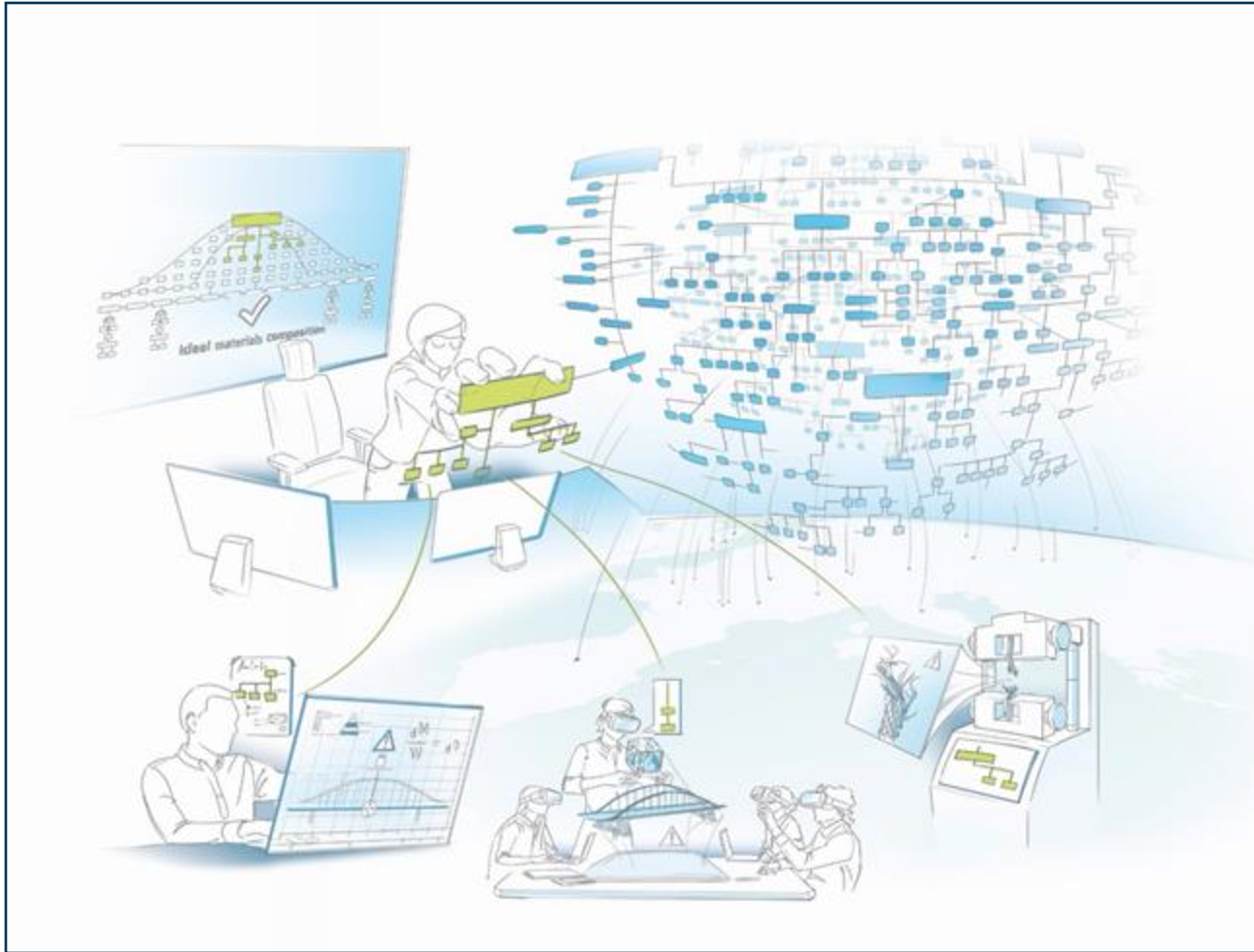
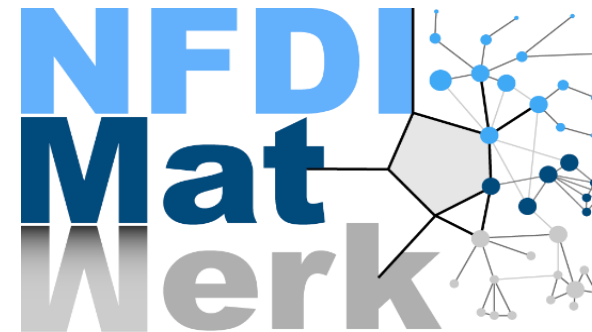


# Metadata Management: Key Essentials



Sabrine Chelbi  
Scientific Computing Center (SCC)  
Karlsruhe Institute of Technology (KIT)

NATIONAL RESEARCH DATA  
INFRASTRUCTURE FOR  
MATERIALS SCIENCE &  
ENGINEERING



4

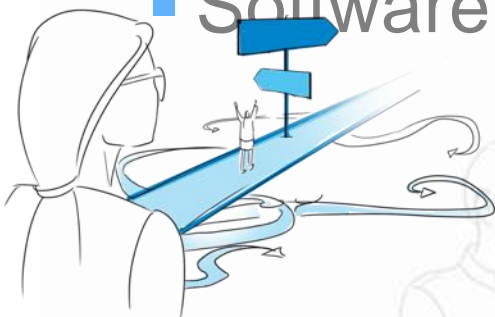
- Understand the importance and added value of metadata.
- Get an overview of JSON Schema.
- Learn how to create, edit, and manage metadata.
- Search for data from existing metadata.

## First part (Theory Part):

- Motivation
- Recap of FAIR principles
- Basics of metadata
- Basics of JSON Schema and metadata documents
- Use case Introduction
- Software Installation

## Second part (Practical Part):

- Metadata management in practice
  - Mapping Service
  - Metadata Editor
  - NFDI-MatWerk Metadata Repository
- Survey



You might want to:

- Compare your results with similar ones in the literature
- Reproduce/reuse results available in the literature
- Take delivery of the project handed over by a student/colleague who left
- Exchange data with your colleagues to collaborate on a research project
- Allow others to reproduce/reuse your results to be cited



# Data-related research questions

- Which data do support the results in this paper?
- How can I reproduce the data?
- How were these measures achieved/performed?
- How can I search for data with specific features?
- How can I publish my data in such a way that others can reuse it and cite it?



# The FAIR Guiding Principles



Findable



Accessible



Interoperable



Reusable

<https://www.go-fair.org/fair-principles/>

# The FAIR Guiding Principles



## Findable

*(Meta)data should be easy to find  
for both humans and computers*



PIDs

- F1: (Meta)data are assigned globally unique and persistent identifiers (**PIDs**)
- F3: Metadata clearly and explicitly include the identifier of the data they describe

# The FAIR Guiding Principles



## Accessible

*It should be known how  
(meta)data can be accessed*



Metadata  
repositories

- A2: **Metadata should be accessible** even when the data is no longer available
- A1.2: The protocol allows for an **authentication and authorization (AAI)** procedure where necessary.



# The FAIR Guiding Principles



Interoperable

*Data should be exchanged and interpreted by humans and computers*



Structured  
metadata

- I1: (Meta)data use a formal, accessible, shared, and broadly applicable **language** for knowledge representation

# The FAIR Guiding Principles



Reusable

*It should be clear how data can be reused and/or replicated*



Standards,  
Licence

- R1: Metadata should richly describe the data with a plurality of **accurate and relevant attributes**
- R1.1: (Meta)data are released with a clear and accessible data **usage licence**

# Basics of metadata: concepts

Data vs Metadata

Structured  
Metadata

Metadata  
Standards

Metadata  
Repositories

PIDs

Licences

# Data vs Metadata

Metadata: data describing data.

I told you,  
metadata is data!      It's so much more...



<https://imgflip.com/i/92po3j>



*“**Data** is stuff. It is raw, unprocessed, possibly even untouched by human hands, unviewed by human eyes, un-thought-about by human minds”.*

J. Pomerantz (2015). *Metadata*. The MIT Press.

*“**Metadata** is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage an information resource”.*

National Information Standards Organization (2004) from “Big Data, Little Data, No Data”, C. L. Borgman (2015)

# Data vs Metadata

	Data	Metadata
<b>Nature and Content</b>	<b>Raw facts</b> , measurements, observations	Information <b>providing context</b> and attributes
<b>Usage</b>	Analysis, decision making, <b>research</b>	Data management, discovery, <b>interpretation</b>
<b>Representation</b>	Mostly <b>unstructured</b>	<b>Structured</b>
<b>Relationship</b>	Independent and <b>stand-alone</b>	<b>Linked</b> to data/other metadata
<b>Purpose and Function</b>	Primary <b>source of information</b>	<b>Supporting framework</b> for organization, management, interpretation

# Basics of metadata: concepts

Data vs Metadata

Structured  
Metadata

Metadata  
Standards

Metadata  
Repositories

PIDs

Licences



**Format:** the language for knowledge representation and exchange



**Semantics:** the (agreed) terminology to describe the attributes



**Schema:** the template which specifies the expected attributes and how they should be structured



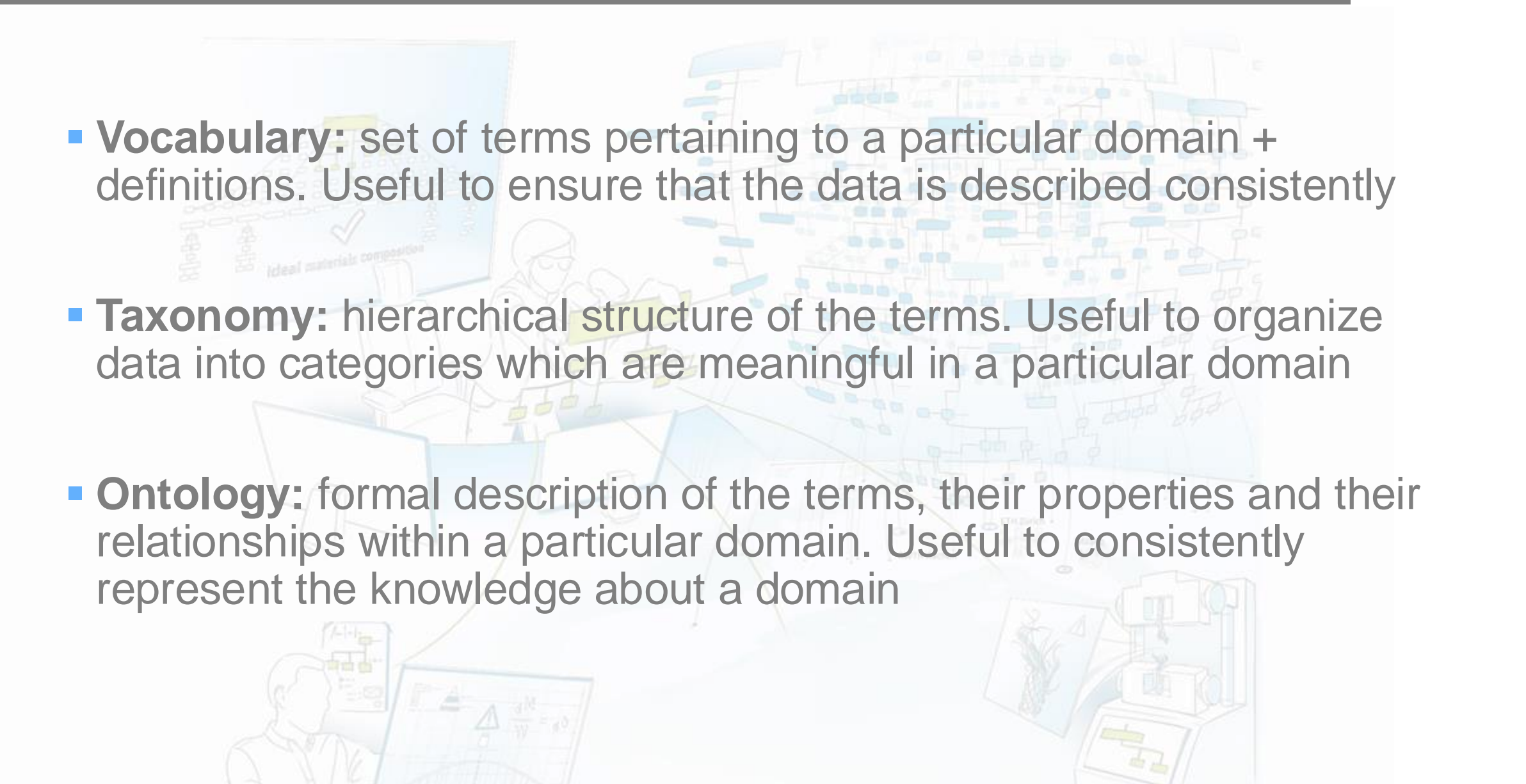
- eXtensible Markup Language
- Main purpose: transfer and storage of arbitrary data on the World Wide Web
- Human- and machine-readable
- Hierarchical (tree-like) structure
- Elements are wrapped in start `<...>` and end `</...>` “tags”

```
<example>  
  <title>This is the example title</title>  
  <description>A simple XML example</description>  
  <wordCount>1</wordCount>  
</example>
```

```
<person>  
  <firstName>John</firstName>  
  <lastName>Doe</lastName>  
</person>
```

- JavaScript Object Notation
- Main purpose: transfer and storage of data
- Human- and machine-readable
- Hierarchical structure
- Elements are defined in key:value pairs
- Elements are wrapped as {objects} or [arrays]

```
{  
  "key":"value",  
  "aString":"string",  
  "anInteger":5,  
  "aFloat":0.5,  
  "aBoolean":true,  
  "anArray": ["item1", "item2", "item3"],  
  "anObject": {  
    "key1":"value1",  
    "key2":"value2",  
    "key3":"value3"  
  }  
}
```

- 
- **Vocabulary:** set of terms pertaining to a particular domain + definitions. Useful to ensure that the data is described consistently
  - **Taxonomy:** hierarchical structure of the terms. Useful to organize data into categories which are meaningful in a particular domain
  - **Ontology:** formal description of the terms, their properties and their relationships within a particular domain. Useful to consistently represent the knowledge about a domain

- Template which specifies the expected elements and how they are structured:
  - Names (vocabulary)
  - Value types
  - Rules
  - Mandatory/optional
- XML Schema Definition (XSD): less frequently used
- JSON Schema: uses the JSON Schema Vocabulary <https://json-schema.org> to specify & syntactically validate the structure



# Metadata Schema vs Metadata Document

## Metadata Schema



Photo by M. Coghlan on Flickr (licence CC-BY-SA 2.0)

## Metadata Document



Photo by M. Carrati on Unsplash

## Metadata Schema

Template which specifies the expected elements and how they are structured

```
"givenName": {  
  "type": "string",  
  "description": "(Optional) - Given name of the user"  
},  
"familyName": {  
  "type": "string",  
  "description": "(Optional) - Family name of the user"  
},  
"age": {  
  "type": "number",  
  "description": "(Optional) - Age of the user"  
},
```

## Metadata Document

An *instance* of a Metadata Schema which describes a given resource and conforms to the specified definitions

```
"givenName": "John",  
"familyName": "Doe",  
"age": 20
```

More on this later



- Same parameters for all data, more harmonized description
- Structured metadata can be more easily interpreted (also by machines)
- Results can be more easily reproduced/reused
- Data can be more easily compared/exchanged
- Data can be found based on their attributes (search, filter)
- Metadata can be validated.
  - Note: schema validation only checks for **syntactical validity** (required property, corresponding value, and whether the value conforms with the expected data type)

# Basics of metadata: concepts

Data vs Metadata

Structured  
Metadata

Metadata  
Standards

Metadata  
Repositories

PIDs

Licences



- Metadata schemas which are well-established, endorsed, and widely accepted by the user community

General purpose

 **DublinCore** <http://dublincore.org/schemas/>

**DataCite** <http://schema.datacite.org>

**Schema.org** <https://schema.org>

```
<xs:sequence>
  <xs:choice minOccurs="0" maxOccurs="unbounded">
    <xs:element ref="any"/>
  </xs:choice>
</xs:sequence>
```

```
<xs:element name="givenName" minOccurs="0"/>
<xs:element name="familyName" minOccurs="0"/>
```

```
"colleague": [
  "http://www.xyz.edu/students/alicejones.html",
  "http://www.xyz.edu/students/bobsmith.html"
],
```

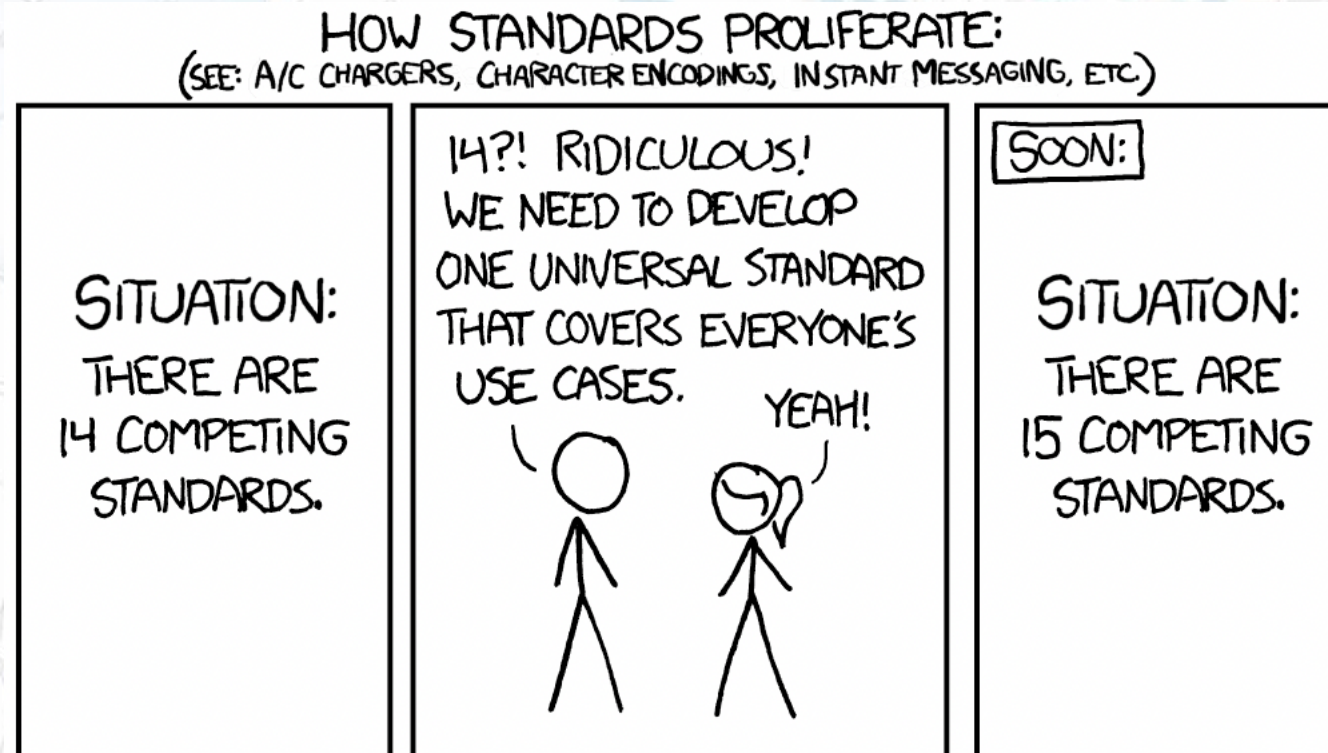
Neutron, x-ray, muon

**NeXus** <http://www.nexusformat.org>

```
entry:NXentry
  raw:NXsubentry
    definition="NXsas"
  reduced:NXsubentry
    definition="NXcanSAS"
  fluo:NXsubentry
    definition="NXfluo"
```

# Take-home message

Before describing your data on your own, you should look for existing metadata schemas or standards.



What if a  
standard  
does not  
exist?

Where do I  
publish my  
metadata?

# Questions?

How do I link  
my data to  
metadata?

How can my  
(meta)data be  
reused and  
cited?

# Basics of metadata: concepts

Data vs Metadata

Structured  
Metadata

Metadata  
Standards

Metadata  
Repositories

PIDs

Licences



- *“Information system used to store, manage and provide access to Metadata, following a policy or a set of rules that define storage and access norms.”* Aversa R., et al. (2024) DOI: 10.5281/zenodo.10663833
- Register/find metadata schemas
- Register/find metadata documents
- Validate metadata documents against the schema
- Versioning
- Access control management
- User authentication

More on this with later

# Basics of metadata: concepts

Data vs Metadata

Structured  
Metadata

Metadata  
Standards

Metadata  
Repositories

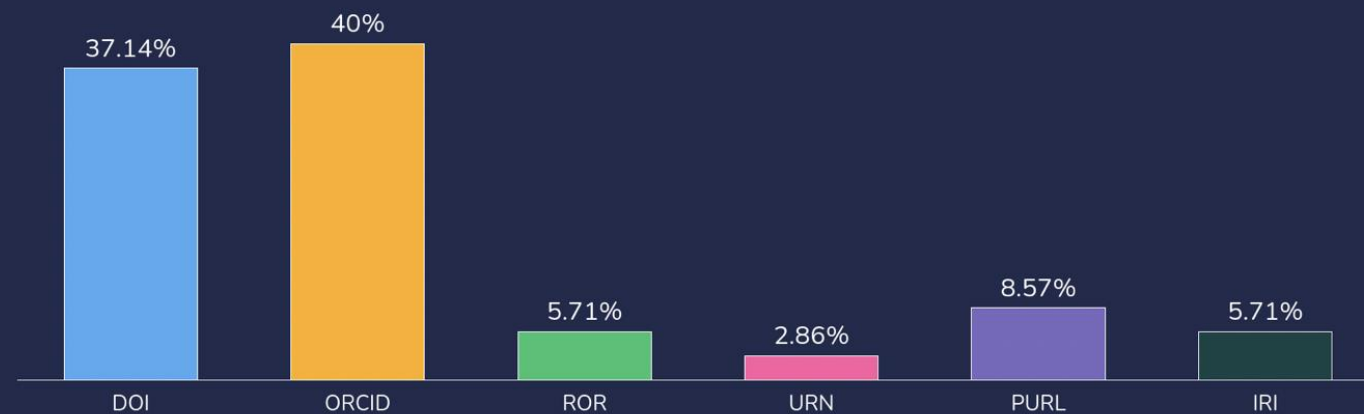
PIDs

Licences

- **Persistent identifier:** long-lasting reference to locate and identify a resource, even if it changes over time
- Globally unique and persistent over time (until the PID provider is maintained!)
- Connected to a set of metadata describing a resource rather than to the resource itself
- **Benefit of PIDs:** allow different platforms to exchange information consistently and unambiguously, e.g. to track citations and reuse.

Mint, manage and resolve PIDs.

Does it ring a bell? Which of these terms do you know?



- DOI: Digital Object Identifier
- ORCID: Open Researcher and Contributor ID
- ROR: Research Organization Registry
- URN: Uniform Resource Name
- PURL: Persistent Uniform Resource Locator
- IRI: Internationalized Resource Identifier



# Take-home message

- PIDs are largely used to identify researchers, institutions, research articles, data resources, metadata records, code, software, ...
- Ready to publish?



# Basics of metadata: concepts

Data vs Metadata

Structured  
Metadata

Metadata  
Standards

Metadata  
Repositories

PIDs

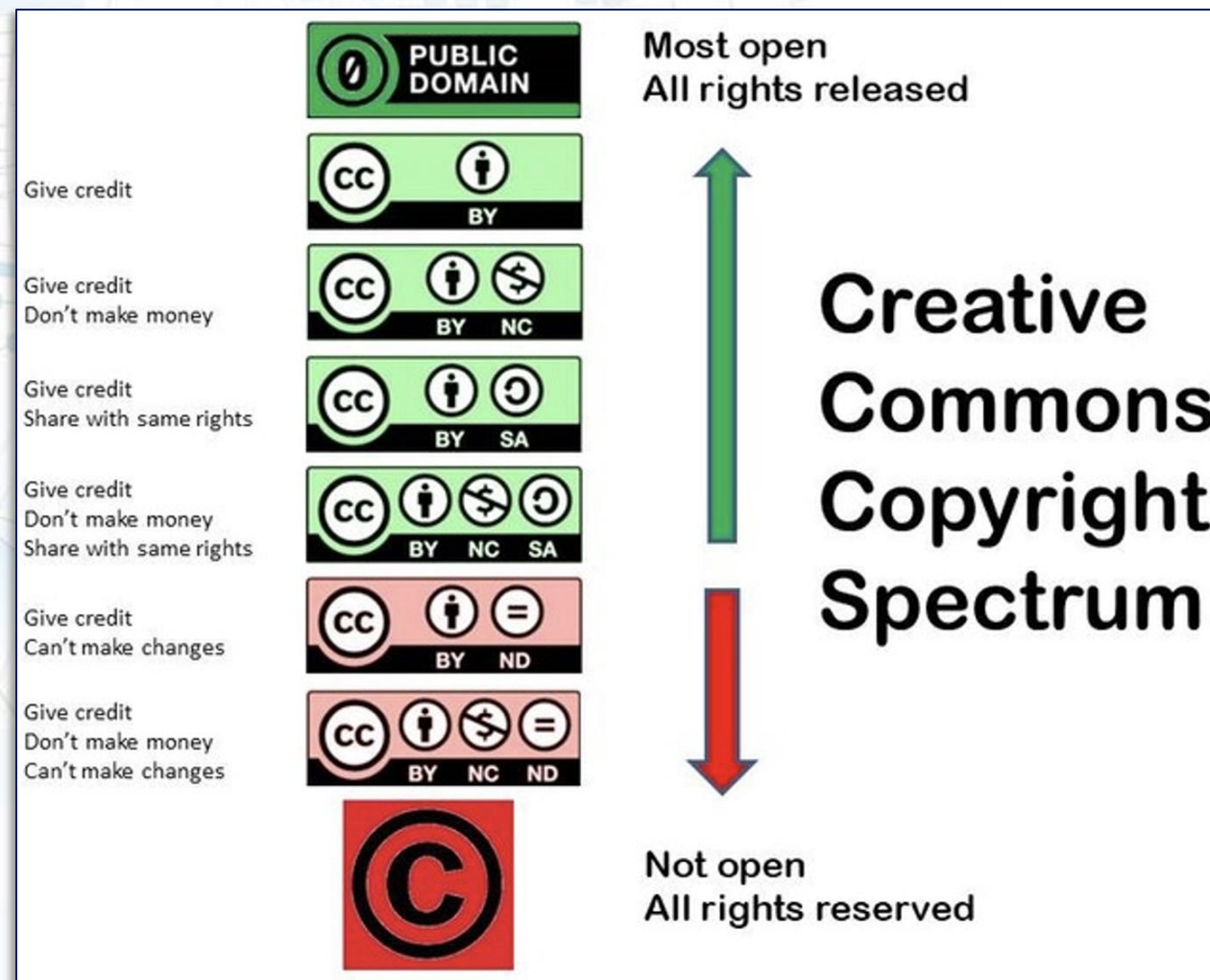
Licences

Legal arrangement between the creator of the data and the end-user, or the place where the data will be deposited, specifying what users can do with the data.

## Rights



Creative Commons Attribution 4.0 International

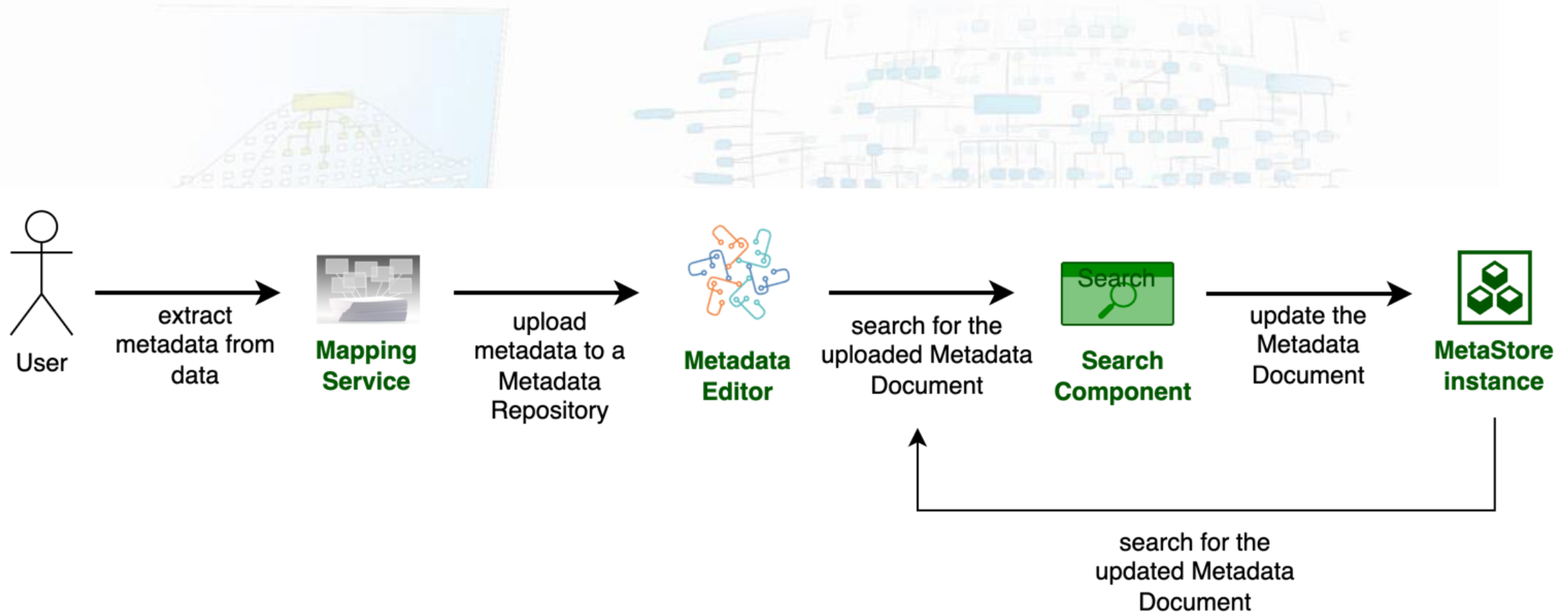


- FAIR Principles are guidelines for (meta)data management
- Structured metadata are helpful for better data interpretation, exchange, reuse
- Adoption of existing standards or community best practices avoids proliferation of descriptions
- Persistent and globally unique identifiers allows you to link resources and cite them
- Licences are keys when you publish or reuse results

Let's put it in practice!



# Use Case Introduction



# Registration to the NFDI-MatWerk Metadata Repository



- <https://t1p.de/schema-management>

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

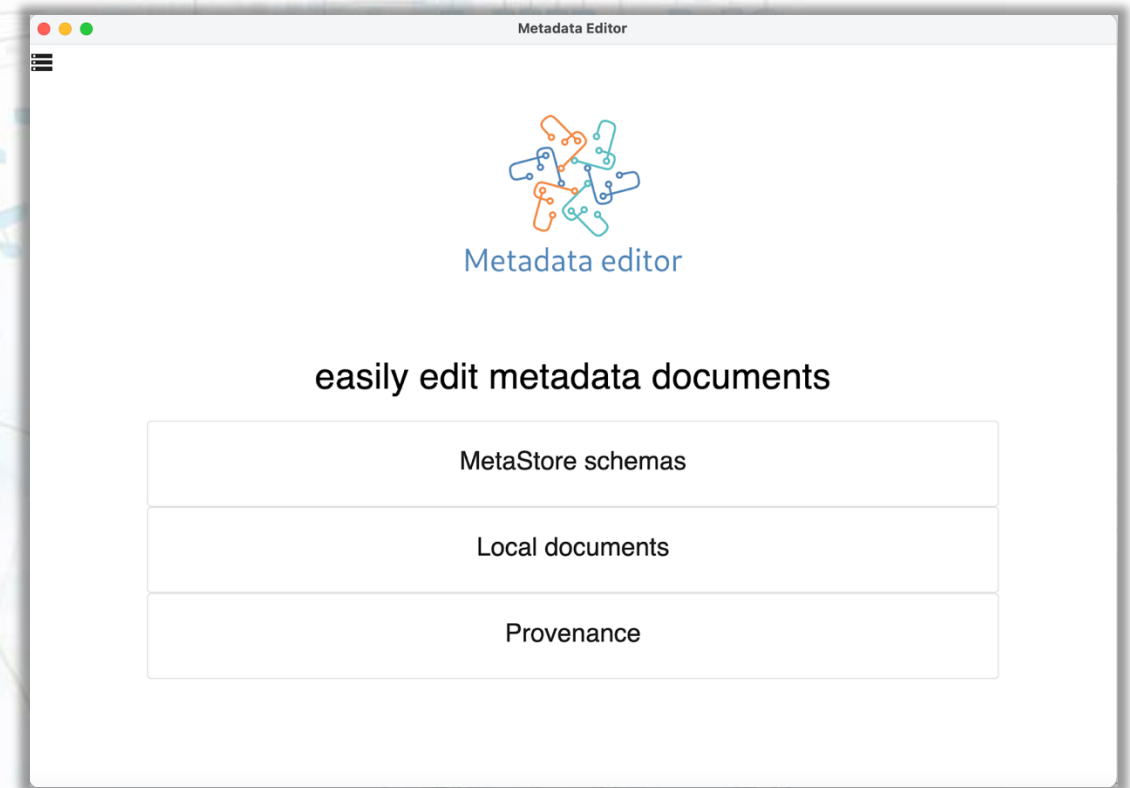
Schema Management Metadata Management Search Show/Hide Filters Not logged in **Login**

**Schema Documents**

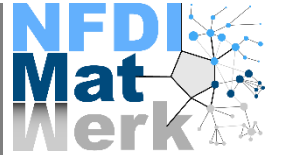
Identifier	Version	Type	Label
▼ <b>ape-he (14)</b> (1 item)			
ape-he	14	JSON	raw data
Date Updated 2024-11-29 09:03			
▼ <b>iuc02_creep_data (1)</b> (1 item)			
iuc02_creep_data	1	JSON	IUC02 - Reference Creep Data
Date Updated 2025-01-13 08:18			
▼ <b>lab_ct (1)</b> (1 item)			
lab_ct	1	JSON	raw data
Date Updated 2025-01-10 14:30			
▼ <b>mri_schema (1)</b> (1 item)			
mri_schema	1	JSON	raw data
Date Updated 2024-09-04 18:29			
▼ <b>pp13-draft-dataset-nested (2)</b> (1 item)			
pp13-draft-dataset-nested	2	JSON	PP13 Draft for dataset (nested)
Date Updated 2022-12-15 13:20			
▼ <b>pp13-draft-main-nested (2)</b> (1 item)			
pp13-draft-main-nested	2	JSON	Draft for main Schema (nested)
Date Updated 2022-12-15 13:22			
▼ <b>pp13-dummy-v2 (1)</b> (1 item)			
pp13-dummy-v2	1	JSON	
Date Updated 2022-10-14 09:16			
▼ <b>pp13-xct-draft (3)</b> (1 item)			

# Installation of the Metadata Editor

- Link: <https://t1p.de/md-download>
- Select the suitable one for your operating system.
- All instructions are described in details and should be followed.



# Acknowledgements



## Contributions:

Rossella Aversa, Thomas Jejkal, Andrea Recchia

## Used material:

Aversa, A. (2024). The journey towards Metadata Management. DOI: 10.5445/IR/1000174166

Gerlich, S., Strupp, A., Hofmann, V., Sandfeld, S. (2023). Fundamentals of Scientific Metadata. The Carpentries Incubator. DOI: 10.5281/zenodo.10091708

## Founded by:

the Joint Laboratory Model and Data driven Materials Characterization (JL MDMC), a cross-centre platform of the Helmholtz Association; the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project; the research program “Engineering Digital Futures” of the Helmholtz Association of German Research Centers; the Helmholtz Metadata Collaboration Platform.

## Funded by



Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure – NFDI 38/1 – project number 460247524

Illustrations by:  
© Fraunhofer IWM, Illustrations: Gebhard|Uhl Freiburg