Integration of controlled vocabularies or ontologies in metadata schemas

Gulzaure Abdildina, Rossella Aversa, Thomas Jejkal, Philipp Ost

Metadata is an essential component of research data management. Metadata enhance compliance of the data with the FAIR principles. One of the most important methods for improving the quality of metadata and increasing the level of machine readability and actionability is the integration of vocabularies and ontologies in metadata (see Interoperability principle I2, https://www.go-fair.org/fair-principles/).

Metadata-Schemas-for-Materials-Science / SEM_schema.json

1237 lines (1237 loc) · 73.5 KB

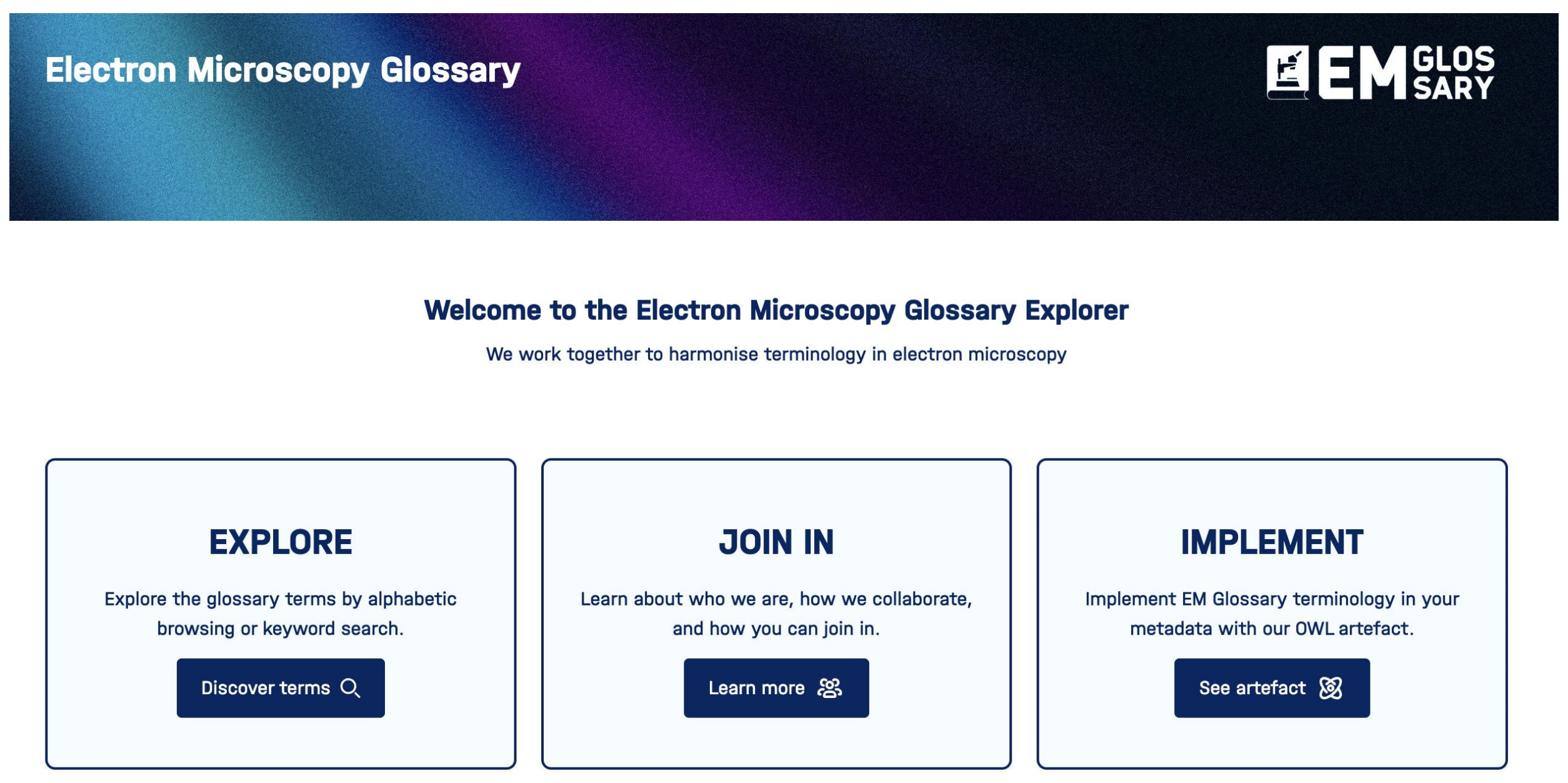
```
{
    "$schema":"https://json-schema.org/draft/2019-09/schema",
    "$id":"https://mdmc.kit.edu/schemas/sem_metadata",
    "type":"object",
    "description":"This JSON Schema describes a Scanning Electron Microscopy
    "properties":{
        "entry":{
            "$ref":"#/$defs/entry"
        }
    },
    "required":[
        "entry"
    ],
    "title":"sem",
```

https://metarepo.nffa.eu/frontend/dashboard.html

The Nanoscience Foundries and Fine Analysis (NFFA)-Europe Pilot (NEP), the Helmholtz Joint Lab "Integrated Model and Data-driven Materials Characterization" (JL-MDMC), and the National Research Data Infrastructure for Materials Science & Engineering (NFDI-MatWerk) collaboratively created **the Scanning Electron Microscopy (SEM) metadata schema**, to describe SEM images in order to facilitate data reuse and exchange.

Challenges:

- The metadata consists of hundreds of terms.
- Maintaining it is a time-consuming process.
- The FAIRness of the metadata should be increased.



Electron Microscopy Glossary

Welcome to the Electron Microscopy Glossary Explorer

We work together to harmonise terminology in electron microscopy

EXPLORE

Explore the glossary terms by alphabetic browsing or keyword search.

JOIN IN

Learn about who we are, how we collaborate, and how you can join in.

IMPLEMENT

Implement EM Glossary terminology in your metadata with our OWL artefact.

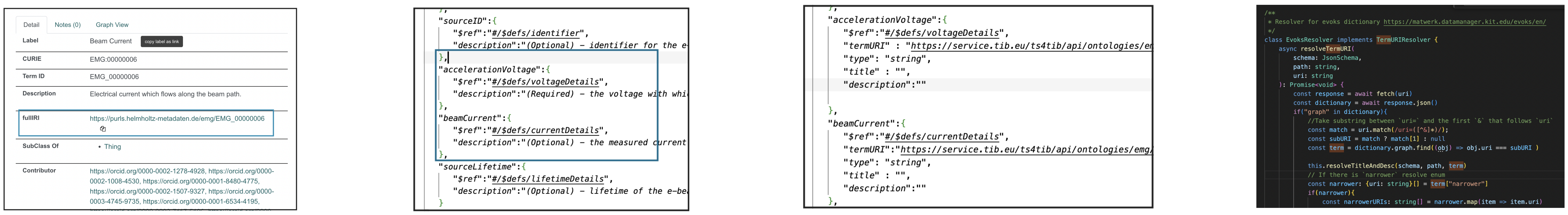
https://emglossary.helmholtz-metadaten.de/

To describe the terminology and semantics of Electron Microscopy techniques at the application level, the Helmholtz Metadata Collaboration (HMC) developed **the Electron Microscopy glossary (EM glossary)** using the Web Ontology Language (OWL).

Each term of the EM glossary has an Internationalized Resource Identifier (IRI) associated to it, which allows it to be unambiguously resolved in a machine-readable way.

Integration of the EM glossary with SEM Metadata Schema

The EM glossary is registered and published in the Terminology Service (TIB). TIB provides an API through which information can be retrieved from each term using its IRI. By integrating them into the metadata schema, hard-coding of term properties is no longer required.



Label	Beam Current
CURIE	EMG:00000006
Term ID	EMG_00000006
Description	Electrical current which flows along the beam path.
fullIRI	https://purls.helmholtz-metadaten.de/emg/EMG_00000006
SubClass Of	Thing
Contributor	https://orcid.org/0000-0002-1278-4928, https://orcid.org/0000-0002-1008-4530, https://orcid.org/0000-0001-8480-4775, https://orcid.org/0000-0002-1507-9327, https://orcid.org/0000-0003-4745-9735, https://orcid.org/0000-0001-6534-4195,

EM glossary

```
"sourceID":{
    "$ref":"#/$defs/identifier",
    "description":"(Optional) - identifier for the e
},
"accelerationVoltage":{
    "$ref":"#/$defs/voltageDetails",
    "description":"(Required) - the voltage with whi
},
"beamCurrent":{
    "$ref":"#/$defs/currentDetails",
    "description":"(Optional) - the measured current
},
"sourceLifetime":{
    "$ref":"#/$defs/lifetimeDetails",
    "description":"(Optional) - lifetime of the e-be
}
```

SEM metadata schema

```
"accelerationVoltage":{
    "$ref":"#/$defs/voltageDetails",
    "termURI" : "https://service.tib.eu/ts4tib/api/ontologies/em
    "type": "string",
    "title" : "",
    "description":""

},
"beamCurrent":{
    "$ref":"#/$defs/currentDetails",
    "termURI":"https://service.tib.eu/ts4tib/api/ontologies/emg/
    "type": "string",
    "title" : "",
    "description":""
},
```

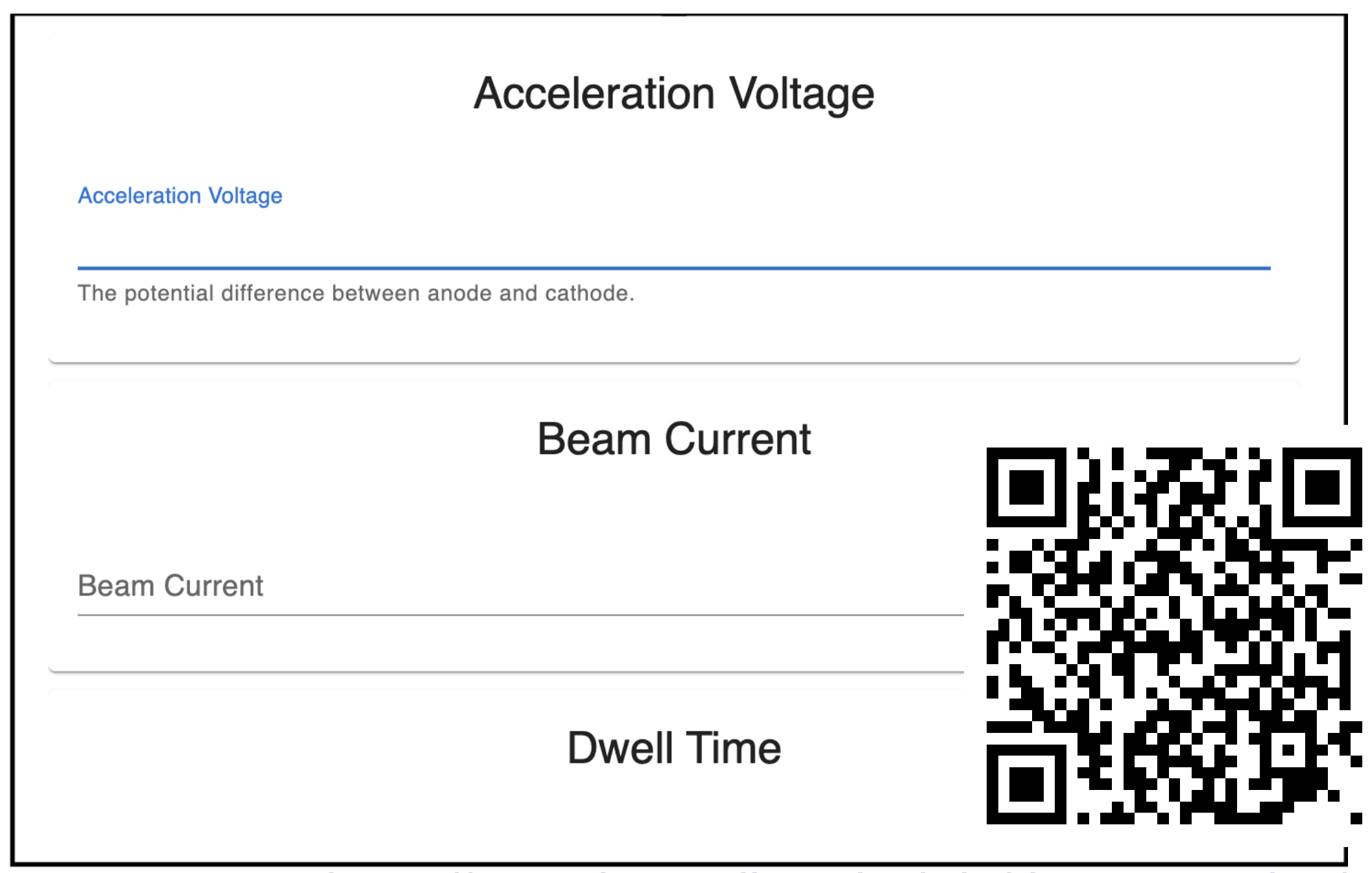
SEM metadata schema with integrated term IRIs

```
/**
 * Resolver for evoks dictionary https://matwerk.datamanager.kit.edu/evoks/en/
 */
class EvoksResolver implements TermURIResolver {
    async resolveTermURI(
        schema: JsonSchema,
        path: string,
        uri: string
    ): Promise<void> {
        const response = await fetch(uri)
        const dictionary = await response.json()
        if("graph" in dictionary){
            //Take substring between `uri=` and the first `&` that follows `uri`
            const match = uri.match(/uri=([^&]*)/);
            const subURI = match ? match[1] : null
            const term = dictionary.graph.find((obj) => obj.uri === subURI )

            this.resolveTitleAndDesc(schema, path, term)
            // If there is `narrower` resolve enum
            const narrower: {uri: string}[] = term["narrower"]
            if(narrower){
                const narrowerURIs: string[] = narrower.map(item => item.uri)
```

Resolving the IRIs in the NEP Metadata Editor

Integration of the term IRIs into the SEM metadata schema and resolving them in the NEP Metadata Editor form



Acceleration Voltage

Acceleration Voltage

The potential difference between anode and cathode.

Beam Current

Beam Current

Dwell Time

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

The Metadata Editor form automatically resolves the terms in the metadata schema and retrieves their properties (e.g., title, description, controlled list of items) from the EM glossary, supporting the user in filling the corresponding fields.

Results:

- use of controlled vocabularies and ontologies that follow FAIR principles;
- terms in schema referenced through the IRI rather than hard-coded;
- centralized maintenance of the terms;
- automatic propagation of any changes performed in the vocabulary or ontology;
- enabling the machine readability and actionability of the descriptive metadata.

This work was supported by NFDI-MatWerk [DFG, n. 460247524]; NFFA-Europe-Pilot [EU H2020 – n. 101007417]; joint Laboratory Model and Data-driven Materials Characterization (JL MDMC), a cross-centre platform of the Helmholtz Association; the Helmholtz Metadata Collaboration Platform (HMC).

Contact and feedback: gulzaure.abdildina@kit.edu