DISO: A Domain Ontology for Modeling Dislocations in Crystalline Materials

Ahmad Zainul Ihsan
Forschungszentrum Jülich, Institute of Advanced Simulation – Materials Data Science and Informatics (IAS-9), Germany
a.ihsan@fz-juelich.de

Said Fathalla
Forschungszentrum Jülich, Institute of Advanced Simulation – Materials Data Science and Informatics (IAS-9), Germany
Faculty of Science, University of Alexandria, Egypt
s.fathalla@fz-juelich.de

Stefan Sandfeld
Forschungszentrum Jülich, Institute of Advanced Simulation – Materials Data Science and Informatics (IAS-9), Germany
RWTH Aachen University, Faculty 5, Germany
s.sandfeld@fz-juelich.de

ABSTRACT

Crystalline materials, such as metals and semiconductors, nearly always contain a special defect type called dislocation. This defect decisively determines many important material properties, e.g., strength, fracture toughness, or ductility. Over the past years, significant effort has been put into understanding dislocation behavior across different length scales via experimental characterization techniques and simulations. This paper introduces the dislocation ontology (DISO), which defines the concepts and relationships related to linear defects in crystalline materials. We developed DISO using a top-down approach in which we start defining the most general concepts in the dislocation domain and subsequent specialization of them. DISO is published through a persistent URL following W3C best practices for publishing Linked Data. Two potential use cases for DISO are presented to illustrate its usefulness in the dislocation dynamics domain. The evaluation of the ontology is performed in two directions, evaluating the success of the ontology in modeling a real-world domain and the richness of the ontology.

CCS CONCEPTS

• **Computing methodologies** → **Ontology engineering**; • **Information systems** → **Web Ontology Language (OWL)**; **Ontologies**;

KEYWORDS

Ontology, Dislocation, Crystallographic Defects, Linked Data, Materials Science and Engineering

1 INTRODUCTION

A dislocation is a one-dimensional lattice defect responsible for plastic deformation in metals and other crystalline materials. It plays an essential role in determining mechanical properties such as strength and ductility. Since the existence of the dislocation was postulated in the 1930s by Orowan [21], Taylor [34], and Polanyi [23], significant efforts have been undertaken to understand all details of systems of dislocations based on, e.g., dedicated microscopy techniques for visualizing dislocations. These techniques and various simulation methods were developed to predict the evolution of dislocation.

In the recent years, data-driven approaches have brought new methods and tools for analyzing and understanding the evolution of dislocation systems [25, 26, 32]. Similarly, the whole MSE field, a parent domain of the specialized domain of dislocations, is undergoing an unprecedented change. This change brings simulations and experiments together and ultimately makes digital transformation in the field of Materials Science and Engineering (MSE) possible [1, 17, 24].

While dislocations can be directly represented through the positions of atoms, in many cases, the idealized representation of the dislocation as a mathematical line is preferred as it allows to consider larger systems with fewer degrees of freedom. The "scale" on which these mathematical lines are defined is larger than the atomic scale (atoms are not visible there) but still contains effective atomic scale information in the form of the dislocation microstructure. Therefore, this is often called the "mesoscale" even though this is not a well-defined notion. To fully understand the behaviour of materials, aspects from different length scales need to be considered. This makes the knowledge representation of systems of dislocations challenging, even though so far this has not been perceived as a significant research hindrance in materials science. Formal knowledge representation, i.e., via ontologies, enables interoperability and data handling between related MSE domains, thus enabling machine actionability. Ontology also allows the domain knowledge to be represented by logical axioms that the machine can understand. Ultimately, an ontology is an important component to enable the establishment of the FAIR principles (Findable, Accessible, Interoperable, and Reusable) [36] in the MSE domains.

This paper represents the dislocation ontology (DISO), an ontology representing the concepts and relationships in linear defects

in crystalline materials. DISO is an extended version of the ontology presented in Ihsan et al. [15]. Indeed, DISO is a part of the Dislocation Ontology Suite (DISOS)[1]. DISOS is an ontology suite comprising several modules describing materials scientific concepts, representations of dislocations, and different simulation models in the dislocation domain. The contributions can be summarized as follows:

- covering further characteristics of dislocation, including the mathematical and numerical representation of the dislocation line,
- adding new concepts, including Bravais Lattice, Space Group, Point Group, Crystal System, and Vector,
- demonstrating the usefulness of the new version by showing its ability to answer a set of predefined queries in various complexity,
- creating human-readable documentation for the ontology as well as adding ontology metadata (e.g. contributor, title, date created, etc.), which will greatly improve its findability and reusability,
- presenting two potential use cases for DISO: Dislocation dynamics data and dislocation experiment data,
- creating RDF dataset of dislocation structure data produced by MoDELib software [22] by mapping data stored in Hierarchical Data Format 5 (hdf5)[2] files to RDF triples using DISO classes, and
- evaluating the ontology using two different evaluation strategies.

DISO is developed and maintained on a GitHub repository[3]. The ontology is available (in several RDF serializations) via a persistent identifier (i.e., https://purls.helmholtz-metadaten.de/disos/diso) provided by PIDA (Persistent Identifiers for Digital Assets)[4]. PIDA is a service that provides persistent identifiers for digital assets. It employs content negotiation [4], which is a mechanisms defined as a part of HTTP serve different versions of a resource (i.e. the HTML documentation or an RDF representation) at the same URI in accordance with the client's request. The dereferenceability of the IRIs of the ontology terms over the HTTP protocol (cf. [19]) has been checked using a Linked Data validator (i.e. Vapour[5]). Furthermore, DISO is syntactically validated by the W3C RDF validation service[6] to conform with the W3C RDF standards. To make it easier to understand and reuse our ontology, human-readable documentation of the ontology is generated and can be found online via its URI.

The rest of this paper is organized as follows: Section 2 presents related work of materials science domain ontologies and points out the existing gaps. Section 3 describes the physical aspects of dislocations and in section 4, the development of the ontology is described. Section 5 shows two potential use cases and section 6 presents the evaluation of DISO using a criteria-based strategy. Finally, section 7 concludes and sketches the envisioned future work.

2 RELATED WORK

In the past decades, several researchers have been involved in the knowledge representation of various fields of science [7], including Materials Science and Engineering (MSE), through developing domain ontologies.

Plinius [35] is an ontology developed for describing ceramic materials covering the conceptualization of chemical compositions ranging from the single atom to complex chemical substances. In the "Materials Ontology" [3], a detailed structure of materials information consisting of substances, processes, environment, and properties was conceptualized.

The "*International Union of Crystallography*"[7] published and distributed the Crystallographic Information File (CIF) and the further developed CIF2 which currently co-exists with the CIF file format. CIF serves as a *de facto* standard for crystallographic information format exchange. The authors of CIF in [13] also have further developed the STAR/CIF Ontology that was written in the mathematical symbolic script language called dREL [31].

Another ongoing effort to establish semantic standards that apply at the highest possible level of abstraction is the *Elementary Multi-perspective Material Ontology* (EMMO)[8]. EMMO has been developed in the context of the *European Materials Modelling Council*[9] (EMMC) and provides a common semantic framework for describing material models, characterization, and data with the possibility of extension and adaptation to other domains. However, EMMO currently contains only a small number of sub-domains. Furthermore, it does not include the domain of dislocations which is of great importance for materials scientists in the context of mechanical deformation of nano- and micometer sized specimens.

Recent development work in materials design (e.g., with regards to ab-initio methods) and crystallographic information resulted in the *Materials Design Ontology* (MDO) [20] and the *Crystal Structure Ontology* (CSO)[10], respectively. The former represents information on the atomic structure of materials via its "structure" ontology module. The latter represents crystallographic information needed to describe the crystallographic defect. However, both do not explicitly represent the physical and conceptual aspects of the crystal structure that are directly related to the representation of defects in crystals, particularly dislocations.

In summary, it can be concluded that even though significant progress has been made concerning the ontology design in a number of related domains, it becomes clear that a level-domain ontology of dislocations in crystalline materials is still missing. Furthermore, there are several attempts to represent materials science phenomena in the literature, even though few researchers have addressed the problem of semantically representing crystalline defects. Therefore, developing a domain ontology of dislocations in crystalline materials fills a clear gap and is an important step for the domain of materials science.

---

[1]https://purls.helmholtz-metadaten.de/disos
[2]https://www.hdfgroup.org/solutions/hdf5/
[3]https://github.com/Materials-Data-Science-and-Informatics/
Dislocation-Ontology-Suite/tree/main/DISO
[4]https://purls.helmholtz-metadaten.de/
[5]http://linkeddata.uriburner.com:8000/vapour?
[6]https://www.w3.org/RDF/Validator/

[7]https://www.iucr.org/
[8]https://emmo-repo.github.io
[9]https://emmc.eu/
[10]https://purls.helmholtz-metadaten.de/disos/cso

3 DESCRIPTION OF THE DOMAIN

In real crystalline materials, atoms are typically not perfectly ordered or positioned – at least not everywhere. Typically, various types of crystallographic defects in a piece of material destroy the local order of the crystal structure (on top of thermal fluctuations that also affect the atomic positions). Such a defect might be a point defect (e.g., a missing or extra atom) or, as shown in the left panel of Figure 1, a dislocation (a strongly localized, tube-like region of disorder containing the highly disordered dislocation core at the center).

According to the geometry, there are two types of dislocations, the so-called "edge" and "screw" dislocation. The latter is characterized by having a line sense parallel to its Burgers vector, whereas in the former, the line sense is perpendicular to its Burgers vector. Furthermore, also the atoms in the vicinity of the dislocation are displaced from their perfect lattice sites. This lattice distortion subsequently results in a stress field in the crystalline material around the dislocation.

From a *mesoscopic* point of view, details from the smallest considered scale, the scale of the individual atoms are no longer visible. Nonetheless, the dislocation line still can be observed: the tube-like defect "region" is reduced to an idealized mathematical line as shown in Figure 1 (see the center figure). Thus, moving from the atomic scale to such a mesoscale is an idealization and strongly reduces the amount of information to only the utmost necessary information.

A wide range of experimental and computational techniques are used to observe and predict dislocations in crystalline materials. E.g., on the atomic scale, field ion microscopy or high-resolution transmission electron microscopy is used to image the arrangement of atoms. Then on the mesoscale, these techniques determine the properties of individual dislocations and study the arrangement, distribution, and density of dislocations in crystalline materials. Examples of the techniques at the mesoscale are Transmission Electron Microscopy (TEM) and, as a simulation method, Discrete Dislocation Dynamics (DDD).

TEM is a microscopy technique that uses a particle beam of electrons, transmitted through a specimen and passing some lenses, ultimately generating a higher resolution (highly-magnified) image. The transmitted beam is deflected once it hits the dislocation. The deflection reduces the intensity of the transmitted beam and increases the intensity of the diffracted beam. The dislocation then appears as a dark line in the bright-field image.

The motion and interaction of dislocation lines during plastic deformation mainly follow the governing equations of (linear) elasticity theory. The DDD simulation method treats dislocations as mathematical lines represented as polygons or splines and moves them according to elastic interactions and further "local rules".

Both techniques mentioned above observe the dislocation as a mathematical line. However, they apply different approaches to represent the mathematical line. In TEM, the image containing the dislocation lines is represented by a pixels image. In DDD, the mathematical dislocation line needs to be discretized to solve the governing equations numerically. The particular type of discretization represents the mathematical line through the *numerical*
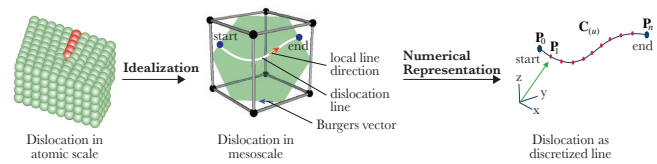
Figure 1: The idealization represents the dislocation in the mesoscale. Here, the individual atoms are no longer visible. This idealization reduced the tube-like defect "region" to a mathematical line. On the rightmost, the numerical representation of a mathematical line as the discretized line is illustrated.

representation. The discretization steps [12] are (also see the rightmost figure in Figure 1): the oriented curve of the dislocation line, $C(u)$, where $0 \leq u \leq 1$, is discretized into a number of segments, n_s. The position vector for any point on the segment is defined by $C(u) = P(u)$ and the shape of a segment is defined by choosing the *shape function*, $N_i(u)$, e.g., linear, circular, cubic spline segment. However, this is only one possible discretization type, and also others are commonly used in the DDD simulation community.
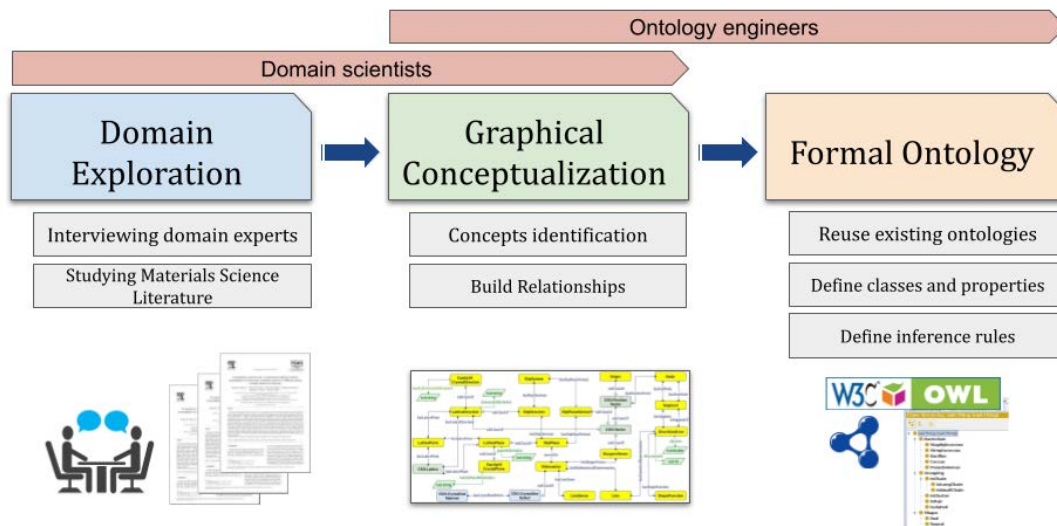
4 ONTOLOGY DEVELOPMENT

Developing domain ontologies attracts considerable interest from several private and public sector organizations to capture their knowledge about their domain of interest in a form that machines can understand. In fact, the main goal of developing ontologies is to share conceptualizations through which humans express meanings of things that can act as communication interfaces between humans and machines.

This section describes the development process of DISO, which is an iterative approach, which means we started with an initial version and then revise and refine the evolving ontology. Figure 2 illustrates the workflow of this process which is inspired from several well-known ontology development methodologies [18]. This figure presents the main phases, their sub-tasks, and the roles involved in the whole process. We have frequently interviewed domain scientists as well as ontology engineers during the whole process in order to improve the final outcome. This continues through the entire development process.

4.1 Ontology Metadata

An important step towards improving ontology reuse is to provide a systematic and comprehensive description of ontologies [29], i.e., ontology metadata. Often, ontologies are not well documented or miss information about the ontology itself. Consequently, these ontologies are therefore suffering from several problems, including 1) being not optimally accessible to potential users, 2) hindering their reusability, and 3) identifying required ontologies for specific applications efficiently. Accordingly, several DCMI Metadata Terms[11] are added to the ontology, involving `terms:contributor`, `terms:created`, `terms:title`, and `vann:preferredNamespacePrefix`. As a next step, providing

---

[11]https://www.dublincore.org/specifications/dublin-core/dcmi-terms/

Figure 2: The workflow of the dislocation ontology development, illustrating the main phases, subprocesses, and roles involved in the whole process.

human-readable documentation for the ontology is also crucial because it is one of the main reuse problems identified in [11].

4.2 Reuse of existing models

One of the initial steps in the ontology development is to reuse terms (i.e., classes or properties) from existing ontologies that describe the same domain or topic. However, it is challenging for ontology engineers to decide which of the existing ontologies is suitable to be reused. In fact, the more ontologies a model reuses, the higher the value of its semantic data is. Therefore, reusing terms from other ontologies greatly increases semantic data value [10] of the developed ontology.

After studying the literature, we decided to reuse several concepts from two related ontologies, namely, *Crystal Structure ontology* (CSO) and *Crystalline Defect ontology* (CDO). The former is an ontology developed to represent crystallographic information needed to describe the dislocation. The latter is an ontology connecting the physical materials entity to the crystal structure and several defects in crystalline materials, e.g., point defect, dislocation, and grain boundary.

Crystalline Defect Ontology (CDO)[12]. In CDO, the `EMMO:Material`, which is a class from EMMO[13], is reused to describe the physical entity of crystalline materials. The `CDO:CrystallineMaterial` class is subsequently a subclass of `EMMO:Material`.

Crystal Structure Ontology (CSO)[14]. In CSO, several MDO [20] classes are reused to describe the crystal coordinate system, motif (an arrangement of chemical species in the crystal structure) in a crystal structure, point groups, and space groups. The standard coordinate system is defined by `MDO:Basis` and `MDO:CoordinateVector` classes that CSO reuses. The motif reuses `MDO:Occupancy`, `MDO:Site`, and `MDO:Species`. Subsequently, to

define the point groups and space groups of a crystal structure, `MDO:PointGroup` and `MDO:SpaceGroup` are reused. Lastly, to define the unit quantity of a property in CSO, CSO reuses several classes from QUDT (Quantities, Units, Dimensions and Data Types Ontologies) [14]. Classes that are reused from QUDT are the `QUDT:Quantity`, `QUDT:QuantityKind`, and `QUDT:QuantityValue`.

4.3 Main Classes

The classes in our ontology are divided into two sets; 1) imported from the ontologies introduced in subsection 4.2 and 2) newly defined classes that are not explicitly defined in the existing related ontologies.

Imported classes. Several classes have been imported from the crystal structure ontology (CSO), involving `CSO:Lattice` which represents the periodic arrangement of one or more atoms, `CSO:Vector` which represents quantities that have both magnitude and direction, and imported from crystallographic defect ontology (CDO), involving `CDO:CrystallographicDefect` which represents lattice irregularity/lattice defects by having one or more of its dimensions on the order of an atomic diameter.
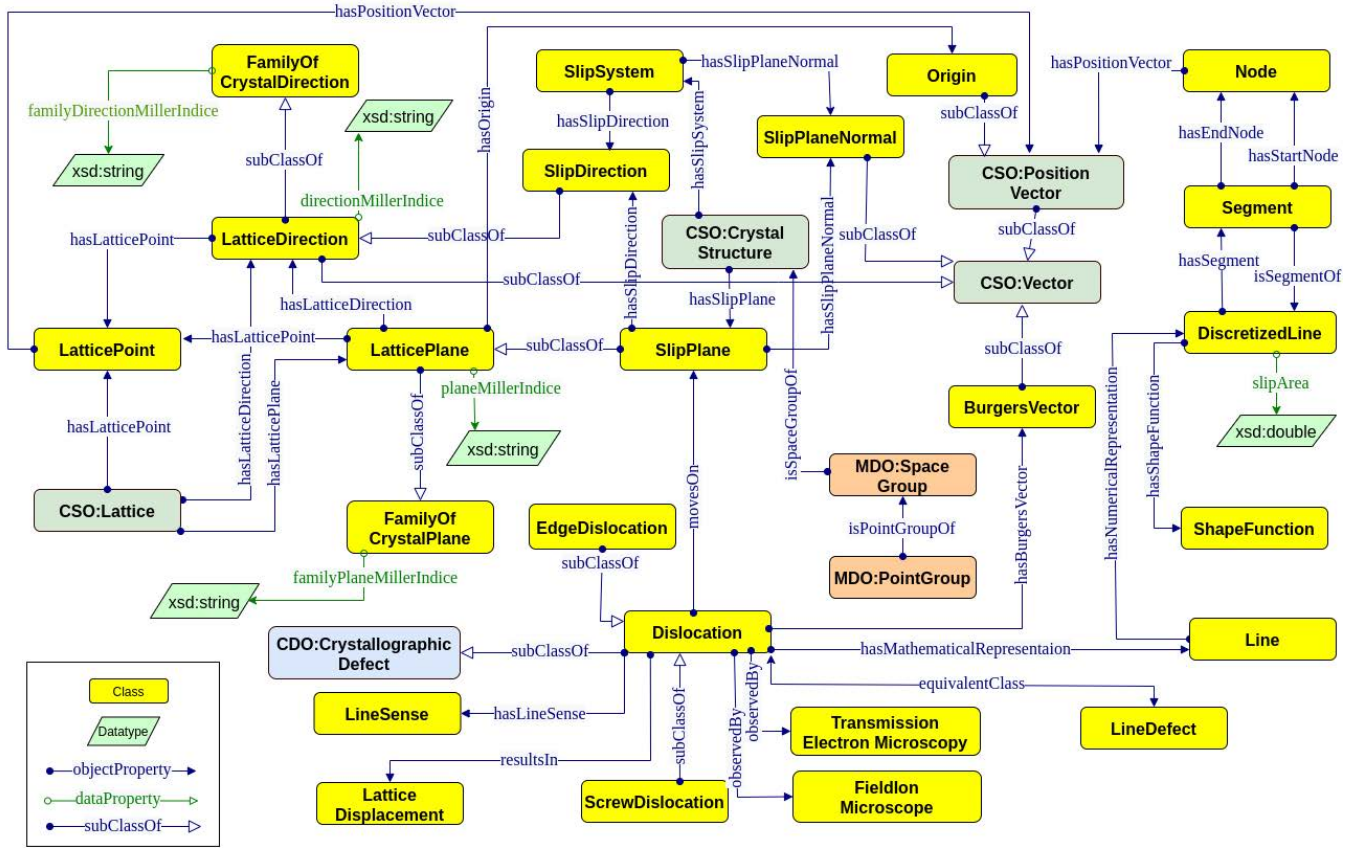
New classes. Here, the focus is put on some chosen classes of the main classes in the crystalline materials and line defect: 1) `Dislocation`, as the entity of main interest, which models a linear or one-dimensional defect around which some of the atoms are displaced, 2) `SlipPlane`, which models the lattice plane to which the dislocation is constrained to move, 3) `SlipDirection`, which models the lattice direction where the slip occurs in the crystalline materials, 4) `LatticePlane`, which models the lattice plane where it forms an infinitely stretched plane that cuts through lattice points. 5) `LatticeDirection`, which models the direction inside the lattice that connects two lattice points, and 6) `DiscretizedLine`, which models the numerical representation of the dislocation line as a mathematical line, e.g. an oriented curve, that is discretized into a number of segments.

---

[12]https://purls.helmholtz-metadaten.de/disos/cdo
[13]https://emmo-repo.github.io
[14]https://purls.helmholtz-metadaten.de/disos/cso

Figure 3: Core concepts and interconnected relationships in the DISO ontology. Arrows with open arrow heads denote `rdfs:subClassOf` properties between classes. Regular arrows visualize *rdfs:domain* and *rdfs:range* restrictions on properties. Furthermore, colored boxes represent different ontologies, e.g., MDO, CDO, CSO, and DISO.

4.4 Properties

Both data and object properties in DISO are divided into two categories: newly defined properties and reused ones:

New properties. The relationships between concepts are implemented as object properties, for example, the relationship between `Dislocation` and `TransmissionElectronMicroscopy` can be represented by `observedby` property. Similarly, the relationship between `Dislocation` and `LineSense` is represented by `hasLineSense` property. Furthermore, several data properties also have been defined, such as `directionMillerIndice`, `planeMillerIndice`, and `slipArea`.

Reused properties. Property `cso:hasPositionVector` from the CSO ontology and several data properties from DCterms for adding ontology metadata (see subsection 4.1) are reused. After defining new properties and identifying reused ones, the domain and range for each property using `rdfs:domain` and `rdfs:range` are defined, respectively. For instance, the domain of the data property `diso:slipArea` is `diso:DiscretizedLine` and the range is `xsd:double`.

Restricting properties. Several DISO classes use property restrictions, e.g., value constraints. For instances, the `observedby` property which connects `Dislocation` and `TransmissionElectronMicroscopy` is restricted by a value

constraint of `owl:someValuesFrom` and `hasBurgersVector` which connects `Dislocation` and `BurgersVector` is restricted by a value constraint of `owl:allValuesFrom`.

4.5 Instantiation of the Dislocation Ontology

As shown in Figure 4, an excerpt of the simulation data of dislocation structure annotated by DISO is illustrated. Due to the space limit, we only added a sample of crystallographic and dislocation information. We instantiate several crystallographic-related classes: crystal structure, Bravais lattice, crystal system, and unit cell. Furthermore, we instantiate several dislocation-related classes: dislocation, slip plane, slip direction, Burgers vector, and the representation of dislocation in numerical simulation, e.g., line, segment, and node. Lastly, the instance of crystalline material connects the crystallographic information, i.e., crystal structure and dislocation instances.

5 POTENTIAL USE CASES

In this section, we present two potential use cases for DISO: dislocation dynamics simulation data and dislocation experiment data.

Use case 1. Dislocation dynamic (DD) simulation is a computational technique/simulation to determine individual dislocation properties and study the arrangement, distribution, and density
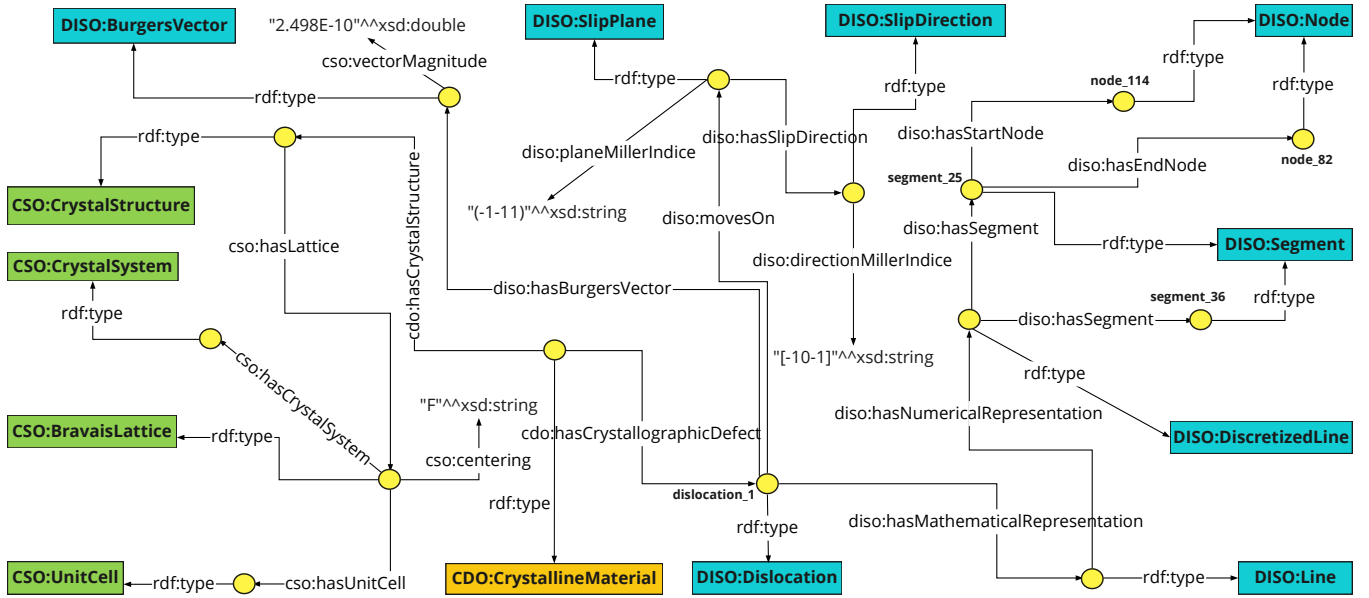
Figure 4: Sample of the instances of the dislocation dynamic data. Yellow points denote individuals of classes. Each individual is defined by an arrow having *rdf:type* relationship to the respective class and individuals are connected by object properties defined in DISO. Colored boxes represent classes from different ontologies, e.g., CDO, CSO, and DISO.

of dislocations in crystalline materials. In the dislocation domain, there are various DD software available, including MoDELib [22], ParaDIS [2], and microMegas [6]. Each software has a different set of metadata to structure the input and output of the simulation. In this regard, annotating/mapping the DD data/metadata to DISO enables data interoperability between various DD software. In addition, it enables semantic search over materials science databases, particularly in the dislocation domain.

Use case 2. An example of dislocation experiment data comes from Transmission Electron Microscopy (TEM). TEM is a microscopy technique that uses a particle beam of electrons transmitting through a specimen. In TEM, the generated data is a higher resolution (highly-magnified) image/video that observes dislocation in a crystalline material. While the TEM image/video already has some metadata to describe the data, the descriptive metadata about the observed dislocation is still lacking. In this regard, leveraging DISO along with the research data packager, such as RO-Crate [30], give the formal metadata description to the dislocation experiment data at ease, and ultimately it enables the data interoperability. Moreover, since RO-Crate is based on JSON-LD, annotating data with DISO is straightforward.

6 EVALUATION

The evaluation of the quality of an ontology can be performed by calculating metrics that evaluate the richness of the given ontology (i.e. criteria-based evaluation) [10]. In this section, we are evaluating our ontology in two directions, evaluating the success of the ontology in modeling a real-world domain and the quality of the ontology. In one direction, we carried out the following steps in order to evaluate the effectiveness of DISO: 1) defining a set of competency questions (CQs) based on domain expert feedback during several interviews, 2) formulating SPARQL queries corresponding

to CQs, 3) running the resultant SPARQL queries against the ontology instances (cf. subsection 6.1) using a SPARQL endpoint, 4) analyzing the query results by comparing them to the correct answers given by domain experts. On the other hand, we assess the ontology using OntoQA [33] evaluation model.

6.1 Data

The dislocation dynamics data produced is a virtual specimen of Nickel that consists of a number of dislocation lines. To fully describe the virtual specimen, the details of the crystal structure of Nickel, such as unit cell lattice parameters, point group, and space group data are taken from the Materials Project [16] repository.

In order to carry out the aforementioned steps, we prepared a dislocation dynamics dataset[15] that can be used for illustrating how DISO is useful by answering competency questions prepared by domain experts. Competency questions (CQs) determine what knowledge has to be entailed in the ontology as well as a way of evaluating how useful the ontology is by showing its capability to answer. Therefore, we write Python scripts[16] (using rdflib 6.0 [5] Python library) that map DDD MoDELib [22] software output with DISO classes and generate RDF triples ($\sim 4K$ triples), which can be found on DISO GitHub repository.

6.2 Competency Questions

Evaluating ontologies with competency questions is among the most widely used types of evaluating ontologies [8, 9, 27, 28].

A total of 20 CQs (defined in the domain exploration step (cf. Figure 2)) were collected and categorized into three categories: the crystal structure, the dislocation, and the vector information.

For illustration, we pick CQ1 from the CQs set (listed in Table 1). The corresponding SPARQL query to CQ1 is listed in Listing 1.

Table 1: Samples of competency questions.

No.	Competency Question
CQ1	What are the slip systems of a given crystal structure?
CQ2	What are the slip planes of a given crystal structure?
CQ3	In which slip planes does a dislocation move on?
CQ4	What is the Burgers vector of the dislocation?
CQ5	What is the Burgers vector magnitude of the dislocation?
CQ6	Given a slip plane of a crystal structure, what is the slip direction?
CQ7	What is the family of slip direction, given a slip direction in the crystal?
..	...
CQ20	Given a Burgers vector, what is its unit?

Listing 1: SPARQL query corresponding to CQ1.

```
1  PREFIX diso: <https://purls.helmholtz-metadaten.de/disos/diso#>
2  PREFIX cso: <https://purls.helmholtz-metadaten.de/disos/cso#>
3  PREFIX ex: <http://example.org/>
4  SELECT ?cryst_struc ?slip_sys ?slip_plane_normal_val
5  ?slip_direction_val WHERE{
6  ?crystal_structure a cso:CrystalStructure ;
7      diso:hasSlipSystem ?slip_system .
8  ?slip_system diso:hasSlipPlaneNormal ?slip_plane_normal ;
9      diso:hasSlipDirection ?slip_direction .
10 ?slip_plane_normal diso:directionMillerIndice ?slip_plane_normal_val.
11 ?slip_direction diso:directionMillerIndice ?slip_direction_val.
12 }
```

Table 2 shows the result of executing the above SPARQL query on the generated RDF data in which the same crystal structure consists of several slip systems.

Table 2: The result of CQ1 in Listing 1

Crystal Structure	Slip System	Plane Normal	Slip Direction
cryst_struc_0	slip_sys_0	[-1-11]	[0-1-1]
cryst_struc_0	slip_sys_1	[-11-1]	[110]
cryst_struc_0	slip_sys_2	[-111]	[-10-1]
cryst_struc_0	slip_sys_3	[1-1-1]	[01-1]
cryst_struc_0	slip_sys_4	[111]	[1-10]
cryst_struc_0	slip_sys_5	[-1-11]	[-10-1]
cryst_struc_0	slip_sys_6	[11-1]	[-101]
cryst_struc_0	slip_sys_7	[-11-1]	[011]

The complete set of the competency questions and the corresponding SPARQL queries can also be found in the DISO GitHub repository[17].

6.3 OntoQA Evaluation

In this section, we assess the richness of DISO by using a criteria-based evaluation called OntoQA [33]. OntoQA evaluates the ontology using schema metrics and population/instance metrics. To

evaluate the ontology design and its potential for rich knowledge representation, we use the following metrics:

- *Relationship richness (RR)* shows the diversity of relations and placement of relations in the ontology. Formally, it is defined by

$$RR = \frac{|P|}{|SC| + |P|} \qquad (1)$$

where **P** is the number of relationships in the ontology and **SC** is the number of sub-classes. The more relations the ontology owns, except *is-a* relations, the richer it is.

- *Attribute richness (AR)* shows the more slots/attributed that are defined, the more knowledge the ontology delivers. Formally, AR is defined by

$$AR = \frac{|AT|}{|C|} \qquad (2)$$

where **AT** is the number of attributes for all classes and **C** is the number of classes.

- *Inheritance richness (IR)* describes the distribution of information across different levels of the ontology inheritance tree. IR indicates how knowledge is grouped into different classes and sub-classes in the ontology. Formally, IR is defined by

$$IR = \frac{|SC|}{|C|} \qquad (3)$$

As shown in Table 3, we compare the results of DISO with MDO [20] and CSO[18]. The former is an ontology representing the domain of solid-state physics simulation in materials science. The latter represents crystallographic information needed to describe the dislocation. While the MDO ontology has a larger *RR*, i.e., it has more diversity in relations in the ontology. DISO and CSO have similar values in terms of *IR*, which means they represent a wider range of knowledge compared to MDO. Concerning *AR*, DISO has a larger *AR* than MDO but is slightly smaller than CSO, enabling more knowledge per instance, which is more useful in the DD data.

Table 3: OntoQA Evaluation of DISO

Ontology	C	SC	AT	P	RR	AR	IR
MDO	37	49	14	32	0.40	0.38	1.33
CSO	30	49	19	24	0.33	0.63	1.63
DISO	38	62	23	33	**0.35**	**0.61**	**1.63**

7 CONCLUSION AND OUTLOOK

In this paper, we presented the dislocation ontology, which is an ontology that defines the linear defect concepts and relations in crystalline materials. DISO is developed using a top-down approach and is published through a persistent identifier, following W3C best practices for publishing Linked Data. Two potential use cases for DISO are presented to illustrate its usefulness in the dislocation dynamics domain. We followed two evaluation strategies to demonstrate the usefulness and the quality of the ontology. We developed Python scripts to annotate the dislocation data generated from the DDD MoDELib software with DISO classes resulting in an RDF dataset.

---

[17]https://github.com/Materials-Data-Science-and-Informatics/
Dislocation-Ontology-Suite/blob/main/DISO/CQs/CQs.md

[18]https://purls.helmholtz-metadaten.de/disos/cso

We are planning to improve DISO by modeling the linear elasticity theory of dislocations, other defects (e.g., point defects, grain boundaries), and the application use cases in the dislocation domain, e.g., TEM data and dislocation simulation. In addition, ontology alignment into EMMO would also be a very worthwhile undertaking. This alignment extends the interoperability between domain ontologies, particularly within the MSE community. The ontology will continue to be maintained and extended in the context of the Helmholtz Metadata Collaboration efforts of facilitating machine readability and reuse of research data.

Since modularity is considered an emerging approach for developing ontologies, we are in the process of developing a dislocation ontology suite (DISOS) which comprises several modules each representing concepts in a specific area in the dislocation dynamics domain.

8 ACKNOWLEDGMENTS

REFERENCES

[1] Nadja Adamovic, Jesper Friis, Gerhard Goldbeck, Adham Hashibon, Kersti Hermansson, Denka Hristova-Bogaerds, Rudy Koopmans, and Erich Wimmer. 2020. The EMMC Roadmap for Materials Modelling and Digitalisation of the Materials Sciences. (Nov 2020). https://doi.org/10.5281/zenodo.4272033
[2] Athanasios Arsenlis, Wei Cai, Meijie Tang, Moono Rhee, Tomas Oppelstrup, Gregg Hommes, Tom G Pierce, and Vasily V Bulatov. 2007. Enabling strain hardening simulations with dislocation dynamics. *Modelling and Simulation in Materials Science and Engineering* 15, 6 (2007), 553.
[3] Toshihiro Ashino. 2010. Materials ontology: An infrastructure for exchanging materials information and knowledge. *Data Science Journal* 9 (2010), 54–61.
[4] Diego Berrueta, Sergio Fernández, and Iván Frade. 2008. Cooking HTTP content negotiation with Vapour. In *Proceedings of 4th Workshop on Scripting for the Semantic Web (SFSW2008)*, Vol. 72. Citeseer.
[5] Carl Boettiger. 2018. *rdflib: A high level wrapper around the redland package for common rdf applications.* https://doi.org/10.5281/zenodo.1098478
[6] Benoit Devincre, Ronan Madec, Ghiath Monnet, Sylvain Queyreau, Riccardo Gatti, and Ladislas Kubin. 2011. Modeling crystal plasticity with dislocation dynamics simulations: The 'microMegas' code. *Mechanics of Nano-objects* 1 (2011), 81–100.
[7] Said Fathalla, Sören Auer, and Christoph Lange. 2020. Towards the semantic formalization of science. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing.* 2057–2059.
[8] Said Fathalla, Christoph Lange, and Sören Auer. 2019. EVENTSKG: A 5-star dataset of top-ranked events in eight computer science communities. In *European semantic web conference.* Springer, 427–442.
[9] Said Fathalla, Sahar Vahdati, Sören Auer, and Christoph Lange. 2018. SemSur: A Core Ontology for the Semantic Representation of Research Findings. In *Proceedings of the 14th International Conference on Semantic Systems, SEMANTICS 2018, Vienna, Austria, September 10-13, 2018* (2018-01-01) *(Procedia Computer Science)*, Vol. 137. Elsevier, 151–162. https://doi.org/10.1016/j.procs.2018.09.015
[10] Said Fathalla, Sahar Vahdati, Christoph Lange, and Sören Auer. 2019. SEO: A Scientific Events Data Model. In *The Semantic Web – ISWC 2019*, Chiara Ghidini, Olaf Hartig, Maria Maleshkova, Vojtěch Svátek, Isabel Cruz, Aidan Hogan, Jie Song, Maxime Lefrançois, and Fabien Gandon (Eds.). Springer International Publishing, Cham, 79–95.
[11] Mariano Fernández-López, María Poveda-Villalón, Mari Carmen Suárez-Figueroa, and Asunción Gómez-Pérez. 2019. Why are ontologies not reused across the same domain? *Journal of Web Semantics* 57 (2019), 100492.
[12] Nasr M Ghoniem and LZ Sun. 1999. Fast-sum method for the elastic field of three-dimensional dislocation ensembles. *Physical Review B* 60, 1 (1999), 128.

[13] Sydney R Hall and Brian McMahon. 2016. The implementation and evolution of STAR/CIF ontologies: Interoperability and preservation of structured data. *Data Science Journal* 15 (2016).
[14] Ralph Hodgson, Paul J Keller, Jack Hodges, and Jack Spivak. 2014. Qudt-quantities, units, dimensions and data types ontologies. *USA Available http://qudt. org March* 156 (2014).
[15] Ahmad Zainul Ihsan, Danilo Dessi, Mehwish Alam, Harald Sack, and Stefan Sandfeld. 2021. Steps towards a dislocation ontology for crystalline materials. In *2nd International Workshop on Semantic Digital Twins, SeDiT 2021*, Vol. 2887. CEUR-WS.
[16] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin a. Persson. 2013. The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* 1, 1 (2013), 011002. https://doi.org/10.1063/1.4812323
[17] Julian Kimmig, Stefan Zechel, and Ulrich S. Schubert. 2021. Digital Transformation in Materials Science: A Paradigm Change in Material's Development. *Advanced Materials* 33, 8 (2021), 2004940. https://doi.org/10.1002/adma.202004940 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/adma.202004940
[18] Konstantinos I Kotis, George A Vouros, and Dimitris Spiliotopoulos. 2020. Ontology engineering methodologies for the evolution of living and reused ontologies: status, trends, findings and recommendations. *The Knowledge Engineering Review* 35 (2020).
[19] Rhys Lewis. 2007. Dereferencing http uris. *Draft Tag Finding, http://www.w3.org/ 2001/tag/doc/httpRange-14/2007-05-31/HttpRange-14.html* (2007).
[20] Huanyu Li, Rickard Armiento, and Patrick Lambrix. 2020. An Ontology for the Materials Design Domain. In *International Semantic Web Conference.* Springer, 212–227.
[21] Egon Orowan. 1934. Zur Kristallplastizität. *Zeitschrift für Physik* 89, 9-10 (1934), 605–613.
[22] Giacomo Po and Nasr Ghoniem. 2014. A variational formulation of constrained dislocation dynamics coupled with heat and vacancy diffusion. *Journal of the Mechanics and Physics of Solids* 66 (2014), 103–116. https://doi.org/10.1016/j.jmps.2014.01.012
[23] Michael Polanyi. 1934. Über eine Art Gitterstörung, die einen Kristall plastisch machen könnte. *Zeitschrift für Physik* 89, 9-10 (1934), 660–664.
[24] A. Prakash and S. Sandfeld. 2018. Chances and Challenges in Fusing Data Science with Materials Science. *Practical Metallography* 55, 8 (2018), 493–514. https://doi.org/10.3139/147.110539
[25] Henri Salmenjoki, Mikko J Alava, and Lasse Laurson. 2018. Machine learning plastic deformation of crystals. *Nature communications* 9, 1 (2018), 1–7.
[26] Stefan Sandfeld and Giacomo Po. 2015. Microstructural comparison of the kinematics of discrete and continuum dislocations models. *Modelling and Simulation in Materials Science and Engineering* 23, 8 (2015), 085003.
[27] Aysegul Say, Said Fathalla, Sahar Vahdati, Jens Lehmann, and Sören Auer. 2020. Semantic representation of physics research data. In *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management Vol. 2.* Setúbal, Portugal: Science and Technology Publications, Lda.
[28] Zeynep Say, Said Fathalla, Sahar Vahdati, Jens Lehmann, and Sören Auer. 2020. Ontology design for pharmaceutical research outcomes. In *International Conference on Theory and Practice of Digital Libraries.* Springer, 119–132.
[29] Elena Simperl, Cristina Sarasua, Rachanee Ungrangsi, and Tobias Bürger. 2011. Ontology metadata for ontology reuse. *International Journal of Metadata, Semantics and Ontologies* 6, 2 (2011), 126–145. https://doi.org/10.1504/IJMSO.2011. 046579
[30] Stian Soiland-Reyes, Peter Sefton, Mercè Crosas, Leyla Jael Castro, Frederik Coppens, José M Fernández, Daniel Garijo, Björn Grüning, Marco La Rosa, Simone Leo, et al. 2021. Packaging research artefacts with RO-Crate. *Data Science* Preprint (2021), 1–42.
[31] Nick Spadaccini, Ian R Castleden, Doug du Boulay, and Sydney R Hall. 2012. dREL: a relational expression language for dictionary methods. *Journal of chemical information and modeling* 52, 8 (2012), 1917–1925.
[32] Dominik Steinberger, Hengxu Song, and Stefan Sandfeld. 2019. Machine Learning-Based Classification of Dislocation Microstructures. *Frontiers in Materials* 6 (2019), 141. https://doi.org/10.3389/fmats.2019.00141
[33] Samir Tartir, I Budak Arpinar, Michael Moore, Amit P Sheth, and Boanerges Aleman-Meza. 2005. OntoQA: Metric-based ontology quality analysis. (2005).
[34] Geoffrey Ingram Taylor. 1934. The mechanism of plastic deformation of crystals. Part I.–Theoretical. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 145, 855 (1934), 362–387.
[35] Paul E van der Vet, Piet-Hein Speel, and Nicolaas JI Mars. 1994. The Plinius ontology of ceramic materials. In *Eleventh European Conference on Artificial Intelligence (ECAI'94) Workshop on Comparison of Implemented Ontologies.* 8–12.
[36] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 3, 1 (2016), 1–9.