A Concise Ontological Model of Properties in the Quantum Cascade Laser Domain

Deperias Kerre [a,b,*], Anne Laurent [b], Kenneth Maussang [c] and Dickson Owuor [a]
[a] *SCES, Strathmore University, Nairobi, Kenya*
E-mails: dkerre@strathmore.edu, dowuor@strathmore.edu
[b] *LIRMM, Univ Montpellier, CNRS, Montpellier, France*
E-mail: anne.laurent@umontpellier.fr
[c] *IES, Univ Montpellier, CNRS, Montpellier, France*
E-mail: kenneth.maussang@umontpellier.fr

Abstract. Terahertz quantum cascade lasers are semiconductor laser devices that operate in the the far infrared (in the frequency range from about 100GHz to 10THz). Information regarding the quantum cascade laser (QCL) design is quite crucial in understanding the various laser designs and their implication on the laser performance. Maintaining knowledge bases or ontologies with this information is therefore useful in supporting data mining activities that seek to retrieve useful information on the various quantum cascade laser designs and their respective performance. The ontologies and knowledge bases can also be used to generate Knowledge Graphs (KGs) that can support queries on QCL designs and performance. Most of the existing ontologies and knowledge bases in the material design domain do not capture this crucial information. In this paper, we present a semantically enriched ontological model of properties in the quantum cascade laser domain. The properties of interest include the design of the laser (Heterostructure), working mode of the laser and the corresponding opto-electronic characteristics. We evaluate the ability of ontological representation to model the quantum cascade laser properties using properties from sample scientific articles documenting the various QCL designs and their properties.

Keywords: Knowledge Bases, Knowledge Graphs, Material Design, Ontologies, Semiconductor Devices

1. Introduction

The QCL semiconductor lasers have been utilised in various applications for instance, in screening various types of abnormal tissues [1] and in configuring high speed networks in the electronics field [2]. An efficient design process of these devices is therefore highly desired in order to maximize their application potential.

The quantum cascade laser design structure is made up of complex hetero-structures. In most cases, the properties of the laser are defined by the growth sheet which gives information on the heterostructure thickness, the materials combined and their respective combination order. The QCL properties can be broadly categorized into two: Design which includes the Hetero-structure i.e the material design properties capturing the various material combination used in constructing the semiconductor laser device together with the specification of the layer sequence and secondly the opto-electronic characteristics i.e the laser performance behaviour as a result of injecting current into the laser for instance the working temperature, power, lasing frequency etc. Some of the laser properties are dependent on other properties and the working mode of the laser. For instance, the semiconductor laser device working temperature may vary based on whether the device is working at continuous or pulse mode.

---

[*] Corresponding author. E-mail: dkerre@strathmore.edu.

Structured information capturing the various QCL laser device designs and their corresponding performance characteristics is very crucial in deciphering the complex structure of the laser. This is useful for instance in understanding the laser structure in relation to its performance. Information on the QCL laser device properties exist in varied sources. Well structured knowledge on semiconductor laser designs and performance is important in optimizing the design process of the lasers as there is availability of answers on various QCL laser device design queries. This will also address issues related to FAIR principles (Findable, Accessible, Interoperable, and Reusable), which will enable automatic sharing and use of data in the quantum cascade laser domain by machines and humans [3]. Further, materials data access, acquisition, representation and sharing are also identified as critical tasks for the materials science community [4, 5].

The existing ontologies and Knowledge bases could not be readily instantiated and used to capture these information. This is also attributed to the complex nature of the relationships between the various quantum cascade laser device properties for instance, between the working modes, properties and designs. There is therefore need for an enriched, formal representation of the QCL device properties that capture the physical properties and the various designs in terms of the hetero-structure.

In this paper, we present a semantically enriched ontological model of semiconductor laser properties in the quantum cascade laser domain. The main focus of the ontology is to formalize the relationship between laser designs and the performance characteristics. We also validate the consistency of the ontological representation with a logical reasoner and sample data mined from scientific articles using a text mining pipeline for QCL properties proposed in [6]. The main contributions of this paper are therefore as follows: (i) A comprehensive review of the state of the art on ontologies and vocabularies in material design, in relation to our domain of interest, (ii) A semantically enriched ontological modelling of properties in the quantum cascade laser domain and (iii) A comprehensive evaluation of the ontology based on a data driven approach.

The rest of the paper is organized as follows: we briefly explain the motivation scenario for the ontological model in section 2, an overview of ontologies and standards in material design in relation to the quantum cascade laser domain in section 3, the development approach of the qcl properties ontological representation, the concepts, relations and the axiomatization in section 4, the evaluation approach in section 5,results in section 6, the the technical specifications of the ontology in section 7 and lastly conclude in Section 8.

2. Motivation Scenario

The motivation for an ontological model of properties in the semiconductor lasers domain arises from the need of semiconductor fabrication with desired performance characteristics. We present a scenario where a semiconductor laser engineer intends to fabricate a heterostructure with desired optoelectronic properties such as working temperature and optical power for an optimized operation. Also for the existing semiconductor laser devices, a design expert may want to quickly understand the relationship between the various design and performance parameters in order to get insights for future semiconductor fabrication processes. The design expect may also be interested in understanding trends in semiconductor laser fabrication over a given period of time. This creates the need for valid references to various sources documenting the laser device design and performance characteristics. In the process of undertaking these tasks, the following issues emerge:

i The semiconductor laser design and optoelectronic characteristics data exists in dispersed sources such as lab notebooks, manuals and scientific articles reporting the various semiconductor devices.
ii Decisions regarding designing of semiconductor laser heterostructures with target properties are usually carried on experimental basis involving manual analysis of experimental data which takes time hence delaying the design process.

With these issues in mind, there is therefore the need for a solution that enables a structured representation of design and optoelectronic characteristics for the semiconductor laser domain and that also provide provenance information for the various properties. This will serve to provide a platform for organization of experimental data from various sources together with respective links to the sources. This will also provide a standard way of exploring the data and understand the inherent relationships in a quicker way hence enhancing efficiency in the semiconductor laser fabrication process. Semantic enrichment will also provide links to other related data such as formal definitions of properties and their corresponding units on the web. Lastly, the formal representation can also be used to create data models for Knowledge graphs which can be used to represent massive experimental data to allow a quicker analysis of the same.

3. Related Work

3.1. Introduction

In this section, we give a detailed overview of ontologies and standards in materials science. The guiding intentions in the analysis are as follows: (i) To analyse the ability of the ontologies to represent the relationship between design and performance properties in the quantum cascade laser domain and (ii) The ability of the ontologies to be instantiated with sample data on quantum cascade semiconductor laser properties and provide answers to queries regarding the laser properties. In order to achieve this, we use search services such as MatPortal[1], BioPortal[2], Linked Open Vocabularies[3] and search engines such as Google.

3.2. Ontologies and Standards in Material Science

In the materials science domain, there is progress in the use of semantic technologies for various applications such as representation of complex domain knowledge. This enables sharing and utilization of complex information in an open and agreeable way by both machines and humans. Ontologies are one of the technologies being widely adopted with the focus being on representing specific sub-domain and general material domain knowledge.

A couple of general ontologies that represent general materials domain Knowledge have been proposed. These includes Chemical Entities of Biological Interest (ChEBI) [7], a freely available dataset of molecular entities e.g atom, molecular ion, etc. , Basic Formal Ontology(BFO) [8], Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [9], General Formal Ontology(GFO) [10] and the Elementary Multiperspective Material Ontology(EMMO) [11].

Other ontologies have also been developed with specific domains or particular interests in mind. The application scenarios ranges from giving a general representation of concepts in a domain of interest to activities such standardization of data curation and sharing in the material design databases. We present examples of the ontologies in table 1.

MatOnto ontology [12] , based on DOLCE, is used for representing oxygen ion conducting materials for the fuel cell domain, Materials Ontology [13] for data exchange among thermal property databases and MDO ontology [14] for materials design field, representing the domain knowledge specifically related to solid-state physics and computational materials science. An ontology for a Polymer Nanocomposite Community Data Resource is also proposed in [15]. The NanoParticle Ontology [16], based on BFO, gives a presentation of nanoparticles properties with the aim of designing new nanoparticles while eNanoMapper Ontology [17] gives an assessment of risks related to the use of nanoparticles. Also, an ontology for design pattern for modelling material transformation for the sustainable construction domain is presented in [24] and an ontology for representing knowledge on simulation, modelling and optimization in molecular engineering sciences is presented in [25]. MatML [26] is an extensible mark up language for exchaning materials information. MatOWL [18], based on MatXML Schema is used to facilitate ontology-based data access. The MMOY ontology [19] is used to represent materials knowledge from Yago [27], a knowledge base capturing many topics including material properties. The Dislocation Ontology [20] reuses some concepts from MDO to represent knowledge on crystalline materials. There are also platforms which functions as a prototype to describe materials science experiments, for instance, the MaterialDigital Ontology [21]. The Materials and

---

[1] https://matportal.org/

[2] https://bioportal.bioontology.org/

[3] https://lov.linkeddata.es/dataset/lov/

Table 1

Some Ontologies in Material Science

Ontologies	Ontology Metrics	Language	Domain	Intended Application
MatOnto [12]	606 Concepts, 31 relations, 488 instances	OWL	Materials	Materials Discovery
Materials Ontology [13]	78 Concepts, 10 relations, 24 instances	OWL	Crystals	Semantic Querying
MDO [14]	37 Concepts, 64 relations	OWL	Materials Design	Semantic Querying over multiple databases
Polymer Nanocomposite [15]	Being expanded asp per use cases	OWL	Polymer Nanocomposite	Knowledge Representation
NanoParticle Ontology [16]	1904 Concepts, 81 relations	OWL	Nanotechnology	Data Integration, Search
eNanoMapper Ontology [17]	12781 Concepts, 5 relations, 464 instances	OWL	Nanotechnology	Data Integration
MatOWL [18]	(not available)	OWL	Materials	Semantic Querying
MMOY Ontology [19]	2325 Concepts, 9 relations, 1738 instances	OWL	Metals	Knowledge Extraction
Dislocation Ontology [20]	18 Concepts, 16 relations	OWL	Crystalline Materials	Knowledge Representation
MaterialDigital Ontology [21]	13 Concepts, 7 relations	OWL	Material Experiments	Knowledge Representation, Data Curation
MAMBO Ontology [22]	26 Concepts, 33 relations	OWL	Molecular-based Materials	Knowledge Representation
ELSSI-EMD Ontology [23]	35 Concepts, 37 relations,33 instances	OWL	Molecular-based Materials	Knowledge Representation

Molecules Basic Ontology (MAMBO) [22], integrates EMMO, CheBI and MDO to represent concepts and relations emerging on materials with a focus on the relationships between molecular aggregation and properties of the system and lastly the ELSSI-EMD ontology [23], provides guidelines for material testing standardization.

One of the key issues that arise is whether the existing ontologies can be able to model/represent the complex relationships between the design and performance characteristics in our domain of interest. The existing ontologies cannot be readily used to present a formal representation of properties in the quantum cascade laser domain due to some reasons: Some of the ontologies give a more general formal representation while others are restricted to specific domain concepts that do not fit in our scope. To mention a few, concepts that capture the laser design types, working mode cannot be readily captured. The focus of this work is more on the representation of “wafer fabrication” or hetero-structure properties, which is a critical step in the semiconductor laser device development and the relation to the performance characteristics of the laser devices.

4. Ontological Modelling of the Quantum Cascade Laser Properties

In this section, we give an overview of the development methodology for the QCL properties ontological representation and the description of the ontological representation

4.1. Development of the Quantum Cascade Laser Properties Ontological Representation

In the development phase of the QCL properties ontological model, we adopt the NeOn ontology engineering methodology [28]. This methodology consists of a list of scenarios mapped from a set of common ontology development activities in the ontology engineering life cycle. The set of scenarios capture most of the ontology development activities suitable for our target domain. We particularly focus on applying scenario i (From Specification to Implementation), Scenario iii (Reusing Ontological resources), Scenario iv (Reusing and re-engineering ontological resources) and Scenario viii (Restructuring ontological resources).

We use the Web Ontology Language (OWL 2) [4] as the representation language of the ontological representation using the RDF/XML syntax. The choice of OWL 2 DL was critical so as to enable reasoning and consistency checks

---

[4] https://www.w3.org/TR/owl2-overview/

on the ontological representation using the available standard reasoners. We utilise two tools in the development of the ontology: Repairing Ontological Structure Environment (RepOSE, [29]) which allows ontology debugging and proposal of additional knowledge to the ontology and OntOlogy Pitfall Scanner [30], for detecting some of the common pitfalls encountered during ontology development. Throughout the development process, input from domain experts in semiconductor heterostructure laser fabrication and knowledge engineers is considered. In the remaining part of this section, we detail the key aspects in the development process of the QCL properties ontological model.

4.1.1. Requirements Analysis

In this step, we clarify the requirements of the ontological representation in relation to scenario i of the Neon methodology for ontology engineering. This involves proposing use cases(UC), Competency questions (CQ) and additional restrictions on the knowledge representation schema.

The use cases for our proposed QCLontological model are identified through literature and from discussions with domain experts in the field of quantum cascade lasers and are as follows:

i UC1: The ontology model will be used to represent knowledge about the various QCL designs (in form of the hetero-structure/material design) and the optoelectronic characteristics (such as output power, working temperature, lasing frequency) based on the various QCL designs.
ii UC2: The ontology model will be used for representing in addition to the QCL designs and properties, the various working modes at which the properties are achieved based on the designs.
iii UC3: The ontology model will be used to maintain provenance information about the various QCL designs and performance characteristics. This will enable tracking of data on QCL development with information such as the developers/ authors, year in which the design was proposed and useful, permanent links to the resources such as the DOI.

The competency questions are also agreed upon based on discussions with domain experts. We propose a list of ten competency questions for the qcl ontology model. The questions are as follows:

i CQ1: What are the material composition and sequence layer of a heterostructure with a particular design type?
ii CQ2: For a particular design type, what are the possible layer sequences and material composition?
iii CQ3: What is the material composition of a particular hetero-structure with a particular sequence layer?
iv CQ4: What are the resultant performance characteristics of a QCL laser working in a particular working mode?
v CQ5: For a particular heterostructure(as described by the sequence and material composition), what are the resultant laser performance characteristics?
vi CQ6: For a particular laser performance characteristics, what are possible design properties (i.e layer sequence and material composition)?
vii CQ7: What is the sequence layer of a heterostructure with a particular material composition?
viii CQ8: Who are the authors of the particular laser device having certain properties?
ix CQ9: When was the laser device proposed?
x CQ10: Where is the information published?

We also provide a list of additional restrictions to help define the concepts as below:
-A hetero-structure corresponds to one particular design type.
-A property can relate to the working mode of the laser.
-A property corresponds to only one laser working mode.
-A property can also relate to the hetero-structure.
-A Layer sequence corresponds to one material combination formula.
The full list of additional restrictions can be found at the GitHub repository [5].

---

[5] https://github.com/DeperiasKerre/qcl_Onto

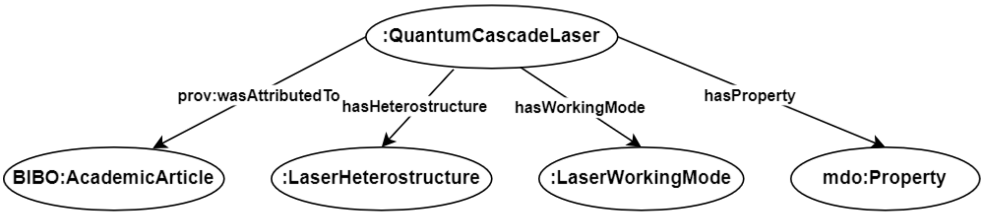


Fig. 1. Upper Concepts in the Ontology.

4.1.2. Modular Development for Building Design Patterns

We adopt a pattern related to provenance information in the repository of Ontology Design Patterns that leads to the re-use of entities from the PROV-O ontology. The ontology is also developed in a modular way where the development is based on the categories of information to be represented i.e the design, provenance, working mode and performance properties.

4.1.3. Reusing, Reconstructing and Re-engineering Ontological Resources

Our proposed ontological representation reuses some terms and concepts from well established ontologies such as EMMO by reusing the concept ‘Material’ and CheBI by reusing the term ‘Atom’. We also reuse the concepts of ‘Agent’ the PROV-O ontology to represent provenance information [31] and the term ‘ Property’ from the MDO ontology in order to represent information on the various qcl properties such as working temperature, power etc. In order to represent the units, we reuse the terms ‘Quantity’,‘Quantity Value’,‘QuantityKind’ and ‘Unit’ from the QUDT (Quantities, Units, Dimensions and Data Types Ontologies) [32]. We use the term ‘AcademicArticle’ from the BIBO vocabulary[6] to represent an academic journal documenting the qcl properties. We also use the metadata terms from the Dublin Core Metadata Initiative (DCMI)[7] to represent the metadata of the ontological representation.

4.2. Ontology Concepts and Relations

The ontological representation of properties in the quantum cascade laser domain contains a total of 15 concepts, 23 relations and 11 instances. The information captured by the ontological representation can be categorized as follows: **Design information** (to capture the heterostructural design information of the laser), **Properties** (to capture the laser performance/optoelectronic characteristics), **Working mode** (to capture the nature of laser beam emission) and **Provenance information** to provide references to sources containing the laser information. The upper concepts in the ontology are therefore the **Quantum Cascade laser**, **LaserHeterostructure**, **WorkingMode**, **Property** and **AcademicArticle**.

The quantum cascade laser concept represents the qcl semiconductor laser device. The concept AcademicArticle is used to represent provenance information about the quantum cascade laser designs and performance properties. The LaserHeterostructure defines the laser design in form of the material stacking and sequencing characteristics. We denote the description logic axioms for the upper concepts as U. A *quantum cascade laser* has a *laser heterostructure* (U1), a *Property* (U2), a *working mode* (U3) and it is attributed to an *AcademicArticle* (U4) which gives link to resources documenting the laser. Figure 1 shows an overview of the concepts and relationships in the for the upper concepts in the ontology and figure 2 shows the corresponding description logic axioms.

4.2.1. Laser Optoelectronic Characteristics

In order to represent knowledge on the quantum cascade laser performance/optoelecronic properties, we use the following concepts: **Property** (from MDO ontology), **Quantity**, **QuantityValue**, **QuantityKind** and **Unit** (from QUDT ontology). The description logic axioms are denoted as P. The concept *Property*, which is viewed as a quantifiable aspect of a material system, is used to represent the optoelectronic characteristics of the laser and its defined as a sub concept of *Quantity* (P1). *Quantity* represents the measurement of an observable property and it has the *QuantityValue* and *QuantityKind*. *Quantity value* represents the value measured (in form of a unit and a

---

[6] https://www.dublincore.org/specifications/bibo/

[7] http://purl.org/dc/terms/

(U1) QuantumCascadeLaser ⊆ =1 hasHeterostructure.LaserHeterostructure

(U2) QuantumCascadeLaser ⊆ ∀ hasProperty.Property

(U3) QuantumCascadeLaser ⊆ =1 hasWorkingMode.LaserWorkingMode

(U4)) QuantumCascadeLaser ⊆ wasAttributedTo.AcademicArticle

Fig. 2. Description Logic Axioms for the Upper Concepts in the Ontology

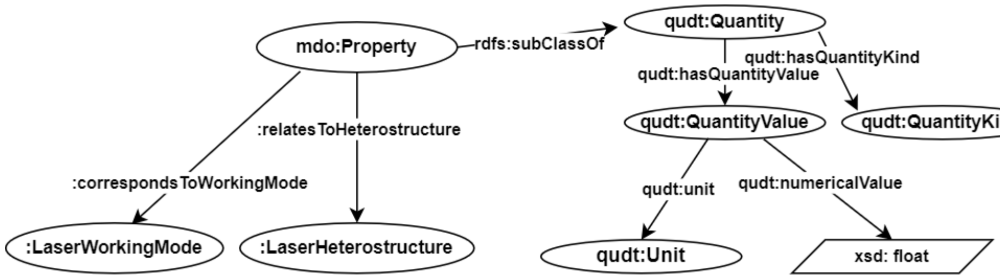


Fig. 3. Ontology Section for Quantum Cascade Laser Optoelectronic Characteristics.

(P1) Property ⊆ Quantity

(P2) Property ⊆ ∀ relatedToHeterostructure.LaserHeterostructure

(P3) Property ⊆ = 1 correspondsToWorkingMode.LaserWorkingMode

Fig. 4. Description Logic Axioms for the Laser Physical Properties

numerical value) and the *QuantityKind* represents any observable property that can be measured and quantified numerically, for instance temperature, power etc. More information about these relations can be found at the QUDT ontology [8]. The definition of *Property* as a sub concept of *Quantity* enables representation of information such as values, kinds and units of the quantum cascade laser properties. A property has a name in a string using the *PropertyName* data property. Properties are related to the *laser heterostructures* (P2) and corresponds to a *working mode* (P3). Figure 3 shows a detailed view of the ontology section describing quantum cascade laser performance properties and figure 4 shows the description logic axioms.

4.2.2. Laser Heterostructure/Design Properties

The quantum cascade laser design information is captured using the following concepts: **LaserHeterostructure**, **DesignType**, **MaterialComposition**, **Materials** and **LayerSequence**. We denote the laser design description logic axioms as D. The laser heterostructure represents the laser layer design comprising of the various semiconductor materials. A *laser heterostructure* has a *design type* (D1), *material composition* (D2), and the *layer sequence* (D3). The *design type* of the laser refers to the geometrical arrangement of materials in the laser design while the *material composition* represents the various materials included in the heterostructure and their respective ratio of combination. *Material composition* consist of *materials* (D4) and the *materials* are composed of *atoms* (D5). The *layer sequence* is based on *materials* (D6) and it has a *unit* (D7). Additionally, the layer sequence has the *sequence* in a string and materials also has a *matFormula* in a string which captures the chemical elements and their ratio of combination. The laser design type has two instances i.e BoundToContinum and LOPhonon depopulation describing the

---

[8] https://qudt.org/

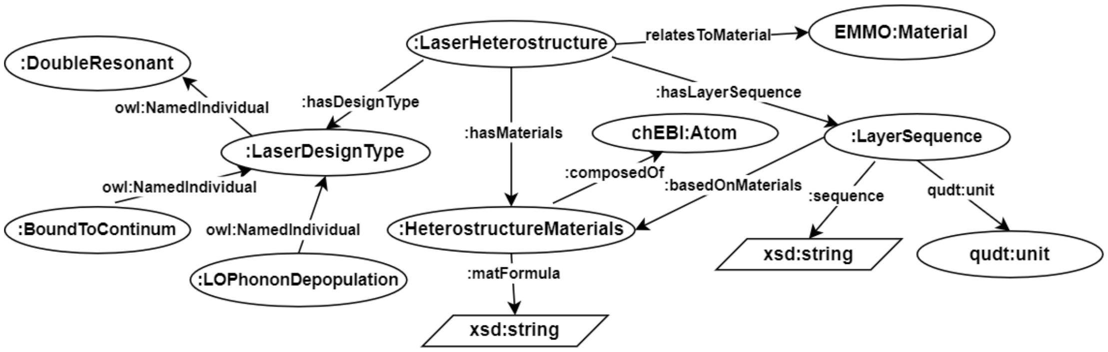


Fig. 5. Ontology Section for the Heterostructure Design Properties.

(D1) LaserHeterostructure $\subseteq =1$ hasDesignType.DesignType

(D2) LaserHeterostructure $\subseteq =1$ hasMaterialComposition.MaterialComposition

(D3) LaserHeterostructure $\subseteq =1$ hasLayerSequence.LayerSequence

(D4) MaterialComposition $\subseteq \forall$ consistOf.Materials

(D5) Materials $\subseteq \forall$ composedOf.Atoms

(D6) LayerSequence $\subseteq \forall$ basedOnMaterials.Materials

(D7) LayerSequence $\subseteq =1$ unit.Unit

Fig. 6. Description Logic Axioms for the Laser Heterostructure Design Properties.

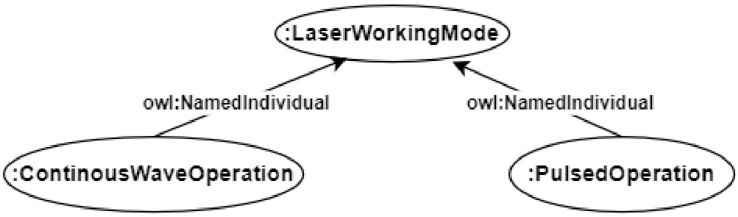


Fig. 7. Ontological Representation for the Quantum Cascade Laser Working Mode.

laser design types. Figure 5 shows the ontology section describing the quantum cascade laser design and figure 6 shows the description logic axioms for this description.

4.2.3. Laser Working Mode

The laser working mode information is represented using the *LaserWorkingMode* concept. This refers to the mode in which the laser beam is released i.e it can either be in continuous or pulse mode as indicated by the instances of the laser working mode class. In continuous mode, the laser beams are emitted continuously while in pulsed mode in pulses. The semiconductor laser working temperature varies depending on the working mode under which the laser beams are emitted. Figure 7shows the ontological representation of the laser working mode.

4.2.4. Provenance Information

For provenance information, we use the terms **Agent** and **ArcademicArtle** and denote the description logic axioms as PR. An Agent from the PROV-O ontology represents something that bears some form of responsibility

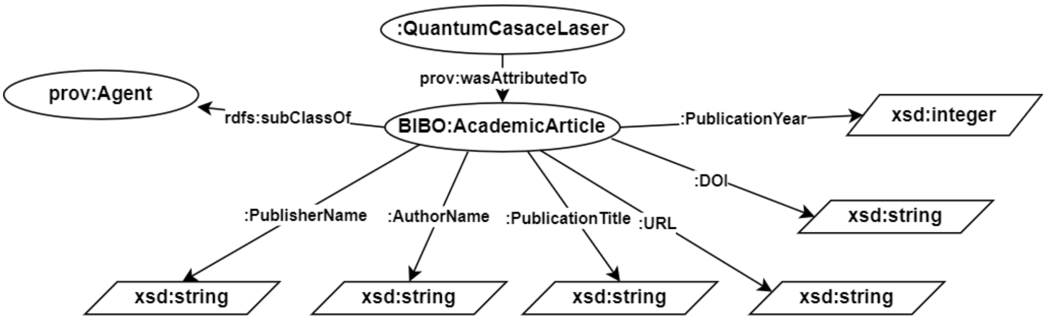


Fig. 8. Ontology Section for Provenance Information.

$$(PR1)\ AcademicArticle \subseteq Agent$$

Fig. 9. Description Logic Axioms for Provenance Information.

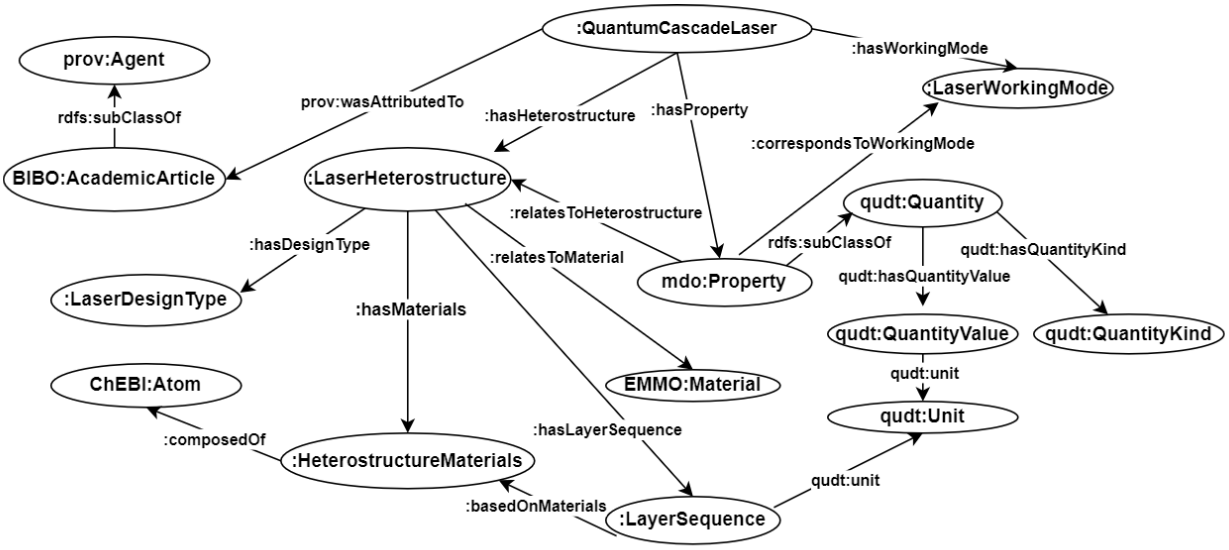


Fig. 10. Overview of the Quantum Cascade Laser Properties Ontological Representation.

for an activity taking place, for the existence of an entity, or for another agent's activity[9]. The ArcademicArticle concept from the BIBO ontology refers to a peer reviewed article documenting the quantum cascade laser properties. An ArcademicArticle is set as as sub concept of the Agent (P1) and quantum cascade laser device is attributed to an ArcademicArticle as seen before. An ArcademicArticle has the PublisherName, URL, DOI, PublicationTitle, AuthorName as a string and the PublicationYear as an integer. This information enables tracking of information sources on the various semiconductor laser properties. Figure 8 shows the ontology section for laser provenance information and figure 9 shows the respective description logic axioms.

Figure 10 shows an overview of the entire ontological representation of properties in the quantum cascade laser domain with all the concepts interconnected together to model the complete relationship between the laser design and performance properties.

---

[9] https://www.w3.org/TR/prov-o/

5. Ontology Evaluation

The evaluation of the ontological representation is done based on i. Consistency of the ontology and ii. Completeness and Correctness i.e evaluating the success of the ontology in modelling the domain of interest (Formative Evaluation) and iii. The richness of the ontology based on an ontology quality evaluation metric.

5.1. Consistency

In order to evaluate the consistency of the ontological modelling, we use the following ontology checking tools: Pitfall scanner [30] and the HermiT Reasoner [33] embedded in protege software[10]. Any inconsistencies detected by the HermiT reasoner were identified and resolved. For the OOPS! Scanner, critical and important pitfalls were also considered and rectified.

5.2. Formative Evaluation

5.2.1. Experimental Setup

In order to design our experiments, we use data composing of sample semiconductor quantum cascade laser design and optoelectronic properties. The properties are the laser heterostructure (material composition and sequence), working temperature, lasing frequency, optical power, laser working mode, and laser design type together with the respective units. We also include data on provenance such as article DOI and publication URL. The individual properties are mined from a sample of scientific articles using a text mining tool proposed in [6], except for the laser design type, working mode and URL which are included using the human in the loop approach. This constitutes a total of 181 quantum cascade laser property instances documenting over 12 quantum cascade laser devices. The units are linked to specific URIs in the QUDT ontology to provide a reference to their description. For data preprocessing, we semantically enrich the data with URIs to the respective resources describing the data elements for instance the working mode, laser design type and the units. Table 2 shows the statistics of the test data. We define data mapping rules for mapping the data to the ontology schema using the RDFlib library[11]. We generate a sample knowledge graph(KG) containing a total of 831 triples, on which we run the validation scripts in form of Competency questions as detailed in Ontology Verification and Validation in section 5.2.2. Figure 11 shows a visualization of a section of the sample KG generated.

Table 2

Summary Statistics for the Test Data

Instance Type	**Number**
Heterostructure	15
Design Type	13
Working Temperature	18
Frequency	13
Layer Sequence	15
DOI	15
URL	15
Working Mode	18
Laser Power	9
Units	50
Total	**181**

---

[10] https://protege.stanford.edu/

[11] https://github.com/RDFLib/rdflib

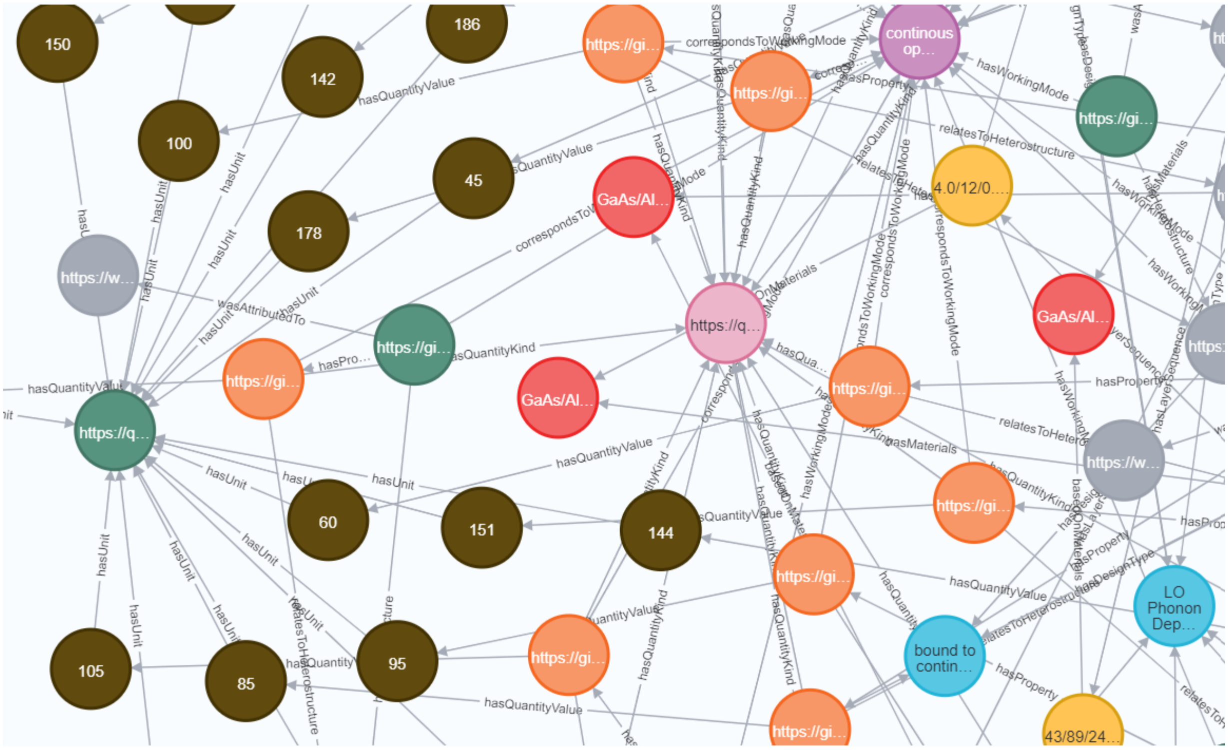


Fig. 11. Sample Knowledge Graph for the QCL Properties

5.2.2. Ontology Verification and Validation

The evaluation steps in this phase are carried out based on data driven evaluation principles proposed in [34]. Verification entails ensuring that the ontology is built correctly while ontology validation (using test cases) aims at ensuring that the ontology fulfills its intended purpose. A set of competency questions(CQs) are defined from the use case scenarios defined in section 4. The CQs are set such that they capture all the information represented in the use cases and are therefore used as functional requirement specification for the ontology. The verification and Validation is performed with the sole aim of ensuring that ontology fully conforms to these requirements and should therefore be able to answer all the CQs correctly. This entails an evaluation of the completeness of the ontology in capturing the domain of interest. The CQs are represented in SPARQL, the formal RDF query language. Table 3 shows the general classes of CQs used in this paper. W, X, Y and Z are place holders for any suitable values for the laser design and performance properties. For the CQs classes in table 3, we run several possibilities of queries capturing the various combination of information needed by the CQs on the knowledge graph and evaluate the output compared to actual values with the data intances.

Table 3

Competence Questions (CQs)

ID	Question Text
CQ1	What is the material composition/sequence layer of a heterostructure with a design type X ?
CQ2	For a particular heterostructure X, what are the possible design type?
CQ3	What is the layer sequence of a heterostructure with a material composition X ?
CQ4	What is the performance property X of a laser working in mode Y ?
CQ5	What is the performance property X of a laser having a heterostructure with material composition Y or layer sequence Z?
CQ6	For a particular performance property X, What are the corresponding heterosturcture designs?
CQ7	What is the DOI and/ the URL of the scientific article documenting a laser with performance property W or with heterostructure X or working mode Y or design type Z.

5.3. Metric Based Evaluation

In this evaluation phase, we assess the richness/quality of the ontology schema by using the evaluation mode adopted in [35]. This approach evaluates the quality of the ontology based on schema and instance metrics. These metrics provide varied information for assessing the various richness within an ontology. In this study, we adopt the *Inheritance Richness(IR)* assessment metric to assess the quality of the ontology model. This metric describes the distribution of information across different levels of the ontology inheritance tree. Formally, the IR metric for a class C_i is defined as the average number of subclasses per class in the a given ontology subtree which is determined as $|H^C(C_1, C_i)|$. The IR for the entire ontology schema is therefore defined as follows:

$$IR = \frac{\sum_{C_i \in C'} |H^C(C_1, C_i)|}{|C|} \tag{1}$$

Where H is the number of inheritance relationships and C is the number of classes in the ontology. With this metric, we assess the extent of the ontology in covering the domain of interest at a detailed level. With regard to this evaluation metric, the lower the IR, the more detailed and specific an ontology is and vice versa.

6. Results and Discussions

6.1. Verification and Validation

For verification and validation, we run a total of 12 queries on the KG, for 12 possibles specific competency questions defined for the CQs specified in section 5.2.2. This results to a total of 144 records fetched. The competency questions for each of the CQ classes are as follows: CQ1 (2), CQ2 (1), CQ3(2), CQ4(2) and CQ5(3). The queries are designed to retrieve relevant information as per the defined CQs classes. The specific questions for the query classes are presented in table 4. We compare the queries' responses with the actual values in the data to determine the precision of query answering. All the 150 records returned by the queries match with the expected data values for the specific relations of interest hence resulting to a precision of 1. This demonstrates the ability of the ontology model in capturing the domain relationships.

Table 4

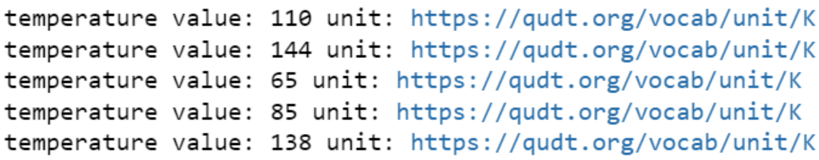
Competence Questions (CQs)

Competency Question ID	Competency Question
CQ1.1	What is the possible heterostructure material composition(s) of a semi-conductor laser with a bound to continum design type?
CQ1.2	What is the possible heterostructure layer sequence(s) of a semi-conductor laser with an LO Phonon design type?
CQ2.1	What is the design type for a heterostructure with material composition X?
CQ3.1	What is the possible layer sequence(s) of a heterostructure with material composition GaAs/Al0.15Ga0.85As?
CQ4.1	What are the operating temperatures of semiconductor laser devices working in a continous wave operation?
CQ4.2	What are the operating temperatures of semiconductor laser devices working in a pulse mode operation?
CQ5.1	What is the lasing frequency of a semiconductor laser with a material composition GaAs/Al0.15Ga0.85As?
CQ5.2	What is the output power of a semiconductor laser with a a layer sequence of 54/78/24/64/38/148/24/94 Å?
CQ6.1	What is the possible material composition for heterostructures with a lasing frequency greater than 3 THz?
CQ7.1	What are the DOIs and URLs of scientific articles documenting semi-conductor laser devices with a working temperature greater than 40K at continous wave mode?
CQ7.2	What are the DOIs and URLs of scientific articles documenting semi-conductor lasers with a material composition of GaAs/Al0.15Ga0.85As?
CQ7.3	What are the DOIs and URLs of scientific articles documenting semi-conductor lasers with An LO Phonon Design Type?

The CQS shows the ability of the ontology model to model various laser fabrication scenarios. For instance, its is possible to analyze various trends such as ranges of semiconductor laser working temperature in various working modes, possible layer sequences for certain heterostructure materials and their corresponding lasing capabilities such as the lasing frequency. It is also possible to access references to documents for specific semiconductor laser

```
    PREFIX QpOnto:<https://github.com/DeperiasKerre/qcl_Onto/blob/main/qclontology/version-1.0/qclonto.owl#>
    PREFIX qudt:<https://qudt.org/schema/qudt/>
    SELECT ?value ?unit
WHERE
{
 ?wt QpOnto:correspondsToWorkingMode QpOnto:PulsedOperation;
 <https://qudt.org/schema/qudt#hasQuantityValue> ?qv.
 ?qv qudt:numericValue ?value;
 qudt:hasUnit ?unit.
}
```

Fig. 12. Query for Competency Question 4.2



```
temperature value: 110 unit: https://qudt.org/vocab/unit/K
temperature value: 144 unit: https://qudt.org/vocab/unit/K
temperature value: 65 unit: https://qudt.org/vocab/unit/K
temperature value: 85 unit: https://qudt.org/vocab/unit/K
temperature value: 138 unit: https://qudt.org/vocab/unit/K
```

Fig. 13. Result for the Query for Competency Question 4.2

properties, for instance the DOI of a scientific document with specific properties. This presents a good step in providing a structured way of accessing all these information. The complete list of queries for the CQs and their results are published together with the ontology as specified in section 7. Figure 12 shows the query for CQ4.2 (*What are the operating temperatures of semiconductor laser devices working in a continuous wave operation?*) and figure 13 shows the result for the query.

6.2. Metric Based Evaluation

The ontology model achieves an Inheritance Richness metric of 0.133. This shows the level of detail of the ontology model in capturing the domain requirements. The ontology model is therefore suitable to represent concepts in the domain at data level, which enables efficient exploration of the data to derive useful inference regarding the semiconductor waver farbrication with target properties. The relationships correspond to the data population requirements making it easier to populate and analyze the data. This is crucial in optimizing the design process of the lasers. The ontology model does not therefore capture a lot of general information although it still captures definitions of high level concepts.

7. Technical Specifications

i **Interoperability:** The ontology model is implemented using OWL2, hence achieving FAIR'S Interoperability(I1). This also permits reasoning on the ontology model using OWL based reasoners.
ii **Indexing and Availability:** The ontological modelling is licensed under CC-BY 4.0 license. The ontology source code, data, evaluation scripts, queries and other related materials are also publicly available at our GitHub repository: https://github.com/DeperiasKerre/qcl_Onto, hence achieving FAIR'S Reusability (R1) and for replication purposes.
iii **Metadata Completion:** We used a checklist[12] for completing the vocabulary metadata to complete the metadata of the ontological representation. With regard to this, we represent the ontology authorship, version and license information using the Dublin Core metadata terms. This enables achieving FAIR'S Findability (F2 and F3) to make it easier to find and reuse the ontology model easily.

---

[12] https://w3id.org/widoco/bestPractices

8. Conclusion and Future Works

In this paper, we propose a concise ontological model of properties in the quantum cascade semiconductor laser domain. The ontology model aims to formalize the representation of the relationship between the design (semiconductor heterostructure) and the optolectronic characteristics of semiconductor laser devices. The formal representation is semantically enriched with information from relevant sources on the web. The Neon design methodology is adopted for the ontology design in order to capture the various ontology development scenarios and the FAIR data principles are also adopted for the publication of the ontology model. The ontology model is evaluated on the basis of three strategies: (i) Consistency and (ii) Ontology Verification and Validation and (iii) The richness of the ontology based on a richness evaluation metric. We check the consistency of the ontology and any pitfalls using the ontology checking tools. For verification and validation, we generate a sample knowledge graph from sample QCL properties data on which we run queries for the CQS. This is important in evaluating the ability of the ontology to capture the domain requirements. We compare the queries' output with the actual data for the specific relations of interest. All the competency questions are answered correctly. The ontology model richness evaluation metric also shows its ability to capture the properties of interest in detail. This provides a structured way of exploring the properties for optimizing the heterostructure fabrication process, especially when there is need for heterostructures with target performance properties. The ontology model can also be extended to represent information on other heterostructure semiconductor devices.

Future works will involve extending the ontology model with new concepts and relationships for instance, the heterostucture material and layer sequence doping and barrier properties. Other interesting research perspective will involve the use of large language models (LLMS) for populating the ontology model with textual data to generate KGs. We also aim to develop data models based on the extension of the ontological representation to generate knowledge graphs for the properties for massive exploration of semiconductor laser properties and the relationships within them.

Acknowledgement

This work was funded by the French Embassy in Kenya (Scientific and Academic Cooperation Department) and the CNRS (under the framework “Dispositif de Soutien aux Collaborations avec l'Afrique sub-saharienne"). The authors would also like to thank Strathmore university, School of Computing and Engineering sciences and the Doctoral academy for creating an opportunity for this work to be produced, and lastly Gaoussou SANOU (Institute of Human Genetics, CNRS, University of Montplellier, Montpellier, France) for the insightful discussions.

References

[1] Z. Vafapour, A.Keshavarz and H.Ghahraloud, The potential of terahertz sensing for cancer diagnosis, *Heliyon* **6** (2020).

[2] A.Kanno, P.T.Dat, N.Sekine, I.Hosako, T.Kawanishi, Y.Yoshida and K.I.Kitayama, High-speed coherent transmission using advanced photonics in terahertz bands, *IEICE Transactions on Electronics* **98**(12) (2015), 1071–1080.

[3] M.D.Wilkinson, M.Dumontier, I.J.Aalbersberg, G.Appleton, M.Axton, N.Blomberg, J.W. Boiten, L.B.d.S. Santos, P.E.Bourne, J.Bouwman, A.J.Brookes, T.Clark, M.Crosas, I. Dillo, O.Dumon, S.Edmunds, C.T.Evelo, R.Finkers, A.G.- Beltran, A.J.Gray, P.Groth, C.Goble, J.S.Grethe, J.Heringa, P.A.Hoen, R.Hooft, T.Kuhn, R.Kok, J.Kok, S.J.Lusher, M.E.Martone, A.Mons, A.L.Packer, B.Persson, P.Rocca-serra, M.Roos, R.v. Schaik, S.A.Sansone, E.Schultes, T.Sengstag, T.Slater, G.Strawn, M.A.Thompson, J.v. der Lei, E.v. Mullingen, J.Velterop, A.Waagmeester, P.Wittenburg, K.Wolstencroft, J.Zhao and B.Mons, The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* (2016), 1–9.

[4] S.J.L.Billinge, K.Rajan and S.B.Sinnott, From Cyberinfrastructure to Cyberdiscovery in Materials Science: Enhancing outcomes in materials research, education and outreach, *in Report from NSF-sponsored workshop held in Arlington, Virginia* (2006).

[5] W.Hunt, Materials informatics: Growing from the Bio World., *JOM Journal of the Minerals, Metals and Materials Society* **58**(7) (2006), 88–88.

[6] D.Kerre, A.Laurent, K.Maussang and D.Owuor, A text mining pipeline for mining the quantum cascade laser properties., *In European Conference on Advances in Databases and Information Systems, Barcelona, Spain* (2023), 293–406.

[7] K.Degtyarenko, P.D. Matost, M.Ennis, J.Hastings, M.Zbinden, A.McNaught, R.A. antara, M.Darsow, M.Guedj and M.Ashburner, ChEBI: a database and ontology for chemical entities of biological interest, *Nucleic acids research* **36** (2008), 344–350.

[8] B.Smith, On classifying material entities in basic formal ontology, *in: Proceedings of the Third Interdisciplinary Ontology Meeting* (2012), 1–13.

[9] A.Gangemi, N.Guarino, C.Masolo, A.Oltramari and L.Schneider, Sweetening ontologies with DOLCE, *In International conference on knowledge engineering and knowledge management* (2002), 166–181.

[10] H.Herre, General formal ontology (GFO): A foundational ontology for conceptual modelling, *in: Theory and Applications of Ontology: Computer Applications, Springer* (2010), 297–345.

[11] E.Ghedini, G.Goldbeck, G.Goldbeck and J.Friis, General formal ontology (GFO): A foundational ontology for conceptual modelling , (EMMO) (2021).

[12] K.Cheung, J.Drennan and J.Hunter, Towards an ontology for data-driven discovery of new materials, *In: AAAI Spring Symposium: Semantic Scientific Knowledge Integration* (2008), 9–14.

[13] T.Ashino, Materials Ontology: An Infrastructure for Exchanging Materials Information and Knowledge, *Data Science Journal* **9** (2010), 54–61.

[14] H.Li, R.Armiento and P.Lambrix, An Ontology for the Materials Design Domain, *In The Semantic Web – ISWC 2020* (2020), 212–227–.

[15] V.Rawte, J.P.McCusker, H.Zhao, L.C.Brinson, W.Chen, L.Schadler and D.L.McGuinness, An Ontology for a Polymer Nanocomposite Community Data Resource, *In Proceedings of the 2017 ACM on Web Science Conference* (2017), 411–412.

[16] D.G.Thomas, R.V.Pappu and N.A.Baker, Nanoparticle ontology for cancer nanotechnology research, *Journal of Biomedical Informatics* **44**(1) (2011), 59–74.

[17] J.Hastings, N.Jeliazkova, G.Owen, G.Tsiliki, C.R.Munteanu, C.Steinbeck and E.Willighagen, enanomapper: harnessing ontologies to enable data integration for nanomaterial risk assessment, *Journal of biomedical semantics* **6**(1) (2015), 10.

[18] X. Zhang, C. Hu and H. Li, Semantic query on materials data based on mapping MATML to an OWL ontology, *Advanced Materials Processes* **18** (2009), 1–17.

[19] X.Zhang, D.Pan, C.Zhao and K.Li, MMOY: Towards deriving a metallic materials ontology from Yago, *Advanced Engineering Informatics* **30** (2016), 687–702.

[20] A.Z. Ihsan, D. Dess'ı, M. Alam, H. Sack and S. Sandfeld, Steps towards a dislocation ontology for crystalline materials, *in: Proceedings of the Second International Workshop on Semantic Digital Twins Co-Located with the 18th Extended Semantic Web Conference (ESWC)* **2887** (2021).

[21] M. Alam, H. Birkholz, D. Dess'ı, C. Eberl, H. Fliegl, P. Gumbsch, P. von Hartrott, L.M. adler, M. Niebel, H. Sack and A. Thomas, Ontology modelling for materials science experiments, *in: Joint Proceedings of the Semantics Co-Located Events: Poster Demo Track and Workshop on Ontology-Driven Conceptual Modelling of Digital Twins Co-Located with Semantics 2021* **2941** (2021).

[22] F.L. Piane, M. Baldoni, M. Gaspari and F. Mercuri, Introducing MAMBO: Materials and molecules basic ontology, *in: Proceedings of the Workshop on Domain Ontologies for Research Data Management in Industry Commons of Materials and Manufacturing (DORIC-MM 2021) Co-Located with the 18th European Semantic Web Conference (ESWC 2021)* (2021), 28–39.

[23] European Committee for Standardization, A Guide to the Development and Use of Standards Compliant Data Formats for Engineering Materials Test Data (2010).

[24] C.F.V. II, A.A.Krisnadhi, M.Cheatham, K.Janowicz, H.Ferguson, P.Hitzler and A.P.C.Buccellato, An ontology design pattern and its use case for modeling material transformation, *Semantic Web* **8**(5) (2017), 719–731.

[25] M.T.Horsch, C.Niethammer, G.Boccardo, P.Carbone, S.Chiacchiera, M.Chiricotto, J.D.Elliott, V. Lobaskin, P.Neumann, P.Schiffels, M.A.Seaton, I.T.Todorov, J.Vrabec and W.L. Cavalcanti, Semantic interoperability and characterization of data provenance in computational molecular engineering, *Semantic Web* **65**(3) (2020), 1313–1329.

[26] Kaufman and E.Begley, MatML: A data interchange markup language, *Advanced Materials Processes* **161** (2003).

[27] T.P. Tanon, G.Weikum and F.Suchanek, YAGO 4: A reason-able knowledge base, *in: The Semantic Web, ESWC* (2020), 583–596.

[28] M.C.Suarez-Figueroa, A.Gomez-Perez and M.Fernandez-Lopez, The NeOn method- ology for ontology engineering, *In Ontology engineering in a networked world, Berlin Heidelberg* (2011), 9–34.

[29] P. Lambrix and V. Ivanova, A unified approach for debugging is-a structure and mappings in networked taxonomies, *In Ontology engineering in a networked world, Berlin Heidelberg* **4** (2013).

[30] M.Poveda-Villalon, A.G. omez-Perez and M.C.Suarez-Figueroa, OOPS! (OntOlogy pitfall scanner!): An on-line tool for ontology evaluation, *In Ontology engineering in a networked world, Berlin Heidelberg* **10**(2) (2014), 7–34.

[31] T.Lebo, S.Sahoo, a.K.B. D.McGuinness, J.Cheney, D.Corsar, D.Garijo, S.Soiland-Reyes, S.Zednik and J.Zhao, Prov-o: The prov ontology (2013).

[32] R.Haas, P.J.Keller, J.Hodges and J.Spivak, Quantities, units, dimensions and data types ontologies (QUDT), *Technical report, Top Quadrant and NASA* (2014). http://qudt.org.

[33] B.Glimm, I.Horrocks, B.Motik, G.Stoilos and Z.Wang, HermiT: an OWL 2 reasoner, *Journal of automated reasoning* **53** (2014), 245–269.

[34] R. de Almeida Falbo, SABiO: Systematic Approach for Building Ontologies, *In: 1st Joint Workshop Onto.Com/ODISE on Ontologies in Conceptual Modeling and Information Systems Engineering* (2014).

[35] S.Tartir, I.B.Arpinar, M.Moore, A.Sheth and B.Aleman-Meza, OntoQA:MetricBasedOntologyQualityAnalysis, *In:IEEEICDMWorkshoponKnowledgeAcquisitionfromDistributed,Autonomous,SemanticallyHeterogeneousDataandKnowledge Sources* (2005).