An ontology for units of measures across history, standards, and scientific and technology domains

Oskar B. Andersson[1,2,*], Huanyu Li[2,3], Patrick Lambrix[2,3] and Rickard Armiento[1,2,*]

[1] *Department of Physics, Chemistry and Biology, Linköping University, 581 83 Linköping, Sweden*
[2] *Swedish e-Science Research Centre, Linköping, Sweden*
[3] *Department of Computer and Information Science, Linköping University, 581 83 Linköping, Sweden*

Abstract

Units of measure are central in all areas of science and technology. There are several ontological frameworks aiming to improve interoperability and precision in digital data exchange of quantities involving units. We introduce an ontology that specifically targets challenges for handling units across databases of computational and experimental data from various sources. The ontology is created using definition files from the community-driven OPTIMADE standard for a common API for materials databases. The resulting ontology allows addressing data integration challenges encountered in that effort, including (*i*) referencing both specific and more general instances of units that have changed over time; (*ii*) the use of unit systems to define short domain-relevant identifiers for a collection of units that make sense within a specific subdomain, rather than having to adopt globally standardized naming schemes; (*iii*) specifications of relationships between units that enables tools to convert between them; and (*iv*) units not part of the International System of Units (SI) can be represented without defining them in SI units or using SI system conventions. This paper provides a brief survey of existing ontologies for units of measure and then presents the design and discuss features of an ontology based on the OPTIMADE unit definitions.

Keywords

Ontology, Units of measure, Unit ontologies, Materials Science

1. Introduction

Units of measure are reference quantities used when expressing physical quantities obtained from experimental measurements, theoretical models, or by other means. They essentially underpin all systematic communication of physical quantities throughout all areas of science and technology and are therefore crucial to standardize [1, 2]. The International System of Units (SI), coordinated by the International Bureau of Weights and Measures (BIPM) [3], has long held a uniquely strong position as a system of standardized unit definitions. However, units outside the SI system are also heavily used in many domains. With the increasing adoption of ontologies to facilitate precise digital communication, processing, and storage of information

---

SeMatS 2024: The 1st International Workshop on Semantic Materials Science co-located with the 20th International Conference on Semantic Systems (SEMANTiCS), September 17-19, Amsterdam, The Netherlands.

*Corresponding author.

oskar.andersson@liu.se (O. B. Andersson); huanyu.li@liu.se (H. Li); patrick.lambrix@liu.se (P. Lambrix); rickard.armiento@liu.se (R. Armiento)

0009-0003-6158-1857 (O. B. Andersson); 0000-0003-1881-3969 (H. Li); 0000-0002-9084-0470 (P. Lambrix); 0000-0002-5571-0814 (R. Armiento)

[4, 5, 6], several ontologies have emerged to describe units and their relationships [7, 8, 9, 10]. An incident frequently cited to highlight the importance of standardized ways to handle units in digital systems is the loss of the NASA Mars Climate Orbiter attributed to incorrect assumptions about units between two software components [11].

A key application for ontologies is database integration. Databases can include semantic information through ontologies to help ensure that data from different providers are systematically and correctly interpreted [12, 13]. Reasoning based on these semantics can be used to, e.g., map a database query expressed in data fields available in one database on those available in another. A good example of an application area undergoing an increasing adoption of rigorous data standards is materials science [14, 15]. The Open Databases Integration for Materials Design (OPTIMADE) consortium, which includes the present authors, develops a standard for a common API for materials data[1] [16]. The recently released version v1.2 of the standard introduces machine-readable property definitions for database fields using an extended subset of the JSON schema standard[2] [17]. These property definitions include definitions of units of measure to address interoperability challenges met across the databases included in the consortium. Turning these unit definitions into an ontology expressed using semantic languages, e.g., OWL and RDF, enables aligning these definitions with other unit ontologies. An ontology of these community-agreed definitions may also be useful in semantically guided systems beyond the OPTIMADE API, e.g., for automated GraphQL server generation [18, 19].

A feature of the SI system relevant in the present context is that the units are occasionally redefined to utilize progress in measurement methodology and technology that allows more accurate measurements. For example, when the name SI was established in 1960, the **metre** was defined in terms of wavelengths of radiation from the krypton 86 atom, whereas in 1983, it was redefined in terms of the speed of light. Likewise, in 1960, the **second** was defined to use a particular astronomical definition of the year, and in 1967 it was redefined based on oscillations between levels of the cesium 133 atom. Yet another redefinition of the second based on modern atomic clocks is being prepared for 2030 [20]. These redefinitions of the base SI units also implicitly redefine all related derived units. For example, the **tesla** is a magnetic flux unit equal to $\mathrm{kg/s^2A}$ which consequently has been redefined four times since 1960. The implicit redefinitions also reach outside the SI system, e.g., the **ångström** length unit, which is in ubiquitous use in materials science and spectroscopy, is usually defined as 10^{-10} m. This definition is only unambiguous when combined with a particular definition of the **metre**. Hence, there are at least two versions of this unit, one before and one after 1983, when the SI **metre** was redefined.

The differences between the historical definitions of the SI units may seem inconsequential, given how they are meant to only make the unit definitions more precise, i.e., a new definition is meant to be within the uncertainty of the most accurate measurements of the quantity used for the prior definition. Nevertheless, in particular in the context of databases containing numerical results of computational simulations, values can be computed and recorded to virtually any precision. A redefined unit then fundamentally alters the interpretation of the stored quantities. Hence, especially in this context, it is useful to allow references to particular historical unit

---

[1] https://www.optimade.org

[2] https://json-schema.org/specification

definitions. Furthermore, another useful feature is to include less precise representations that refer to any historical definition. The authors are unaware of any unit ontology currently in use that allows separately handling and referencing different historical SI definitions and more broadly generalized definitions. However, a key feature of the OPTIMADE property definitions is that they are designed to handle multiple units and unit systems with definitions that change over time. The definition files included with version v1.2 of the standard cover the changes in the SI system since 1960. Hence, bringing these definitions and relations into a standard ontology format will allow this feature to be used also in other semantic systems.

2. Related work

In the process of designing an ontology based on the unit definitions provided as part of the OPTIMADE standard, we have surveyed several existing ontologies and related efforts that provide formal representations of units of measure in various ways. We comment on them briefly below. Several more in-depth reviews of ontologies for units are also available in the literature [7, 8, 9, 10].

- Quantities, Units, Dimensions and Data Types Ontologies (QUDT)[3] [21] provides a collection of linked ontologies that is described as an *"architecture for the conceptual representation of quantities, quantity kinds, units, dimensions, and data types."* It is largely based on the SI standard, ISO standards on Units and Quantities (e.g., ISO 80000 [22]), and the NIST Guide for the use of the International System of Units [23]. QUDT is extensible and has a submission system to add units to existing or new vocabularies. However, we have not found information that suggests how to accommodate changes in units over time and the definitions in the current version 2.1.41 appear to be based on the pre-2019 version of the SI system (e.g., the kilogram entry refers to the international kilogram prototype).
- Unified Code for Units of Measure (UCUM)[4] is not presented as an ontology, but rather a *"code system intended to include all units of measures being contemporarily used in international science, engineering, and business."*. It was originally created for clinical information systems [24]. An RDF datatype has been defined to represent physical quantities using this code system [25]. There does not appear to be a clear way to use the UCUM code system to refer to different versions of, e.g., the SI base units.
- Ontology of units of Measure (OM)[5] [26] models concepts and relations for the formulation of quantitative knowledge with a strong focus on units, quantities, measures, and dimensions. The ontology does not appear to have any explicit definitions or mechanisms for changes of unit definitions over time, and also the latest version 2.0 refers to the pre-2019 SI definition of the kilogram.
- Semantic Web for Earth and Environmental Terminology (SWEET)[6] [27] is a highly modular ontology suite covering Earth system science which also covers scientific units.

---

[3] https://qudt.org/

[4] https://ucum.org

[5] https://github.com/HajoRijgersberg/OM

[6] https://github.com/ESIPFed/sweet

- Extensible Observation Ontology (OBOE)[7] [28, 29] is a formal ontology for capturing the semantics of scientific observation and measurement oriented around the concepts of observations, measurements, entity, characteristic, standard, and protocol.
- Measurement Units Ontology (MUO)[8] was created to represent units in a software development framework for mobile devices. It provided definitions of unit instances automatically extracted from UCUM.
- Quantity Kinds and Units (QU)[9] is based on the conceptual framework (or model) Quantity, Unit, Dimension, and Value (QUDV), a part of OMG SysML [30]).
- Units of Measurement Ontology (UO)[10] [31] targets the standardization of units of measurement in the biomedical domain.
- The Wikidata project [32] provides a range of unit definitions used for structured data to be referenced from Wikimedia websites, including Wikipedia.[11]
- The GNU units software package distributes a range of unit definitions in a database file.[12] This collection of units has been adopted to express formalized property definitions in the OpenKIM project [33].

Several of the surveyed ontologies include representations of in-depth concepts in the domain of quantities and units. In particular, many include the concept of unit dimensions [7, 2]. While dimensions and related concepts are critical to describe all aspects of how quantities, units, and unit systems relate, they can be difficult to model in a unit system-independent way. However, these concepts are not strictly necessary to represent units used for quantities and to aid database integration, and are therefore not currently part of the ontology presented here.

The primary motivation for the ontology presented in this work is to encode the community-agreed unit definitions part of the OPTIMADE standard. We do not aim for this work to fully evaluate the advantages and limitations in the resulting design compared to existing ontologies for units of measure.

3. Definition of the ontology

3.1. Requirements

We present an ontology designed around the following principles:

- Allow multiple historical definitions of the same unit to co-exist within the same version of the ontology (i.e., not via the versioning of the ontology itself).

---

[7] https://github.com/NCEAS/oboe

[8] While many works reference the MUO, we have been unable to locate a detailed description online or as a published work. An archived web page with some details can be accessed via the Internet Archive Wayback Machine at https://web.archive.org/web/20130723220432/http://forge.morfeo-project.org/wiki_en/index.php/Units_of_measurement_ontology

[9] https://www.w3.org/2005/Incubator/ssn/ssnx/qu/qu and https://www.w3.org/2005/Incubator/ssn/ssnx/qu/qu-rec20.html

[10] http://obofoundry.org/ontology/uo

[11] https://www.wikidata.org/wiki/Wikidata:Units

[12] https://www.gnu.org/software/units/

- Standards-agnostic design (i.e., not designed around the SI standard).
- Freedom to represent units using relevant domain-specific short symbols.
- Support compact mathematical expressions to express compound units and relationships between units.
- Provide relevant identifiers for various uses: display symbols for mathematical typesetting and latin symbols and identifiers for referencing in computer code and data contexts.
- Entities separated into physical units, constants, and prefixes.
- Support of referencing unit systems as a collection of units with associated short symbols.
- IRIs that refer to the exact same entities as the OPTIMADE unit definitions are retained from those definitions.
- A versioning scheme that handles changes in the ontology and in the referenced unit definitions provided via OPTIMADE.

3.1.1. Use cases

A couple of example use cases that can be addressed by the ontology follows:

1. Specification (via IRI) of the unit of measure that applies to a stored numerical value (e.g., for a database field) at the level of specificity that applies to that quantity. For example, one IRI refers to the SI unit **second** defined in 1967, and another IRI more broadly to any SI definition of the **second** (either the current one or any of the historical SI definitions).
2. Specification (via IRI) of the precise unit used in a computer programming interface, e.g., to avoid ambiguity for values returned by an API.
3. Determination if the numerical values of two quantities can be directly compared based on reasoning to determine if their units are the same or if one can be generalized into the other.
4. Determine an appropriate typographical symbol to use for a unit in a user interface.
5. Reference a set of domain-relevant short identifiers referencing specific unit definitions with a single IRI.
6. Discover mathematical relationships between units (either as defining relationships or approximate, such as for the dalton unit of atomic mass and the kilogram), which can be used by a tool for unit conversions.

3.2. Development method

The ontology has been developed by re-engineering the community-agreed OPTIMADE unit definitions into an ontological representation. The OPTIMADE unit definitions are provided as machine-readable files using a JSON-based format specified by a JSON Schema meta-schema which is part of the OPTIMADE standard. In practice, the JSON files are generated from more user-friendly YAML-formatted source files. These files are also used to generate human-readable HTML and Markdown (MD) files made available for browsing via one of the websites maintained by OPTIMADE.[13] The OPTIMADE standard v1.2 includes unit definition files for all current

---

[13] https://schemas.optimade.org

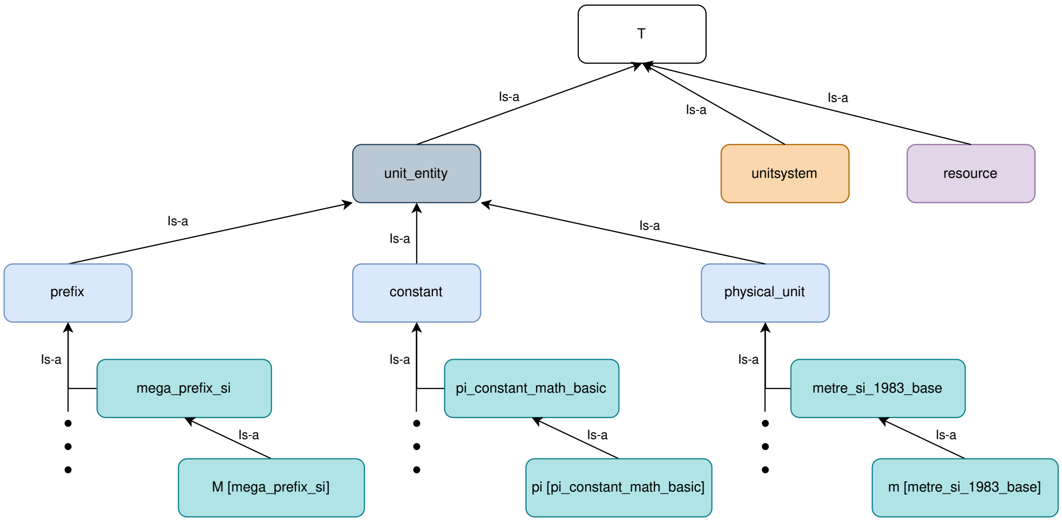

Figure 1: Overview of the concepts of the ontology.

and historical SI unit definitions since 1960, and most other, non-SI, units referenced from the nine editions of the SI brochure [3]. In particular, this includes all units referred to as *"non-SI units accepted to use with the SI system"*. A few other non-SI units are also included, e.g., the digital information units **bit** and **byte**.

In this work, the source format is the JSON unit definitions provided as part of the OPTIMADE standard, which are converted by a special-purpose tool into the ontological representation RDF using turtle syntax [34]. The ontology is intentionally designed to, as closely as possible, reflect the intent and content of the OPTIMADE unit definitions. One materials science domain expert and two ontology knowledge engineers have reviewed the ontology design.

A pre-release version v0.9.0 of the ontology is available via our GitHub repository: https://github.com/LiUSemWeb/units-of-measure.

3.3. Ontology design

We have chosen a class-based design approach consisting of hierarchical levels of concepts where all unit-related data is purely expressed through object properties and data properties in class definitions. The ontology consists of the following top level concepts as illustrated in Fig. 1:

- $unitsystem \sqsubseteq \top$
- $resources \sqsubseteq \top$
- $unit\_entity \sqsubseteq \top$
- $physical\_unit \sqsubseteq unit\_entity$
- $constant \sqsubseteq unit\_entity$
- $prefix \sqsubseteq unit\_entity$

Here, $\top$ is the top concept (*owl:Thing* in OWL). Specific units of measure, e.g., the **metre** unit as defined by SI in 1983, are sub-concepts of *physical_unit*. Sub-concepts of *constant* are numerical and scientific constants, e.g., the mathematical constant π and the **elemental charge** physical constant. The inclusion of constants is arguably not essential for an ontology for units of measure, but is useful by allowing unit definitions to describe relationships to other units using these constants. Sub-concepts of *prefix* are unitless scaling factors used in unit systems to reference units scaled to larger and smaller magnitudes, e.g., the SI prefix **mega** with symbol M denotes a factor of one million.

The concepts *physical_unit*, *constant*, *prefix* are sub-concepts of the more general concept *unit_entity*. Unit entities have an annotation property *description* for expressing the definition of the unit entity with a human-readable string and a set of data and object properties. One of these object properties are to specific individuals of the *resource* concept to express various types of relationships to external resources, e.g., the official standards document that is the source of the definition, or an informational link to the description of the unit in Wikipedia. The sub-concepts of *unit_entity* are *specific unit entity concepts*. Each of these specific unit entity concepts has another level of sub-concepts, *symbol concepts*, for a specific unit entity represented by a specific symbol. The symbol concepts provides a form of reification for the combination of a specific unit entity and a symbol, which allows concepts that are, e.g., the same specific unit entity, but with several different symbols to co-exist in the ontology.

The sub-concepts of *unitsystem* are collections of unit entities, e.g., the concept for the SI 2019 unit system has relationships to all the *physical_unit* and *prefix* symbol concepts with their respective standard SI symbols, defined as part of the SI system of units in the 9th edition of the SI brochure [3]. It is possible to extend the ontology using sub-concepts of *unitsystem* to define other standardized units systems and domain-specific collections of physical units and prefixes with specific choices of symbols.

3.3.1. Labels

All direct sub-concepts of *physical_unit*, *constant*, and *prefix* are given an *rdfs:label* annotation. These labels are taken from the `x-optimade-definition/label` field in the OPTIMADE definition files. They are on the format `<name>_<standard>_<year>_<category>`, e.g., `metre_si_1960_base`. The first segment `<name>` is a short name describing the unit. The second segment `<standard>` is the name of the standards organization from which the definition originates, or the word “independent” if the definition cannot be referenced to such an organization (e.g., the definition of the **bit** unit references an early paper on digital information by Shannon). The third segment `<year>` is the year that the definition was introduced. The fourth segment `<category>` is a relevant categorization either made by the standards organization, or if no such categorization exists, just an informal categorization as a means to disambiguate cases of similarly or identically named units. As stated in a previous example, **minute** is often used to refer to both a time unit and an angular unit, thus it is helpful to know at a glance which kind of **minute** a label applies to.

Furthermore, the symbol concepts (i.e., the second level sub-concepts) of *physical_unit*, *constant*, and *prefix* are given labels on the format `<symbol> [<name>_<standard>_<year>_<category>]` where `<symbol>` is the symbol, e.g., the SI

1960 definition of the metre unit referred to using the symbol `m` is given the label `m [metre_si_1960_base]`.

3.3.2. IRIs

The ontology base IRI is `<https://github.com/LiUSemWeb/units-of-measure#>` for the pre-release of the ontology. However, the concepts *physical_unit*, *constant*, *prefix* reference entities already assigned IRIs by the OPTIMADE standard, which then are retained in the ontology. These IRIs reside under `<https://schemas.optimade.org/defs/v1.2/>`, e.g., *metre_si_1960_base* has IRI `<https://schemas.optimade.org/defs/v1.2/units/si/1960/base/metre>`. The symbol sub-concepts of *physical_unit*, *constant*, and *prefix* for, e.g., a specific unit represented by a specific symbol are assigned the OPTIMADE IRI with an additional anchor, `#symbol`. For example, *metre_si_1960_base* has a sub-concept labeled `m [metre_si_1960_base]` with IRI `<https://schemas.optimade.org/defs/v1.2/units/si/1960/base/metre#m>`. Finally, concepts in the ontology that do not have IRIs assigned by OPTIMADE are assigned IRIs under the ontology base IRI.

3.3.3. Symbols and other identifiers

A key aspect in the use of units is the practice of representing them by symbols which are compact domain-specific identifiers. In practical use, different symbols are often used for the same units in different contexts and domains, e.g., the **nautical mile** navigational length unit has multiple common symbols: M, NM, nmi, and Nm. As outlined above, the primary way symbols are implemented in the ontology is via symbol concepts as sub-concepts of a specific physical unit, constant or prefix.

The symbols used for symbol concepts are taken to be short identifiers composed of Latin characters.[14] The symbol is often just one or a few characters, e.g., “T” for the **tesla** unit and “Bq” for the **becquerel** unit. The symbol can be longer if no clear short Latin symbol exists, e.g. the **ångström** unit uses ”angstrom” rather than the more standard non-Latin character “Å”. When possible, this is the symbol defined in the SI standard for SI units. However, when the SI symbol requires Unicode or mathematical notation, a Latin version is chosen, e.g., “mc” is used for the SI prefix **micro**, rather than “μ”. These limitations gives a symbol that is useful, e.g., for referencing the unit in data and programming contexts, where non-Latin characters may not be available.

The symbol limited to Latin characters is complemented with:

- *title*: A data property for each unit entity concept for the name as a Unicode string with white space and capitalization where relevant, e.g., “nautical mile”, “ångström”, “degree Celsius”.
- *display_symbol*: A data property for symbol entity concepts for an appropriate representation of the unit symbol in Unicode, e.g., “Å” for the **ångström** unit and “Ω” for the **ohm** unit; and, when needed, mathematical expressions using MathJax notation, e.g., “\(\mu_B\)” for the **Bohr magneton** unit.

---

[14] More precisely: lowercase a–z, uppercase A–Z, and the underscore character “_”, but not white-space characters.

As an example, the unit **nautical mile** $nauticalmile\_si\_1970\_temporary$ and its symbol concepts are defined as:

$$\begin{aligned}
&nauticalmile\_si\_1970\_temporary \equiv physical\_unit \sqcap (title\ value\ "nauticalmile") \sqcap \ldots\\
&M\ [nauticalmile\_si\_1970\_temporary] \equiv nauticalmile\_si\_1970\_temporary \sqcap (symbol\ value\ "M")\\
&\quad \sqcap (display\_symbol\ value\ "M")\\
&NM\ [nauticalmile\_si\_1970\_temporary] \equiv nauticalmile\_si\_1970\_temporary \sqcap (symbol\ value\ "NM")\\
&\quad \sqcap (display\_symbol\ value\ "NM")\\
&Nm\ [nauticalmile\_si\_1970\_temporary] \equiv nauticalmile\_si\_1970\_temporary \sqcap (symbol\ value\ "Nm")\\
&\quad \sqcap (display\_symbol\ value\ "Nm")\\
&nmi\ [nauticalmile\_si\_1970\_temporary] \equiv nauticalmile\_si\_1970\_temporary \sqcap (symbol\ value\ "nmi")\\
&\quad \sqcap (display\_symbol\ value\ "nmi")
\end{aligned}$$

Furthermore, in the design of the ontology, the same symbol can be used for different unit definitions to allow adhering to domain-specific conventions that cannot be handled by enforcing unique symbols across the ontology. While it is useful to standardize the use of a specific symbol, the reality is that the same symbols are used for different unit definitions. A few examples include: “minute” used for both **arcminute** and **minute** (for time); “tonne” for **metric ton**, **short ton**, **long ton**; or “billion” which is 10^9 in the short scale and 10^{12} in the long scale.

3.3.4. Standards agnostic design

The formal aspects of representing units of measure in mathematical and scientific frameworks can be handled in more than one way. For example, concerns have been raised about the standing of the **mole** unit in the SI system [35], and one can speculate that similar concerns have led the SI system to not yet adopt any units for digital information, e.g., the **bit** or **byte**. The aim of the ontology presented in this work is not to provide a framework for the formally most rigorous way of representing physical quantities according to current standards, but rather to allow representing existing uses of units as used across history, standards, and domains in a way that preserves the full original meaning. If the representation of a unit in an ontology requires that it is reinterpreted into a different formal system it can subtly alter the meaning of the data (cf., reinterpretation of the unit used for old data using the most recent SI definitions). Our ontology is designed to allow sub-concepts of *physical_unit* to represent anything possible to use as a reference quantity to give meaning to a numerical value from, e.g., a measurement, a calculation, or stored in a database.

More specifically, the strong standing of the SI system has led to multiple ontologies designed around that standard, for example, by including SI-systems-specific categorizations (e.g., base and derived units) or completely adopting the unit dimensions that the SI system is based on. The ontology presented in this work has been designed with the intent not to adopt conventions of the SI system that are different in other unit systems.

3.3.5. Compact mathematical expressions for compound units

The ontology supports expressing compound unit expressions using a domain-specific compact mathematical string language defined in the OPTIMADE specification [16]. Similar languages are found in the H5MD specification UCUM [24] and [36]. The compound unit expressions are also used within the ontology to express the *defining_relation_base_units_expression* and *approximate_relations_base_units_expression* data properties. Each symbol in the base units in the compound unit expression is defined by referencing the symbol concept through the *defining_relation_base_unit* and *approximate_relation_base_unit* object property. This allows the freedom to use any desired symbol in the compound unit expression, as long as the symbol concept is defined to its unit super-concept in the ontology. A *defining relation* is used strictly for units defined in terms of other units represented in the ontology to describe that relationship. An *approximate relation* describes a non-defining relationship that can be used to convert the value of a quantity to express it with a different unit.

3.3.6. Versioning

There are three types of changes that affect the ontology: (*i*) units are occasionally redefined by standard organizations such as SI; (*ii*) the OPTIMADE consortium may change the definition files the ontology is based on; (*iii*) we may make changes to the ontology based on the same OPTIMADE definition files. As an example, consider the 1960 definition of the SI metre unit with IRI `<https://schemas.optimade.org/defs/v1.2/units/si/1960/base/metre>`. Changes in the SI unit definitions are handled by introducing the new ones concurrent with the old ones. The 1960 SI metre thus coexist with the redefined 1983 SI metre with IRI `<https://schemas.optimade.org/defs/v1.2/units/si/1983/base/metre>`. For changes made by the OPTIMADE consortium, the released version of the OPTIMADE standard is also included in the IRIs naming scheme as the `v1.2` in the second path component, which thus will change when the ontology is modified to refer to the updated definitions (which also incurs a version change of the ontology itself).

All changes to the ontology, both in general or in response to changes of the kinds described above, will be reflected by updating the version number in the ontology version IRI, which we intend to assign using the pattern MAJOR.MINOR.PATCH, based on the semantic versioning scheme.[15] The semantic versioning scheme is interpreted according to the following: we will increment the PATCH number for changes that only address unintended errors or omissions, and the MINOR number when expanding the ontology with additional definitions, concepts, and attributes that do not affect the use of those existing in prior releases of the ontology. Major changes that alter the overall design and/or the meaning of concepts, attributes, or their relationships, in ways that may be incompatible with the use or imports of prior versions will increment the MAJOR version number.

---

[15] https://semver.org/

3.3.7. Conversion from OPTIMADE definitions

As explained above, the ontology is created directly from the machine-readable JSON files maintained by OPTIMADE. We have implemented a conversion tool in Python that reads files on the format documented in version 1.2 of the OPTIMADE standard [16] and derives concept definitions and their relationships for specific units as described in the sections above. A more in-depth technical description of this conversion tool will be presented in a separate paper.

4. Applications of the ontology

The ontology allows reference to units used in OPTIMADE property definitions using standard semantic languages. This enables the alignment of units in databases that use OPTIMADE property definitions with existing unit ontologies. Furthermore, ontology reasoners can be applied to these uses of units, and the OPTIMADE unit definitions can be integrated with other semantical frameworks. Section 3.1.1 lists a number of such specific use case examples for the ontology.

More generally, databases need to represent units of measure to provide data about physical quantities. This information is essential to aggregate data from more than one source or query multiple databases with related data in a meaningful way. In the following we propose such a use case in more detail: consider two databases that provide formation energies for materials obtained from different computational software. Both software packages use the non-SI unit **electron volt**, which implicitly depends on the derived SI unit **volt**. If the databases declare their units using the ontology provided in this work, a client accessing the data will be able to make an informed decision on the meaning of a comparison of these data fields. If both databases declare the use of the exact same historical definition of the **volt** unit, it is meaningful to directly compare the numerical values to full precision. If they declare different historical definitions, the client must either perform a unit conversion or utilize the fact that both the specific definitions generalize into *volt_si_general*, the higher level concept for the current, or any one of the historical definitions of **volt**. A comparison on the level of *volt_si_general* yields less precise information, but at least it is clear to the client that this is the case.

As mentioned above, during construction, the ontology was reviewed by one materials science domain expert and two ontology knowledge engineers. We also emphasize that the construction is based on unit definitions in OPTIMADE that have been discussed by that consortium of domain experts. The ontology has not yet undergone validation or performance measurements beyond these considerations.

5. Conclusion

In this paper we have presented an ontology that covers units of measure for physical units, constants, and prefixes. The ontology have been created from a set of community-agreed property definitions in the OPTIMADE materials database API standard. Our ontology engineering has strived to preserve the original design idea of the source definition files as much as possible while finding a way to express that design using standard semantic languages. The result is an ontology that aims to allow the expression of units the way they are used in databases.A few key design decisions are (*i*) a rich framework for representation that can accomodate multiple historical definitions and different domain-relevant symbols for the same unit; (*ii*) a unit system agnostic design to allow representation of unit definitions across multiple standards; *(iii)* unit definitions provided by domain experts that covers domain-relevant definitions, symbols, and other identifiers. The ontology is populated by the definitions provided with the OPTIMADE release v1.2, which includes the SI units in their current and past definitions since 1960, most other definitions provided in the nine editions of the SI brochure [3], and a few additional units in practical use in the fields of materials science and spectroscopy.

Acknowledgments

We acknowledge all contributors to the OPTIMADE property and unit definitions on which this work is based. R.A. and O.B.A acknowledge support from the Swedish Research Council (VR) Grant No. 2020-05402; R.A., O.B.A and P.L. from the Swedish e-Science Research Centre (SeRC); H.L. from the Swedish National Graduate School in Computer Science (CUGS); and P.L. and H.L. from the EU Horizon project Onto-DESIDE (Grant Agreement 101058682).

References

[1] G. Price, On the communication of measurement results, Measurement 29 (2001) 293–305. doi:10.1016/S0263-2241(00)00053-1.

[2] M. P. Foster, Quantities, units and computing, Computer Standards & Interfaces 35 (2013) 529–535. doi:10.1016/j.csi.2013.02.001.

[3] BIPM, Le Système international d'unités / The International System of Units ('The SI Brochure'), ninth ed., Bureau international des poids et mesures, 2019. URL: http://www.bipm.org/en/si/si_brochure/.

[4] F. Probst, Observations, measurements and semantic reference spaces, Applied Ontology 3 (2008) 63–89. doi:10.3233/AO-2008-0046, publisher: IOS Press.

[5] J. M. Keil, S. Schindler, Comparison and evaluation of ontologies for units of measurement, Semantic Web 10 (2018) 33–51. doi:10.3233/SW-180310.

[6] S. Schindler, J. Keil, Building Ontologies for Reuse: Lessons Learned from Unit Ontologies, in: Proceedings of the Joint Ontology Workshops 2019, volume 2518 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019. URL: https://ceur-ws.org/Vol-2518/paper-BOG4.pdf.

[7] I. Johansson, Metrological thinking needs the notions of parametric quantities, units and dimensions, Metrologia 47 (2010) 219. doi:10.1088/0026-1394/47/3/012.

[8] H. Rijgersberg, M. Wigham, J. L. Top, How semantics can improve engineering processes: A case of units of measure and quantities, Advanced Engineering Informatics 25 (2011) 276–287. doi:10.1016/j.aei.2010.07.008.

[9] X. Zhang, K. Li, C. Zhao, D. Pan, A survey on units ontologies: architecture, comparison and reuse, Program 51 (2017) 193–213. doi:10.1108/PROG-08-2015-0056.

[10] M. D. Steinberg, S. Schindler, J. M. Keil, Use Cases and Suitability Metrics for Unit Ontologies, in: M. Dragoni, M. Poveda-Villalon, E. Jimenez-Ruiz (Eds.), OWL: Experiences

and Directions - Reasoner Evaluation, Springer International Publishing, Cham, 2017, pp. 40–54. doi:10.1007/978-3-319-54627-8_4.

[11] Mars Climate Orbiter Mishap Investigation Board - Phase I Report, Mars Climate Orbiter Mishap Investigation Board - Phase I Report, Technical Report 611, NASA, 1999. URL: https://llis.nasa.gov/lesson/641.

[12] J.-L. Hippolyte, M. Romanchikova, M. Bevilacqua, P. Duncan, S. E. Hunt, F. Grasso Toro, A.-S. Piette, J. Neumann, Using Ontologies to Create Machine-Actionable Datasets: Two Case Studies, Metrology 3 (2023) 65–80. doi:10.3390/metrology3010003.

[13] B. G. Pelkie, L. D. Pozzo, The laboratory of Babel: highlighting community needs for integrated materials data management, Digital Discovery 2 (2023) 544–556. doi:10.1039/D3DD00022B, publisher: Royal Society of Chemistry.

[14] P. Lambrix, R. Armiento, A. Delin, H. Li, FAIR Big Data in the Materials Design Domain, in: A. Zomaya, J. Taheri, S. Sakr (Eds.), Encyclopedia of Big Data Technologies, Springer International Publishing, Cham, 2020, pp. 1–8. doi:10.1007/978-3-319-63962-8_293-2.

[15] H. Li, R. Armiento, P. Lambrix, A Method for Extending Ontologies with Application to the Materials Science Domain, Data Science Journal 18 (2019) 50. doi:10.5334/dsj-2019-050.

[16] M. L. Evans, J. Bergsma, A. Merkys, C. W. Andersen, O. B. Andersson, D. BeltrÃ¡n, E. Blokhin, T. M. Boland, R. C. Balderas, K. Choudhary, A. D. DÃaz, R. D. GarcÃa, H. Eckert, K. Eimre, M. E. F. Montero, A. M. Krajewski, J. J. Mortensen, J. M. N. Duarte, J. Pietryga, J. Qi, F. d. J. T. Carrillo, A. Vaitkus, J. Yu, A. Zettel, P. B. de Castro, J. Carlsson, T. F. T. Cerqueira, S. Divilov, H. Hajiyani, F. Hanke, K. Jose, C. Oses, J. Riebesell, J. Schmidt, D. Winston, C. Xie, X. Yang, S. Bonella, S. Botti, S. Curtarolo, C. Draxl, L. E. F. Cobas, A. Hospital, Z.-K. Liu, M. A. L. Marques, N. Marzari, A. J. Morris, S. P. Ong, M. Orozco, K. A. Persson, K. S. Thygesen, C. Wolverton, M. Scheidgen, C. Toher, G. J. Conduit, G. Pizzi, S. Grazulis, G.-M. Rignanese, R. Armiento, Developments and applications of the OPTIMADE API for materials discovery, design, and data exchange, 2024. URL: http://arxiv.org/abs/2402.00572, arXiv:2402.00572 [cond-mat].

[17] F. Pezoa, J. L. Reutter, F. Suarez, M. Ugarte, D. Vrgoč, Foundations of json schema, in: Proceedings of the 25th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2016, pp. 263–273.

[18] H. Li, O. Hartig, R. Armiento, P. Lambrix, Ontology-based GraphQL server generation for data access and data integration, Semantic Web (2024). doi:10.3233/SW-233550.

[19] P. Lambrix, R. Armiento, H. Li, O. Hartig, M. Abd Nikooie Pour, Y. Li, The materials design ontology, Semantic Web 15 (2024) 481–515. doi:10.3233/SW-233340.

[20] N. Dimarcq, M. Gertsvolf, G. Mileti, S. Bize, C. W. Oates, E. Peik, D. Calonico, T. Ido, P. Tavella, F. Meynadier, G. Petit, G. Panfilo, J. Bartholomew, P. Defraigne, E. A. Donley, P. O. Hedekvist, I. Sesia, M. Wouters, P. Dubé, F. Fang, F. Levi, J. Lodewyck, H. S. Margolis, D. Newell, S. Slyusarev, S. Weyers, J.-P. Uzan, M. Yasuda, D.-H. Yu, C. Rieck, H. Schnatz, Y. Hanado, M. Fujieda, P.-E. Pottie, J. Hanssen, A. Malimon, N. Ashby, Roadmap towards the redefinition of the second, Metrologia 61 (2024) 012001. doi:10.1088/1681-7575/ad17d2, publisher: IOP Publishing.

[21] FAIRsharing Team, FAIRsharing.org: QUDT; Quantities, Units, Dimensions and Types, 2015. doi:10.25504/FAIRSHARING.D3PQW7, last Edited: Friday, May 6th 2022, 11:03, Last Accessed: Sunday, June 23rd 2024, 10:43,Last Reviewed: Monday, May 2nd 2022, 12:22.

[22] ISO 80000-1, ISO 80000-1:2022: Quantities and units, 2022. URL: https://www.iso.org/standard/76921.html.

[23] A. Thompson, B. N. Taylor, Guide for the Use of the International System of Units (SI). National Institute of Standards and Technology Special Publication 811 (2008 Edition), National Institute of Standards and Technology / U.S. Department of Commerce, 2008.

[24] G. Schadow, C. J. McDonald, J. G. Suico, U. Fohring, T. Tolxdorff, Units of Measure in Clinical Information Systems, Journal of the American Medical Informatics Association 6 (1999) 151–162. doi:10.1136/jamia.1999.0060151.

[25] M. Lefrancois, A. Zimmermann, The Unified Code for Units of Measure in RDF: cdt:ucum and other UCUM Datatypes, in: A. Gangemi, A. L. Gentile, A. G. Nuzzolese, S. Rudolph, M. Maleshkova, H. Paulheim, J. Z. Pan, M. Alam (Eds.), The Semantic Web: ESWC 2018 Satellite Events, Springer International Publishing, Cham, 2018, pp. 196–201. doi:10.1007/978-3-319-98192-5_37.

[26] H. Rijgersberg, M. Van Assem, J. Top, Ontology of units of measure and related concepts, Semantic Web 4 (2013) 3–13. doi:10.3233/SW-2012-0069.

[27] R. Raskin, M. Pan, C. Mattmann, Enabling Semantic Interoperability for Earth Science Data, NASA's Earth Science Technology Office (ESTO), 2004.

[28] J. Madin, S. Bowers, M. Schildhauer, S. Krivov, D. Pennington, F. Villa, An ontology for describing and synthesizing ecological observation data, Ecological Informatics 2 (2007) 279–296. doi:10.1016/j.ecoinf.2007.05.004.

[29] M. B. J. Mark Schildhauer, Oboe: the extensible observation ontology, version 1.2, 2016. URL: https://github.com/NCEAS/oboe/tree/OBOE_1_2. doi:10.5063/F1125R0F.

[30] OMG Systems Modeling Language 1.2, OMG Systems Modeling Language 1.2, Technical Report, 2010. URL: https://www.omg.org/spec/SysML/1.2.

[31] G. V. Gkoutos, P. N. Schofield, R. Hoehndorf, The Units Ontology: a tool for integrating units of measurement in science, Database (2012) bas033. doi:10.1093/database/bas033.

[32] D. Vrandevic, M. Krötzsch, Wikidata, Communications of the ACM 57 (2014) 78–85. doi:10.1145/2629489.

[33] E. B. Tadmor, R. S. Elliott, J. P. Sethna, R. E. Miller, C. A. Becker, The potential of atomistic simulations and the knowledgebase of interatomic models, JOM 63 (2011) 17–17. doi:10.1007/s11837-011-0102-6.

[34] D. Beckett, T. Berners-Lee, E. Prud'hommeaux, G. Carothers, RDF 1.2 Turtle, 2024. URL: https://www.w3.org/TR/rdf12-turtle/.

[35] I. Johansson, The mole is not an ordinary measurement unit, Accreditation and Quality Assurance 16 (2011) 467–470. doi:10.1007/s00769-011-0804-z.

[36] P. de Buyl, P. H. Colberg, F. Höfling, H5MD: A structured, efficient, and portable file format for molecular data, Computer Physics Communications 185 (2014) 1546–1553. doi:10.1016/j.cpc.2014.01.018.