content of metadata documents to search for relevant data
- What is the data about?  
Is it useful for my needs?
- Full-text search
- (basic, customizable) faceted search
- Private vs Public resources



**MetaStore Frontend for NFFA EU Metadata Search**

**Metadata Details (2b649aca-2d52-47f8-81d7-c8a1d9348bbe)**

**Schema Management** | **Metadata M**

**Search**

Sort Order: [v] \* [v]

**Access**

Open (5)

**Creation**

Last Year (132)

Last 2 Years (0)

Last 3 Years (0)

Older (0)

**Related Resources**

**2b64**

**Created at:** 2023-02-10 08:39:06

**Last Modified:** 2023-02-10 08:39:06

**Version:** 1

**Show more...**

```

1 {
2   "labelProperties": [
3     {
4       "labelDescription": "Biological SEM
5         images.",
6       "labelTerm": {
7         "term": "Biology",
8         "termDefinition": "https://vocabularie
9           .unesco.org/browser/rest/v1/thesauru
10            /data?uri=http%3A%2F%2Fvocabularies
11              .unesco
12                .org%2Fthesaurus%2Fconcept219&format
13                  =application/ld%2Bjson",
14                "termValueDataType": "string"
15              },
16            "imageLabelType": "tag",
17            "correspondingResource": "sha512sum:
18              c111982f51dd9ba40fafee757abb77747507db1
19                b5f95cc4e7bf69ab7f25a9d7d96ca76e40a778
20                  7fc43ad5efc4a0f7ea6c9b991eafc68a3f3e39
21                    e9b7e4aca87"
22            }
23          ]
24        }
25      ]
26    }
  
```

**Metadata**

**Schema:** [mldata\\_basic\\_schema v1](#)

**Hash:** sha1:71aaf27387ee390d11918a2b364b6750c42d423

**Created at:** 2023-09-18 13:49:19

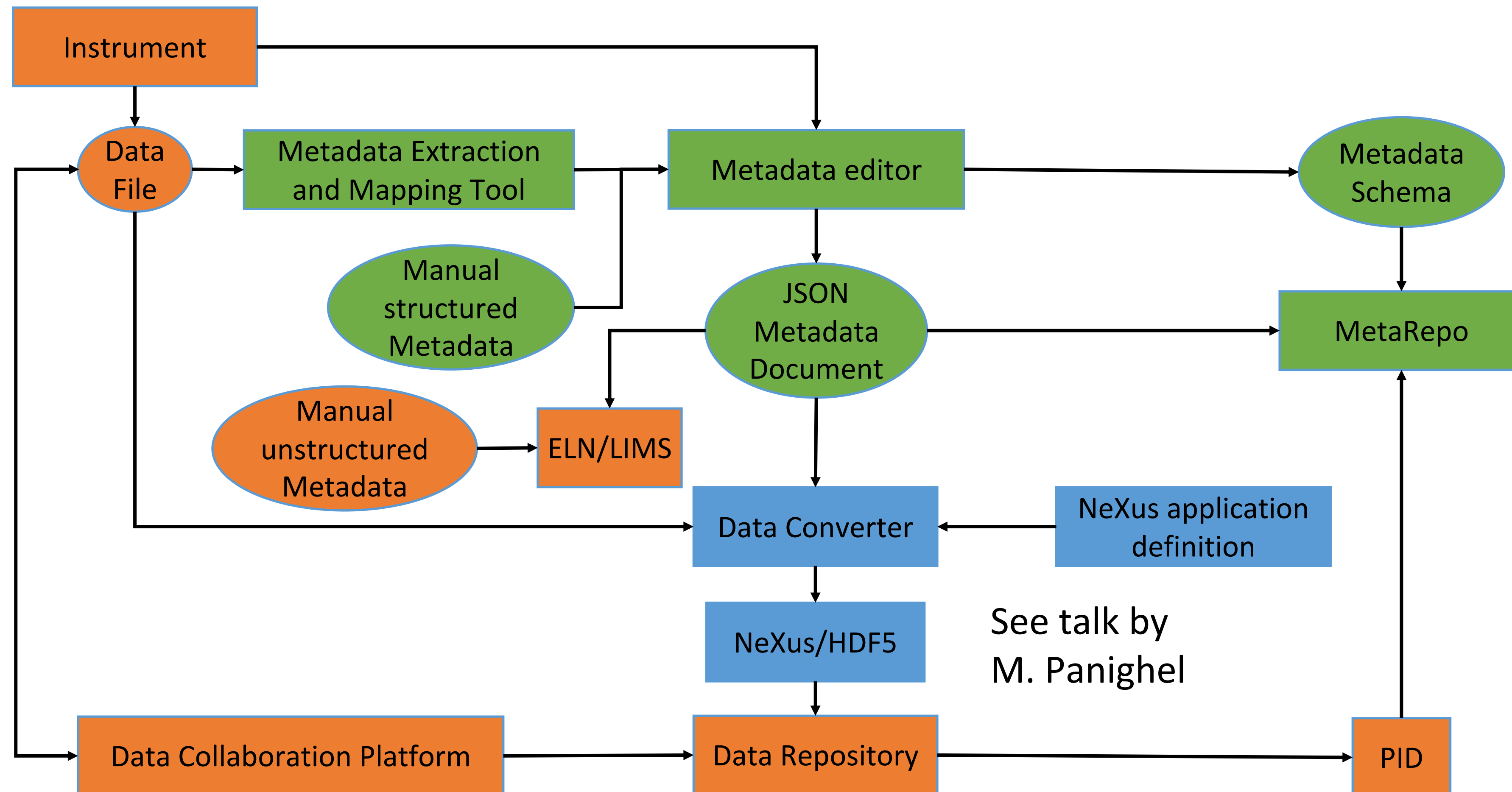
**Last Modified:** 2023-09-18 13:49:19

**Downloads**

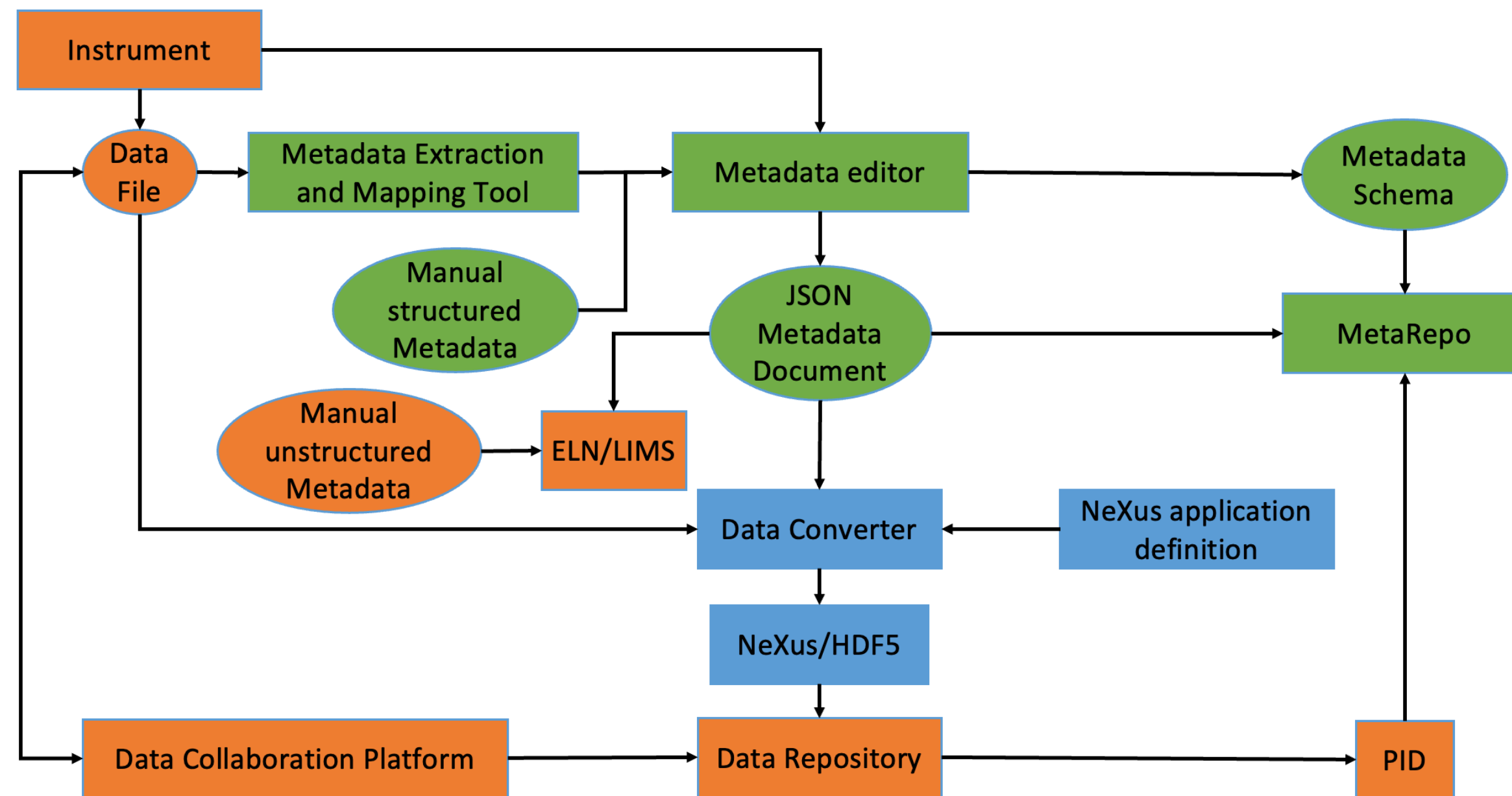
[JSON](#)

**Related Resource: Biological.tar**

# Let's put it all together



# Closing Remarks



- Tools and services to facilitate (meta)data management
- Framed in a (meta)data management workflow
- Targeted to be useful to researchers
- Applied in daily research life



nffa.eu

**PILOT** 2021  
2026

# Contacts

ROSSELLA AVERSA

[rossella.aversa@kit.edu](mailto:rossella.aversa@kit.edu)

Acknowledgements to: G. Abdildina, N. Blumenröhr, F. Ernst, V. Hartmann, M. Inkmann, T. Jejkal, A. Kirar, E. Vitali, JA6 members, JL-MDMC Metadata WG members, and all the colleagues who contributed with tests and consultancy.

Funded by: the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number 460247524; the Joint Laboratory Model and Data-driven Materials Characterization (JL MDMC), a cross-centre platform of the Helmholtz Association; NFFA-Europe-Pilot (EU H2020 - n. 101007417); the research program 'Engineering Digital Futures' of the Helmholtz Association of German Research Centers, the Helmholtz Metadata Collaboration Platform.

