Comparison and Evaluation of Ontologies for Units of Measurement

Jan Martin Keil [a,*], Sirko Schindler [a]

[a] *Heinz Nixdorf Chair for Distributed Information Systems,*
Institute for Computer Science, Friedrich Schiller University Jena, Germany
E-mails: jan-martin.keil@uni-jena.de, sirko.schindler@uni-jena.de

Abstract. Measurement units and their relations like conversions or quantity kinds play an important role in many applications. Thus, many ontologies covering this area have been developed. Consequently, for new projects aiming at reusing one of these ontologies, the process of evaluating them has become more and more time consuming and cumbersome. We evaluated eight well known ontologies for measurement units and the relevant parts of the Wikidata corpus. We automatically collected descriptive statistics about the ontologies and scanned them for potential errors, using an extensible collection of scripts. The computational results were manually reviewed, which uncovered several issues and misconceptions in the examined ontologies. The issues were reported to the ontology authors. This caused new bugfix releases in three cases.

In this paper we will present the evaluation results including statistics as well as an overview of detected issues. We thereby want to enable a well-founded decision upon the unit ontology to use. Further, we hope to prevent errors in the future by describing some pitfalls in ontology development—not limited to the domain of measurement units.

Keywords: measurement unit ontology, ontology comparison, ontology evaluation, ontology quality

1. Introduction

Units of measurement are an essential part in many aspects of modern life: The correct handling of the scale a value is measured in is crucial not only in science, but also in trade, industry and administration. A well documented use of units is especially important, when a project is carried out by different partners with different backgrounds. One of the most prominent examples of neglecting this fact is the crash of the Mars Climate Orbiter in 1999, which the NASA investigation board attributed to a mismatch of used units between two components of its software [1].

Similar integration challenges arise on an even larger scale in the context of Big Data: With the increasing need to integrate datasets of different origins, data annotation—preferably using Semantic Web techniques—gains importance. Using a machine readable annotation is essential for (semi)automatic discovery, verification, and integration.

As part of these semantic descriptions and to cover the field of measurement units and related concepts, over the last years several projects were initiated to create respective ontologies [2–5].

Most of these attempts were embedded in bigger research projects and, thus, catered to their specific needs. This led to a variety of different approaches to model the domain at hand. The created ontologies differ not only in the modeled subset of concepts and relations, but also in the type and number of units included. Engineers or researchers who wish to use an existing ontology in their work are now faced with the choice between several ontologies.

---

*Corresponding author. E-mail: jan-martin.keil@uni-jena.de.

To assist in this decision making process, we analyzed eight ontologies in the field of measurement units and the respective parts of the emerging Wikidata corpus [6]. Our analysis is focuses on the individuals and values in the ontologies, in contrast to earlier efforts that mostly focused only on the schemata. We used a collection of scripts to extract several kinds of statistics and to detect contradictions between the data sources. The manual review of the results confirms not only the different emphases of the ontologies, but also reveals the existence of several issues in all of them. Our contribution is as follows:

- We provide an extensible collection of scripts, which can be used by unit ontology developers to validate their work against other ontologies.
- We provide a mapping between the concepts of all selected ontologies.
- We identified multiple issues in existing ontologies and reported them to the respective authors. This triggered the publication of new releases of OM 1, OM 2, and SWEET to fix the issues found.
- We identified issue classes and hint towards preventive actions.
- We present the analysis of existing ontologies to support potential users in the decision process of selecting a unit ontology for reuse.

The paper is structured as follows: Section 2 will give an overview of the analyzed ontologies and present previous work on ontology evaluation. In Section 3 we will describe the general approach and some implementation details, before in Section 4 the various aspects of the analysis will be presented. Limitations of our approach will be discussed in Section 5. We conclude with suggestions to ontology authors to prevent some of the encountered issues in Section 6 and final remarks in Section 7. The terminology used will be specified in Appendix A.

2. Related Work

The discussion of related work is split into two parts: First, we will present the analyzed ontologies and specify the respective version we used. The second part will then describe previous work on ontology evaluation.

2.1. Ontologies for Measurement Units

Over the last years, several projects were initiated to create ontologies modeling the domain of measurement units. The following selection focuses on ontologies that model a reasonable number of relevant individuals and classes in an OWL compatible format. There are, of course, several other unit ontologies within and beyond the OWL world, but they do not meet this basic condition. In addition, the selection contains an OWL compatible knowledge base which provides data on measurement units.

The selection evaluated in this work includes:

- Measurement Units Ontology (MUO)[1]; result of a project to exploit semantics in mobile environments; the individuals were automatically generated from UCUM [7],
- Extensible Observation Ontology (OBOE)[2]; an ontology suite to represent scientific observations [2],
- Ontology of units of Measure and related concepts (OM 1)[3]; an ontology to model concepts and relations important to scientific research, developed in context of food research [3],
- Ontology of units of Measure (OM 2)[4]; second iteration of the OM ontology,
- Library for Quantity Kinds and Units (QU)[5]; a showcase ontology based on the OMG SysML 1.2 QUDV specifications and the UN/CEFACT Recommendation 20 code list [8],
- Quantities, Units, Dimensions and Data Types Ontologies (QUDT)[6]; developed in context of NASA projects[4],
- Semantic Web for Earth and Environmental Terminology (SWEET)[7]; originally developed in context of NASA projects, now maintained in a community project,

---

[1] `muo-vocab.owl` and `ucum-instances.owl` dated 2008 from http://idi.fundacionctic.org/muo/

[2] Version 1.0 from https://code.ecoinformatics.org/code/semtools/trunk/dev/oboe/

[3] Version 1.8.6 from http://www.wurvoc.org/vocabularies/om-1.8/

[4] Version 2.0.6 from https://github.com/HajoRijgersberg/OM

[5] `qu.owl` dated 2011-06-20 and `qu-rec20.owl` dated 2010-09-28 from https://www.w3.org/2005/Incubator/ssn/ssnx/qu/

[6] Version 1.1 from http://www.qudt.org/

[7] Version 3.1 from https://github.com/ESIPFed/sweet; earlier Version 2.3 from http://sweet.jpl.nasa.gov/

- Units of Measurement Ontology (UO)[8] [5] and Phenotypic Quality Ontology (PATO)[9]; both modules of the OBO family to model units and phenotypic qualities and
- Wikidata (WD)[10]; community-driven knowledge base of factual data for Wikipedia [6].

We are aware of the current efforts towards a second version of QUDT. At the time of writing, however, the ontology has only been partly published. The available parts do not allow a meaningful comparison. Thus, the analysis of QUDT 2 is left open for future work.

DBpedia extracts structured data from Wikipedia[11] and creates an interlinked knowledge base out of it [9]. At the time of writing, however, this extraction does not cover units. Therefore, DBpedia is not included into the selection.

On the other hand, Wikidata—a sister project to Wikipedia—aims at collecting factual knowledge for Wikipedia in a central repository [6]. This information can then be linked directly from different Wikipedia articles and thus provide a consistent view even across several language versions. Wikidata also includes a large number of units and their relations. As this data can also be accessed through a SPARQL endpoint, WD is included in the selection.

In the following "OM" will refer to both OM 1 and OM 2, when a statement applies to both ontologies alike. If only one ontology is affected, we will use "OM 1" or "OM 2" respectively.

The selected ontologies use different terms to refer to the same concepts. For example, the concept "kind of quantity" is denoted as "physical quality" (MUO), "quantity kind" (QU, QUDT), and "physical quantity" (WD). OBOE, OM, SWEET, and UO do not provide an explicit entity for this concept, but model the respective entities as subclasses of "physical characteristic" (OBOE), "quantity" OM, "property" (SWEET), and "quality" (UO). Therefore, a clear specification of the terminology used in this survey is needed. This is provided in Appendix A. Where possible, it is compliant with the International Vocabulary of Metrology [10].

2.2. Ontology Evaluation

Motivated by the lack of published evaluation processes as well as examination results for well known ontologies, an early evaluation approach was developed by Gómez Pérez [11]. This approach consists of two steps: In a first analysis step, the ontology is inspected regarding consistency, completeness, conciseness, expandability, and sensitiveness. *Consistency* is defined as the absence of contradictions between formal definitions and the modeled world, contradictions between informal definitions and the modeled world, contradictions between formal and informal definitions, and contradictions between formal definitions ("inferential consistency"). *Completeness* is informally defined as "each definition is complete" and "all that is supposed to be in the ontology" is included. *Conciseness* is the absence of redundancies and "unnecessary or useless definitions". *Expandability* is an assessment of the effort to add new definitions. *Sensitiveness* describes the impact of small changes to the whole ontology. In a second synthesis step the ontology is then corrected.

This approach was applied to the Standard Unit Ontology (SU), which contains 61 individuals and one class, by two ontology engineers and two domain experts. They used a checklist of eight possible errors in taxonomies to decide on its consistency, completeness and conciseness. As a result, Gómez Pérez concluded that SU is consistent and concise, but incomplete. Further, they identified seven problems by inspecting the ontology's source code. These errors were related to completeness, conciseness, expandability, and missing code conventions. However, the theoretical criteria were not consistently applied in the evaluation process. The choice of SU indicates a particular need for evaluation and improvement in the domain of measurement units.

Based on Gómez Pérez's criteria, Rijgersberg et al. evaluated five ontologies concerned with units of measurement [12]. They used a checklist of 34 requirements related to "completeness of the modeled scope", "quality of formal definitions", "understandability and extensibility", and "completeness in the natural language documentation". The results indicated that none of the ontologies contained all important concepts of this domain.

---

[8]Version 2017-04-24 from http://purl.obolibrary.org/obo/uo.owl
[9]Version 2017-03-22 from http://purl.obolibrary.org/obo/pato.owl
[10]Status as of 2017-06-27; using https://query.wikidata.org/
[11]https://www.wikipedia.org/

Consequently, they developed OM 1. Subsequently, Rijgersberg et al. published a comparison of OM 1 with QUDT 1.0.0 [3]. They compared the schemata of both ontologies and outlined the differences in the modeling. Further, they compared the number of kinds of quantities and measurement units as well as the support of five use cases. Overall, the results indicated that OM 1 outperformed QUDT 1.0.0.

Another domain specific comparison approach was introduced by Marcus P. Foster. He suggested a ranking of measurement ontologies in five levels regarding the supported concepts and use cases [13]. Ontologies of the *first level* "standardize characters for unit symbols and operators". On the *second level* they "additionally represent the concepts of unit and dimension" and "support the operations of unit conversion and unit conformance". On the *third level* they "additionally include the concept of quantity and system of quantities". On the *fourth level* they additionally provide a "full descriptions of quantities, prefixes, ordinal and nominal measurement scales, and systems of quantities and units". On the *fifth level* they "also incorporate the concepts (precision, accuracy, distribution etc.), relations and operations associated with calculating and representing measurement uncertainty". QUDT was classified as a third level and OM 1 as a fourth level ontology.

This ranking provided a fast overview regarding scope and level of development of ontologies. However, the order of requirements for each ranking-level seems biased by the authors background in parts. For example, an ontology might provide concepts of quantity and system of quantities, as required by the third level, without supporting unit conversion, as required by the second level.

Chau Do and Eric J. Pauwels applied MathML to map unit ontologies [14]. They manually created a predefined mapping of the seven International System of Units (SI) base units. Then the definitions of unit composition and unit conversion were used to map the remaining units. This approach was applied to OM 1, QUDT, and SWEET. During the evaluation of their mapping approach, they discovered several problems caused by "incorrect information in the ontologies". Some examples were provided for each ontology. However, as the work was not focused on ontology evaluation, they did not continue to systematically search for errors. Unfortunately, not all ontologies in our selection provide definitions of unit composition and unit conversion. Therefore, this mapping approach is not applicable in our analysis.

Samadian et al. compared the Ontology for Engineering Mathematics (EngMath), MUO, OM, and QUDT [15]. Tailored to the specific requirements of a medical data integration project, the comparison was focused on the conversion of units, the modeling of their mathematical relations, and the extensibility of the ontology. OM was the most suitable ontology for their purpose. But, due to the narrow focus, the benefit for other projects is limited.

In [16] a comprehensive list of 16 use cases in the field of unit ontologies was presented. For each use case, the required concepts and relations were identified and categorized into necessary and optional requirements. While necessary requirements are mandatory for a use case, optional requirements "simplify the implementation of a use case or increase its usefulness". A metric was developed to measure the suitability of a given ontology with respect to those use cases. Finally, a selection of seven ontologies (MUO, OBOE, OM 1, QU, QUDT, SWEET, and UO) was analyzed and compared using this metric. OM 1 surpassed the other ontologies in each use case with the closest overall contenders being QUDT, QU, and SWEET. This evaluation, however, remained on a conceptual level. The individuals of the ontologies have not been considered.

Zhang et al. tried to compare the four unit ontologies OM 1, QUDT, QUDV, and UO [17]. They focused on the OWL standard vocabulary used for modeling and possible reuse of the ontologies. However, their approach lacks a general methodology, which could be extended to include other ontologies in the comparison. Furthermore, the used OLW language features in an ontology do not enable any conclusions regarding its usefulness in other projects.

A domain independent approach of ontology comparison is the usage of general quality criteria as implemented by, e.g., OntoClean [18] and the OntOlogy Pitfall Scanner! (OOPS!) [19]. OntoClean provides a "methodology for validating the ontological adequacy and logical consistency" and "is based on highly general ontological notions" like *identity* and *unity*. It can be used to check for common pitfalls related to taxonomic relationships. On the other hand, OOPS! (semi)-automatically checks for many common pitfalls in OWL ontology development. For instance, this includes unconnected ontology elements or missing human readable annotations. It is easy to use and can give an initial impression of the quality. However, a detailed analysis of domain specific issues is out of the scope for both methods.

Although the comparison in [3] includes some statistics, all mentioned evaluation efforts are manual assessments almost exclusively remaining on a conceptual level or only cover general quality issues. Thus, there is a lack of evaluation concerning the individuals of unit ontologies.

3. Methods

The evaluation of unit ontologies on the level of individuals consist of two major aspects: One is *coverage*, which revolves around what concepts are modeled as well as how many and what kind of individuals are included. This corresponds to the particular aspect of the completeness criteria of [11] to define every concept "that is supposed to be in the ontology". The other one is *correctness*, which is concerned with how reliable the included values are in an application's context. It is related to all aspect of the *consistency* criteria of [11]. Further, it is related to particular aspects of the *completeness* and *conciseness* criteria [11], namely *completeness of definitions* and the *absence of redundancies*.

A strictly manual assessment of the selected ontologies is not appropriate. On the one hand, subtle differences like a missing digit in a 15 digit number, that we discovered using our approach, might be overlooked. On the other hand, the sheer number of concepts and relations to review might lead to a lapse in concentration and thus missing certain issues. Further, an automated approach ensures the identical evaluation of all selected ontologies and as well of new ontologies which might be added in the future.

A reasoner[12] was used to automatically test the inferential consistency. Where possible, we covered the other aspects automatically using a collection of scripts. In particular, that includes the gathering of statistics and detection of most issue classes. The creation process itself was iterative, as many of the issue classes only became apparent during the implementation of other detection or gathering processes. The assessment of most aspects, except inferential consistency and absence of redundancies, requires a comprehensive reference dataset.

3.1. Reference Dataset

Despite an extensive search, we are not aware of a comprehensive reference dataset in the field of measurement units. There are some datasets [7, 8, 21, 22], but those either cover only a subset of the units in use or omit some of the major concepts as listed in the appendix. Further, some of these datasets have been used to create some of the selected ontologies. Therefore, the results would likely be biased in favor of these ontologies.

For a suitable replacement, we turned to the information contained within the selected ontologies. We extracted relevant entities from the ontologies into an ontology-agnostic relational model (cf. Figure 1) and subsequently mapped them semi-automatically to one another. This resulted in a reference dataset for the further analysis.
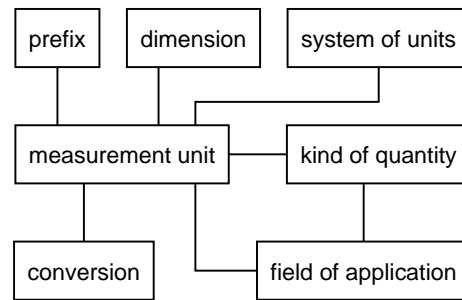

Fig. 1. Ontology-agnostic model of the created reference dataset.

The created dataset can be used to detect deviations between the ontologies. In most cases, these deviations can be backtracked to issues within at least one ontology.[13] This, of course, does not prove correctness of the remaining facts. Issues affecting objects present in only one ontology can not be detected. It is also possible, although unlikely, that all ontologies contain the same error. It is not possible to automatically decide which of the alternatives is correct and which erroneous. For example, two ontologies might provide different conversion factors for the same unit-pairs. Techniques from the field of data integration, like majority decisions or the ranking of data sources, do not ensure a correct result in every case. Therefore, they are not applicable in the context of an evaluation. For this, further manual intervention is needed along with possibly some literature research.

---

[12]Openllet 2.6.3 from https://github.com/Galigator/openllet; an open source fork of Pellet [20].

[13]The other major cause of deviations was a difference in precision of conversion factors.

3.2. Approach

Following the previous discussion, we decided against a strictly manual assessment and for a semi-automatic approach as shown in Figure 2. To apply it to the selected ontologies we created an implementation in the form of a collection of scripts. It is publicly available for free use and extension [23]. To easily be able to run just a subset of scripts, but yet preserve dependencies, each script will store its results into a set of JSON files, which will then be used by its successors.

In the following, we will describe each step in detail. The intermediate results of each step are available at [24]. Note that the received results were not only used to assess the given ontologies, but also to improve upon our analysis process. The results of the first analysis phase, e.g., allow to verify the extraction process and hint towards erroneous queries, which can then be corrected.

3.2.1. Extraction

The first step requires the extraction of individuals. For each ontology we formulated SPARQL queries to obtain the individuals per concept as well as the respective relations (cf. Figure 1). Those ontologies that are given in an OWL-compatible format are added to a Sesame triple store[14] before the corresponding set of queries is executed. An RDF-schema reasoner was used to simplify the extraction queries. For WD, on the other hand, an online SPARQL endpoint is provided[15], which we used in this phase.

Most ontologies only model a subset of the examined concepts, so some SPARQL queries could be omitted. The scripts interpret a missing query as a missing concept or relation in the specific ontology. This removes the need to explicitly state the absence of those concepts and relations. Furthermore, it allows to add queries for new ontologies one by one, instead of having to provide all queries at once. Note that we provided queries for each relevant concept and relation of the analyzed ontologies.

The result of the extraction is a unified, relational representation of the data contained per ontology, as described in Section 3.1. This removes all ontology-specific modeling techniques and ensures a common format for further analysis. As a side effect, the approach may also be used to compare against other datasets not necessarily provided as an ontology. To in-clude those datasets only an extraction process has to be provided, whereas all following steps are independent of the original source format.

In addition to the extraction of individuals and relations, the transitive closure of conversions for each ontology is computed in this step. This significantly expands the number of available conversions: If, for example, the conversions *milligram to gram* and *gram to kilogram* exist, the missing, but inferable conversion *milligram to kilogram* is added. As a consequence, a wider range of conversions can be validated later on as compared to only resorting to the conversions explicitly defined within the ontologies.

3.2.2. Intra-Ontology Analysis

The extracted data allows for a first analysis. This phase is called "intra-ontology analysis", as it is concerned with each ontology individually. Tests that involve multiple ontologies are subject of the "inter-ontology analysis" later on (cf. Section 3.2.4).

Most checks at this phase concern the *completeness of definitions* in the ontologies. This includes the following aspects:

- Is at least one label attached to each individual?[16]
- Are there individuals with multiple labels in the same language?
- Where applicable, is each individual part of at least one relation?
- Are all relevant values attached to each individual?
- Is each individual only assigned to a single concept?

The check for multiple labels per individual seems superfluous at first glance. However, this particular check allowed us to identify several occurrences of labels that were annotated with the wrong language tag.

Furthermore, a first set of statistics can be extracted at this point. The most prominent statistic here is the count of individuals and relations (cf. Table 1). Other examples include the usage of certain keywords like "per" and "reciprocal" to examine the uniformity in naming conventions used or the distribution of languages within the given labels (cf. Table 5).

---

[14]http://rdf4j.org/
[15]https://query.wikidata.org/

[16]This check is also part of the OOPS! pitfall catalogue: "P08. Missing annotations" [19].

```
┌──────────────┐    ┌──────────────┐    ┌──────────────┐    ┌──────────────┐    ┌──────────────┐
│              │    │ Intra-Ontology│   │              │    │ Inter-Ontology│   │              │
│  Extraction  │──▶ │   Analysis   │──▶ │   Mapping    │──▶ │   Analysis   │──▶ │    Manual    │
│              │    │              │    │              │    │              │    │    Review    │
└──────┬───────┘    └──────┬───────┘    └──────┬───────┘    └──────┬───────┘    └──────┬───────┘
       │                   │                   │                   │                   │
       ▼                   ▼                   ▼                   ▼                   ▼
 ╭───────────╮       ╭───────────╮       ╭───────────╮       ╭───────────╮       ╭───────────╮
 │Individuals│       │Statistics and│    │  Ontology │       │Statistics and│    │           │
 │per Concept│       │ poss. Errors │    │  Mapping  │       │ poss. Errors │    │Issue Lists│
 ╰───────────╯       ╰───────────╯       ╰───────────╯       ╰───────────╯       ╰───────────╯
```
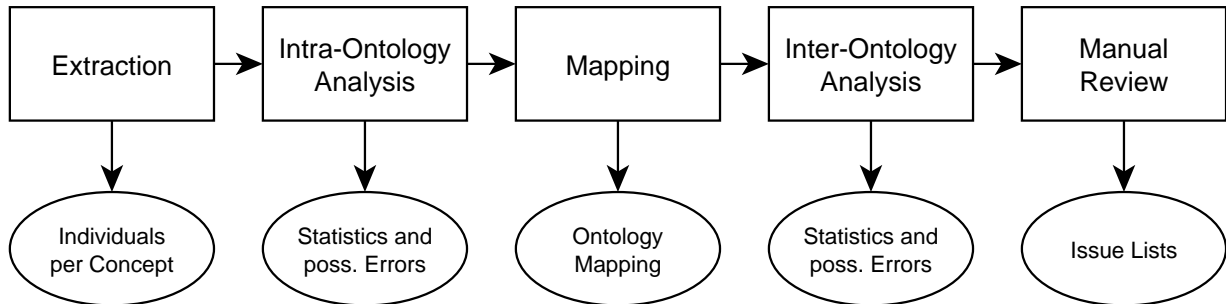
Fig. 2. Schematic of evaluation approach and the respective results.

3.2.3. Mapping

In the third step the individuals for each concept and relation are mapped to one another. Since, e.g., a unit must not be mapped to a dimension, the mappings work on each class of individuals separately. Besides reducing the computational effort, this allows to employ different mapping strategies. The result of the process are multiple mapping sets. Each of these sets includes all individuals that are mutually identical.

The mapping process uses three kinds of sources to determine whether two individuals are identical. The first source are `owl:sameAs` and other similar relations present in the ontology definitions. In the case of WD, this includes also mappings to QUDT and OM 1.

The second and most frequent sources of mappings are the individuals' labels. Whereas often a mere comparison by name might be sufficient, others like the units themselves need a more sophisticated approach. In the following we will highlight some of the challenges encountered and outline our solutions. The solutions are implemented using regular expressions unless noted otherwise.

A first obstacle was due to the differences in British and American spelling of some terms like "metre" vs. "meter". Similarly, different spellings exist for certain prefixes and other modifiers. The prefix "deca", e.g., also appears as "deka". All occurrences were replaced by a uniform variant.

Some ontologies are adding suffixes to the labels of individuals of some classes, which results in labels like "length dimension" or "Antal (jednotka)"[17]. We removed these stop words.

Another challenge was posed by different orderings of certain phrases. The label for the unit *square meter*, e.g., appears as "square meter" or "meter squared". For this and similar pairs like "cube" and "cubed" we

unify the order and spelling. A similar issue arises with the use of systems of units within unit labels. Examples are the two labels "US Survey Foot" and "Foot US Survey". We created a list of such modifiers and harmonized their position within the labels.

As noted before, the labels for reciprocal units include wordings using "per" and "reciprocal". We replaced all uses of "reciprocal" by "per".

The wide range of languages in Wikidata posed a particular challenge for the mapping. We found examples of units where the label in some languages was plain wrong: See, e.g., WD→Q25343, which represents the unit "square metre". The Hindi label घन मीटर, however, translates to "cubic meter".

On the other hand, some individuals have the same label although they depict different measurement units: See, e.g., WD→Q16830567 and WD→Q103510 both sharing the same label in German "Bar", but while one denotes the common unit for pressure the other one represents an old Persian unit of mass.

Another challenge in WD are duplicates with no overlap in the languages of their labels. See WD→Q2649540 and WD→Q10778756, which appear to reference to the same unit "Alma", but currently can not be matched as WD→Q2649540 provides only a German label while WD→Q10778756 only has a Czech one. Finding such duplicates is out of the scope of this paper.

As the other ontologies besides OM and WD only use English labels or omit the language tag altogether (cf. Table 5), we decided just to map measurement units based on three types of labels: English labels, labels without a given language and an artificial label created from the local name of the URI. The published code, however, also includes a mapping implementation that involves all languages which can be activated in the configuration.

---

[17] "Jednotka" is the Czech translation of "unit".

The manual review of the mappings uncovered several missed mappings. As adding specialized code for particular individuals would have made the codebase unnecessarily complex, we predefined certain mappings in separate files. This constitutes the final source of mapping information.

3.2.4. Inter-Ontology Analysis

At this step the reference corpus described in Section 3.1 is available. This allows for the creation of comparative statistics. Here, we focused on the overlap in unit-individuals between the different ontologies (cf. Tables 3 and 4). Similarly, we created a list of common measurement units[18] missing in particular ontologies.

The major results of this step, however, are aimed at the discovery of differences between the ontologies. Oftentimes these point towards issues present in the ontologies. A first example here is the list of conversions. To further expand the validation at this point, also the conversions as given by the prefix $\bowtie$ unit-relation are included. This and the aforementioned computing of transitive closure allows to compare a large fraction of the conversions present.

The comparison process provided one challenge: Some ontologies provide a more precise conversion factor than others. In a first, naïve comparison this fact is disregarded. This results in an extensive list of differences. Many of those are equivalent, however, if rounding values is assumed valid. So in a second comparison we applied a filter to reveal more substantial differences. That filter determines for two values if one is the rounded variant of the other within certain limits. The result only contains the differences that also fail this test.

Another check addresses the units that have varying definitions in different systems. To detect those we mapped the mapping sets of units again—this time ignoring the system modifiers in the labels. This results in sets of mapping sets. If one of at least two contained mapping sets has no system modifier within its labels, it is a candidate of a vague unit designation. To illustrate the approach here, assume the first mapping process created sets for *acre*, *acre (international)*, and *acre (U.S. survey)*. Removing the system modifiers results in a label of "acre" for all three of these sets, so they are mapped to one another by the second mapping. As one of the original sets' labels did

not include a system modifier, we probably identified one case of vague designations. On the other hand, if each of the original sets' labels would have included a system modifier, no issue had been raised.

Besides uncovering differences between ontologies, the mapping also allows to discover duplicates within one ontology. As the mapping algorithm is agnostic to the ontology an individual originates from, it also maps individuals from the same ontology to one another, if they fulfill the aforementioned criteria. Some of these potential duplicates are added to represent different names for the same unit, like "grade" and "gon" [22], and are properly connected by `owl:sameAs`. Others, however, are the result of typographical errors and different orderings of terms (e.g., "metric ton" and "ton metric") and should be fixed or harmonized.

The mapping was also used to identify prefixed units that were not flagged as such by their respective ontology. For units that have not been flagged in any ontology, we also used a heuristic to identify missing prefix associations. This heuristic checks the unit label for appearances of prefixes at the start of each word. That way, e.g., the prefix "kilo" will be detected in the units "kilometer" or "meter per kilogram" and they will be marked as prefixed. Additional refinements for some special cases, e.g., prevent the detection of "milli" in "part per million" and ensure it for "hecto" in "hectare". This heuristic leads to the further detection of missing relations as well as to a more conclusive comparison of the portion of prefixed units (cf. Tables 1 and 3).

3.2.5. Manual Review

The previous steps maintain their results in the form of JSON files. To ease the manual review, the final step will convert all results to HTML files. This includes, e.g., a list of mappings found for each concept, tables showing a statistical overview similar to the ones presented in Section 4.1, or lists of potential issues for various issue class.

For the review of some issue classes, it proved helpful to adapt the presentation of the results or even provide different representations. For example, multiple lists are created to spot mappings that are missed by the algorithm. One list shows the created mapping sets in a lexicographical ordering of the labels. Another list also shows the created mapping sets in an lexicographical ordering of the labels, but the words within each label are ordered lexicographically beforehand. For example, the labels "meter kilogram per second" and "kilogram meter per second" of equivalent measurement

---

[18]"Common" is here defined as being present in at least half of the analyzed ontologies.

units both turn into "kilogram meter per second". Each of these representations can assist in finding additional mappings.

It is important to highlight, that the results just contain statistics and *potential* issues. A manual review of these *potential* issues is indispensable to get a reliable list of issues for the evaluation. For example, the mentioned list with lexicographically sorted words within the labels will group "mole per kilogram" and "kilogram per mole" although they are different measurement units.

While reviewing the existing results oftentimes new issues are identified, which are not yet covered. During the evaluation process this occurred on different occasions. In order to ensure a fair comparison for all involved ontologies we added new scripts to check for these issues and repeated the evaluation process.

4. Results

The analysis of the examined ontologies using the collection of scripts provides a deeper insight into the ontologies with respect to broadness and completeness. Furthermore, it reveals several issues of different types and some pitfalls specific to unit ontologies. The raw results are publicly available [24]. In this section first a statistical overview of the ontologies will be discussed, followed by a description of discovered issue classes and an overview of issues per ontology.

The presented statistics are based on most recent releases of all ontologies at the time of writing. Note that Wikidata is constantly evolving and no specific version-numbers are issued. The numbers stated in the following for Wikidata will probably already have changed by the time of reading. They can, however, give a first impression about the amount of data managed by the project.

4.1. Statistics

For an initial overview the supported concepts and relations in the ontologies as well as the respective number of individuals are given in Table 1. Also included are the number of *non-prefixed* and *prefixed* units, which incorporate the results of the aforementioned heuristic to identify prefixed units.

The first notable observation is the absence of a connection between *prefixed units* and *prefixes* in MUO, QUDT and UO, even though both exist. First of all, despite containing both prefixed units and prefixes MUO,

QUDT and UO do not model the connection between both. This hinders attempts to retrieve multiples or submultiples of a given unit. Furthermore, individuals for the concepts of *field of application*, *dimension* and *system of units* are just included in small subset of the ontologies.

OM is the only ontology with two versions present in this analysis. Most improvements between both OM versions are concerned with the simplification of the ontology structure, which is outside of the scope of this paper, but is documented in the SPARQL queries provided [23]. An important observation in the comparison of both versions is the absence of almost 40% of the individuals for kinds of quantities and systems of units in OM 2. This also has a major impact on the number of app⋈qk and qk⋈unit relations. Remarkably, the numbers for system⋈unit remain almost the same. This suggests, that the removed individuals for systems of units were poorly connected anyways. In contrast to the absence of many kinds of quantities and systems of units, there is a considerable increase of app⋈unit relations, whose number has more than doubled.

The number of units in WD exceeds the number of units in the other ontologies by far. Manual inspection reveals a large number of historic units to be present in WD, which might explain this large number. Another reason are possible duplicates that were not recognized due to the reasons mentioned in Section 3.2.3. The second place is taken by OM, which does not suffer from these limitations. If the prefixed units are discarded, however, the situation changes in favor of QUDT. A closer look at the actual individuals in the ontologies suggests that a systematic way has been used to create the prefixed pendants for all SI units in OM, which vastly increased the total number of units.

In the next step, the individuals for each concept were mapped to one another. Table 2 compares the number of individuals present in the ontologies and the number of mapped distinct units overall. The number of found mappings especially for units is surprisingly low. To eliminate shortcomings of the mapping process itself the list of mapped units was carefully examined by hand, which led to a number of mappings added manually, but to no substantial overall change. Due to the wide range of languages we were not able to perform this manual duplicate search for WD to the same extent as for other ontologies. Therefore in some of the further analysis the values reported for WD are not as reliable as the ones reported for the other ontologies.

Table 1

Number of individuals per concept and number of relations between two concepts per ontology after extraction. (A ⋈ B denotes a relation between concept A and B.)

	MUO	OBOE	OM 1	OM 2	QU	QUDT	SWEET	UO	WD
field of application (app)	-	-	18	17	12	-[21]	49	-	-
conversion	-	64	828	832	674	55	37	-	603
dimension (dim)	-	-	68	76	-	130	-	-	-
prefix (prefix)	24	-	28	28	20	28	12	20	-
kind of quantity (qk)	78	67	879	549	598	235	847	324	1419
system of units (system)	-	-	13	8	-	11	-	-	35
measurement unit (unit)	267	138	1367	1384	801	768	140	221	4404
non-prefixed unit	264	71	371	388	433	652	111	110	4143
prefixed unit	3	67	996	996	368	116	29	111	261
app ⋈ qk	-	-	625	335	653	-	489	-	-
app ⋈ unit	-	-	410	992	-	-	-	-	-
dim ⋈ unit	-	-	1144	1157	-	491	-	-	-
prefix ⋈ unit	-	-	989	866	163	-	9	-	-
qk ⋈ unit	267	140	5157	4027	129	1093	1028	944	4177
system ⋈ unit	-	-	819	821	-	4124	-	-	166

Table 2

Comparison of the total number of individuals as present in all ontologies with the number of distinct objects after the mapping.

	field of application	dimension	prefix	kind of quantity	system of units	measurement unit
extracted	96	274	160	4996	67	9490
mapped	76	140	29	3644	51	6413

Following up on differences in the number of units included and the lack of mappings, we analyzed the frequency in which units appear in the different ontologies. Table 3 lists the number of prefixed and non-prefixed units which are present in a certain number of ontologies. The actual number of widely used units was less than anticipated as over 85 %[19] of all units appeared in just one ontology.

Also notable is the fact that there seems to be little consensus which units are essential to an ontology. Only 17 units are present in all ontologies.[20] This includes, e.g., no unit for *mass* as both *gram* (SWEET) and *kilogram* (MUO) are missing in one ontology.

To explore this even more, the overlap between the ontologies with respect to the included units is given in Table 4. It displays the fraction of both prefixed and non-prefixed units from one ontology, that are included in another.[22]

As can be seen, for any two given ontologies less than three quarters of the units of one ontology are part of the other. Even the smaller ontologies with respect to the number of units include units, that are not covered by the larger ones. There are different possible reasons for this fact. The ontologies were created for different projects or domains and hence needed to meet different demands regarding the included units. This is especially noticeable in the different sets of compound units included.[23] Another reason might be the inclusion of prefixed units. While on the one hand ontologies will include almost no prefixed units (MUO), others seem to have systematically added all prefixed variations for a given unit (OM). Most ontologies, how-

---

[19]Excluding WD lowers this number to 70 %.

[20]Units present in all ontologies: *ampere, candela, day, degree, degree Celsius, hour, kelvin, lumen, meter, minute, mole, newton, pascal, radian, second, siemens* and *volt*.

[21]Information is included in the ontology, but not explicitly modeled using a specific concept or relation.

[22]A similar table showing the same relation but just for non-prefixed units can be created using the provided scripts. Although the exact numbers differ of course, the general trend remains the same.

[23]In the absence of a respective modeling in most ontologies, exact numbers for the share of compound units could not reliably be extracted.

Table 3

Number of units occurring in a certain number of ontologies.

present in #ontologies		1	2	3	4	5	6	7	8	9	total
WD incl.	#units	4771	1037	275	136	73	50	33	21	17	6413
	#non-prefix units	4447	299	111	80	29	34	25	15	17	5057
WD excl.	#units	1086	1008	257	90	65	33	24	17	-	2580
	#non-prefix units	785	198	137	41	46	24	17	17	-	1265

Table 4

Percentage of units from one ontology (row) that are also present in another ontology (column).

	MUO	OBOE	OM 1	OM 2	QU	QUDT	SWEET	UO	WD
MUO	100.0	10.3	40.7	41.1	28.1	31.6	12.9	17.5	32.7
OBOE	19.7	100.0	67.2	67.9	67.9	46.7	27.7	39.4	68.6
OM 1	7.8	6.7	100.0	99.9	21.6	18.4	6.2	9.6	25.9
OM 2	7.8	6.7	98.6	100.0	21.8	18.8	6.2	9.5	25.9
QU	9.3	11.7	36.9	37.8	100.0	28.9	10.3	13.0	26.3
QUDT	10.9	8.4	32.9	34.1	30.2	100.0	10.8	9.8	42.0
SWEET	25.2	28.1	62.2	63.7	60.7	60.7	100.0	43.0	52.6
UO	20.8	24.4	59.3	59.3	47.1	33.9	26.2	100.0	52.0
WD	2.0	2.2	8.1	8.2	4.8	7.3	1.6	2.6	100.0

ever, choose a more relaxed approach adding some prefixed units, but not all of them.

WD provides a wealth of languages for its labels—totaling over 300 different language tags including multiple regions for some of the languages. Following this observation we were interested in the frequency of each individual language. An excerpt of this is given in Table 5. Note that just the five most frequent languages for each type are shown for WD, while all present languages are listed for the other ontologies. Most ontologies choose to only add English labels. When the language tags were missing, the label turned out to be in English in all cases examined (QU, QUDT, UO). The only exception for domain ontologies is OM which also added Dutch and Chinese labels for a subset of individuals. In WD English is not as dominant as in other ontologies: Considering just the subset of analyzed individuals English is surpassed by German and rivaled at least by Russian in numbers. On the other end of the spectrum are OBOE and SWEET which omit labels altogether for most of their individuals.

The final examined aspect is the coverage of relations. It is important not only to include individuals for each type, but also to link them together in an appropriate manner. All units, e.g., could be linked to at least one kind of quantity to represent what physical property can be measured using a particular unit. That statement, however, does not hold true for all relations: Not all units can be linked to a matching prefix as there

is no prefix for non-prefixed units. This also shows that both types within a single relation have to be considered individually in terms of coverage. Hence, in the example for each prefix there should at least be one unit be linked to it. The current state of the ontologies is shown in Table 6.

Note that a frequency of 1.00 does not suggest that all possible links are included, but only that each individual is part of at least one link. To illustrate this restriction, assume an ontology includes just two units EX→meter and EX→foot, as well as two kinds of quantities EX→Breadth and EX→Height. To achieve a frequency value of 1.00 this ontology only needs to include two relations like, e.g., (EX→meter, EX→Breadth) and (EX→foot, EX→Height). The other two possible relations would not be needed, and their absence would not be detected by this metric. An optimal value of 1.00 was only achieved in a few cases. On the contrary, the relation coverage within the different ontologies varies quite a lot. Noteworthy are especially those entries that suggest no link at all, despite both types being modeled. MUO, QUDT and UO, although having both prefixes and units, did not choose to to link them at all. Similarly, QU and SWEET did not link field of application and units. This begs the question why prefixes or fields of application were included in these ontologies in the first place. Other relations are usually given for just a subset of individuals with just few exceptions of a full coverage.

Table 5

Labels per language as present in the ontologies. Only the most frequent languages are shown for WD.
(Σ total number of individuals; $\varnothing$ missing labels; $\Diamond$ labels without language tag)

	MUO	OBOE	OM 1	OM 2	QU	QUDT	SWEET	UO	WD
field of application			Σ: 18 en: 18 nl: 18	Σ: 17 en: 17 nl: 17	Σ: 12 $\varnothing$: 12		Σ: 49 $\varnothing$: 49		
dimension			Σ: 68 en: 68 nl: 15	Σ: 76 en: 76 nl: 15		Σ: 130 $\Diamond$: 130			
prefix	Σ: 24 en: 24		Σ: 28 en: 28 nl: 28 zh: 15	Σ: 28 en: 28 nl: 28 zh: 15	Σ: 20 $\Diamond$: 20	Σ: 28 $\Diamond$: 28	Σ: 12 $\varnothing$: 12	Σ: 20 $\Diamond$: 20	
kind of quantity	Σ: 78 en: 78	Σ: 67 $\varnothing$: 62 en: 5	Σ: 879 en: 879 nl: 256 zh: 24	Σ: 549 en: 549 nl: 228 zh: 26	Σ: 598 $\Diamond$: 598	Σ: 235 $\Diamond$: 235	Σ: 847 $\varnothing$: 847	Σ: 324 $\Diamond$: 324	Σ: 1419 en: 1313 de: 1073 ru: 1014 fr: 1011 nl: 929
system of units			Σ: 13 en: 13 nl: 13	Σ: 8 en: 8 nl: 8		Σ: 11 $\Diamond$: 11			Σ: 35 en: 31 de: 28 es: 28 ru: 24 ja: 23
measurement unit	Σ: 267 en: 267	Σ: 138 $\varnothing$: 138	Σ: 1367 en: 1367 nl: 1214 zh: 104	Σ: 1384 en: 1383 nl: 1224 zh: 106 $\varnothing$: 1	Σ: 801 $\Diamond$: 801	Σ: 768 $\Diamond$: 768	Σ: 140 $\varnothing$: 140	Σ: 221 $\Diamond$: 221	Σ: 4404 de: 3227 en: 1824 ru: 1423 es: 1151 nl: 1142

4.2. Encountered Issue Classes

The analysis revealed several classes of issues affecting the ontologies. Some are relevant to all kinds of ontologies, others are specific to unit ontologies. In the following we describe common issue classes encountered, except the most obvious classes like missing, wrong or duplicated values, assignments or entities. Note that each of these classes only applies to a subset of ontologies. An ontology specific discussion will follow in the next section.

Entity IRIs contain ontology version. The IRIs of entities contain the version number of the ontology. This is in contrast to the idea of Linked Data [25] as an update of the ontology will break references by other ontologies or systems. The version number should be contained in an additional version-specific IRI of the ontology [26].

Label concatenation. A single label contains the concatenation of multiple alternative names of a single entity, e.g., the label "tonne (metric ton)" actually represents two labels: "tonne" and "metric ton". This concatenation might become a problem in automation and should be replaced by multiple or alternative labels.

Mix-up of equivalent and identical units. A conversion of factor one and offset zero does not imply two unit individuals to be the same. A unit also consists of other properties like the use within a system of units or the assignment to a kind of quantity. Consider the following:

```
Individual: :cubicDecimetre
 SameAs:
  :liter
 Facts:
  :unitOfSystem :SI
```

Table 6

Individuals present in at least one relation.
Total represents the number of individuals of that type, while the respective frequencies are given as fraction of that total number.

		MUO	OBOE	OM 1	OM 2	QU	QUDT	SWEET	UO	WD
field of application	total	-	-	18	17	12	-	49	-	-
	⋈ kind of quantity	-	-	1.00	1.00	1.00	-	0.14	-	-
	⋈ measurement unit	-	-	0.28	0.94	0	-	0	-	-
dimension	total	-	-	68	76	-	130	-	-	-
	⋈ measurement unit	-	-	0.99	1.00	-	0.95	-	-	-
prefix	total	24	-	28	28	20	28	12	20	-
	⋈ measurement unit	0	-	1.00	1.00	0.75	0	0.67	0	-
kind of quantity	total	78	67	879	549	598	235	847	324	1419
	⋈ field of application	-	-	0.67	0.58	0.98	-	0.46	-	-
	⋈ measurement unit	1.00	0.51	0.62	0.95	0.21	0.74	0.55	0.30	0.10
system of units	total	-	-	13	8	-	11	-	-	35
	⋈ measurement unit	-	-	0.62	1.00	-	0.73	-	-	0.63
measurement unit	total	267	138	1367	1384	801	768	140	221	4404
	⋈ field of application	-	-	0.25	0.65	0	-	0	-	-
	⋈ dimension	-	-	0.84	0.84	-	0.64	-	-	-
	⋈ kind of quantity	1.00	1.00	0.77	0.77	0.16	0.98	0.61	0.76	0.90
	⋈ system of units	-	-	0.56	0.55	-	0.73	-	-	0.03

This would erroneously imply that liter is an SI unit, which is wrong [22]. Similarly, a fraction is not identical to its simplified form (e.g., $\frac{m^3}{m^2}$ is not identical to m) as discussed in [27].

Namespace proliferation. An ontology or ontology family contains multiple namespace definitions for the same or similar namespaces or the same prefix is used for multiple namespaces. Assume the following two ontology files:

```
a.owl
Prefix ref: <http://example.org/b#>

b.owl
Prefix a:   <http://example.org/a>
Prefix ref: <http://example.org/a#>
```

The namespaces in `b.owl` differ only in whether or not to include the tailing #. Furthermore `ref` is used for two completely different namespaces.

SPARQL incompatible namespace. The namespace may be separated from the local name of an entity by using #, but the namespace definition does not include the #. Thus the namespace can not be reused in SPARQL as # is reserved for comments in SPARQL and has to be escaped each time a entity is referred using the namespace prefix.

Vague designation of units. The existence of several units of the same name in different systems of units has been ignored. For example, *acre* might refer to *acre (international)* as well as *acre (U.S. survey)*, which represent different sizes of an area [7]. The respective systems of units should be explicitly declared or alternatively, the conversions of a unit should be defined with respect to a specific system of units as done in WD. The disadvantage of the second option would be the lack of an IRI for a unit of a certain system of units and therefore impedes referring them.

Unspecified language. An ontology provides literals containing natural language without identifying the used language with a language tag. In automated processing this may cause language mixups if ignoring those literals is not an option.

Depending on the usage context of an ontology or the specific task it was created for, some of the listed issues might not be considered problems at all. For example, concatenated labels might not do harm or even be an advantage, if the ontology is just used for manual labeling. However, if the same ontology is used for automatic annotation, it is hard to work around that issue.

4.3. Ontology Assessment

A full list of issues found was sent to the authors of the respective ontologies. They were also asked about

the status of the development, the intended handling of the issues and were asked for permission to report on their answer. Helpful responses were provided by authors of OM, QUDT, SWEET, and UO. They affirmed the ongoing development of their ontologies. The authors of OM fixed the reported issues immediately. Furthermore, the new community project maintaining SWEET has also started to handle the reported issues. Similarly, the authors of UO have fixed most issues, but so far did not provide a new release. All the other ontologies at the time of writing still contain the reported issues. As WD is a collaborative project we did not report discovered issues to specific authors. However, a more detailed verification and correction will be subject of future work.

In the following we discuss each ontology with regard to encountered issues. If not stated otherwise, the ontologies turned out to be inferential consistent.

MUO is one of the smaller ontologies and almost exclusively includes non-prefixed units. Since the data was imported from UCUM [7], most of the issues seem to be shortcomings of the import process or UCUM itself:

In six labels the Unicode symbol #160 ("no break space") was used instead of a space (e.g. in "calorie at 15 °C"). This might cause problems when comparing labels to different inputs. We found two most likely unintended unit duplicates (MUO→liter-2 and MUO→the-number-ten-for-arbitrary -powers-2). Although UCUM defines both MUO→ calorie and MUO→British-thermal-unit just as aliases to other units, the ontology includes them as independent units. A typing error has been recognized in the IRI and the label of MUO→pound-per -sqare-inch. The entity MUO→horsepower is a case of vague designation.

OBOE is the smallest unit ontology included in the evaluation. Nevertheless, we detected a substantial number of issues. Most critical, ten conversions contained wrong factor or offset values. In eight of these cases the direction of the conversion has been mixed up resulting in an inverse conversion factor. In case of OBOE→NanogramToKilogram, this would, e.g., result in a calculation error of factor 10^{24}. Less critical, one conversion (OBOE→HectopascalToPascal) has been defined without any conversion values. A typing error occurred in the IRI of OBOE→Milliliter PerLIter. Further, we found 14 units with a vague designation (e.g., OBOE→Acre and OBOE→ Gallon). One conversion occurred twice (OBOE →AcreToMeterSquared and OBOE→AcreTo

SquareMeter). Three times equivalent and identical units have been mixed up, e.g. OBOE→MeterCubed and OBOE→Kiloliter have been connected using owl:sameAs but should only have a conversion with factor one and offset zero.

OM is the largest unit ontology. Thus, a larger number of issues is to be expected. However, OM stands out as most reported issues below have been fixed immediately. Due to further development and the resolving of reported issues, updated releases have been published by the authors (1.8.6 and 2.0.6 respectively).

We started our analysis with version 1.8. As of the latest version, 2.0.6, the following issues remain to be addressed:

All IRIs contain the version number of the ontology. Four entities classified as units should be scales instead. For example, OM→_1-10 should by a OM→Ordinal_scale. Five temperature scale individuals were connected to the according temperature scale class by owl:equivalentClass which makes them individuals of themselves. Six temperature quantity classes erroneously contained the according temperature scale classes in their definition. For example, the definition of OM→ CelsiusTemperature contained the following statement:

```
Class: om:CelsiusTemperature
 SubClassOf:
  om:hasValue only (om:usesUnit only
   (om:CelsiusTemperatureUnit or
    om:CelsiusTemperatureScale))
```

That caused the temperature scale classes to be a subclass of OM→Unit.

Further, we found the following issues that have been fixed in the current version (2.0.6). Most of these issues have also been fixed in the current version of OM 1 (1.8.6).

OM 2 was inferential inconsistent due to missing datatype declarations of property values of OM2→hasFactor. Therefore, OM 2 was not usable with a reasoner. Two conversions contained wrong factors. For example, the factor between OM1 →metre_of_mercury and OM1→pascal was 1.33322e2 instead of 1.33322e5. Three dimension individuals had wrong dimension vector values. For example OM1→thermal_resistance -dimension used the exponent −2 instead of 0 for the length component. Nine typing errors in labels, comments, symbols, and alternative symbols have been found like "reciprocal metre kilogram" instead of "kilogram second to the power -2" in OM1→

`kilogram_second-time_to_the_power_-2`.
Seven entity IRIs contained typing errors like in
OM1→`SI_luminous_intensity_dimension`
`_dimension_exponent` a duplicated "dimension"
breaking references to the entity. Four duplicated units
were included like OM1→`metre_to_the_power`
`_-2`, which is a duplicate of OM1→`reciprocal`
`_square_metre`. We recognized 18 units with a
vague designation like OM1→`foot`. The unit OM1→
$g \cdot m^{-2} \cdot m^{-1}$ was missing m^{-1} in its composition def-
inition. Two units contained a wrong dimension like
OM1→`reciprocal_kelvin` with dimension OM1
→`thermodynamic`
`_temperature-dimension`. Nine labels were
falsely tagged English instead of Dutch like "Gaussis-
che eenhedenstelsel" on OM1→`Gaussian_system`
`_of_units`. The property OM2→`hasFactor` con-
tained the class OM2→`Prefix` as an additional do-
main that caused the most units to be a prefix too.
The property OM2→`hasUnit` instead of OM2→
`commonlyHasUnit` was used to assign the unit OM2
→`are` and the quantity kind OM2→`Area`.

QU is one of the medium sized ontologies. We
found in QU→`xo15degCCalorie` one wrong con-
version factor. We also recognized six units with
concatenated labels, e.g., in the label "tonne (metric
ton)" of QU→`tonneMetricTon`. The entity QU→
`micrometreMicron` contains a mix-up of equiv-
alence and identity of units and should be separated,
since "micrometre" is a SI unit, but "micron" is not
[21]. Three units (QU→`reciprocalSquareMetre`
and QU→`perSquareMetre`, QU→`dryQuartUs`
and QU→`quartUsDry`, QU→`dryPintUs` and QU
→`pintUsDry`) occurred twice in QU. Further, we
found 14 units with vague designation.

QUDT is also one of the medium sized ontologies.
The most important issue of QUDT is its inferential
inconsistency. For example, it contains the following
definitions:

```
Class: qudt:Unit
 SubClassOf:
  qudt:typePrefix exactly 1 owl:Thing,
  qudt:typePrefix value "U"^^xsd:string
Class: qudt:ResourceUnit
 SubClassOf:
  qudt:Unit
Class: qudt:FinancialUnit
 SubClassOf:
  qudt:ResourceUnit,
  qudt:typePrefix value "UF"^^xsd:string
```

This means that every individual of the class
`FinancialUnit` has the type prefix "UF". More-
over, because it is also an individual of the class `Unit`,
it has the type prefix "U". As the number of type pre-
fixes is, however, limited to one, this results in a con-
flict. As a result QUDT can not be used with a reasoner.
Also critical, three conversions and two prefixes con-
tained wrong factors. For example, the factor of the bi-
nary prefix QUDT→`Pebi` was 125899906842624
instead of $\underline{1}$125899906842624 (2^{50}). We found
four duplicates of units, like QUDT→`Dalton2` and
QUDT→`AtomicMassUnit` for QUDT→`Dalton`
[21]. Further, QUDT$\xrightarrow{ind.}$`AtomicMassUnit` was clas-
sified as QUDT→`AtomicPhysicsUnit` but should
be a QUDT$\xrightarrow{class}$`AtomicMassUnit`[24]. Typing errors
occurred in QUDT→`RadianPerMinute`, which was
labeled "radians per second" and in the namespace of
QUDT→<u>schema</u>/quantity#`Mass` which should
be QUDT→<u>vocab</u>/quantity#`Mass`. We found 38
units with vague designation.

SWEET contains only a few units more than the
smallest ontology in our test, but likewise contains a
lot of issues:

The prefix SWEET→`micro` has no symbol. There
are two duplicates like SWEET→`FTU` for SWEET
→`FormazinTurbidityUnit`. Three conversions
contain a wrong factor or a wrong offset. Two of
them used the inverse values. The units SWEET→
`FTU` and SWEET→`FNU` are marked as identical but
are not. Further, a mix-up of equivalence and identity
occurred in case of SWEET→`hertz` and SWEET→
`perSecond` as well as SWEET→`micrometer` and
SWEET→`micron`.

Further, we found the following general issues,
that have been fixed in the current version (3.1):
Most adversely, the SWEET unit module file
`reprSciUnits.owl` did not contain valid RDF[25],
due to a duplicated ID. For the further evaluation, we
removed one of the duplicates (line 635–638) to get a
valid file. A second general issue was the namespace
proliferation and the SPARQL incompatibility of some
namespace definitions.

UO contained the following issues: The most im-
portant issue regarding UO is the lack of a strin-
gent version management. Although most of the is-
sues have been fixed or the erroneous concepts have

---

[24]QUDT contains two entities with this name: an individual
`http://qudt.org/`<u>vocab/unit</u>`#AtomicMassUnit` and a
class `http://qudt.org/`<u>schema/qudt</u>`#AtomicMassUnit`

[25]RDF Validation: https://www.w3.org/RDF/Validator

been removed in the development repository[26] the updates were never made available at the location referenced by the ontology's IRI. The development version for example contains most missing links between unit and prefixes, as described in Section 4.1, whereas they are missing in the released version. Our analysis is based on the released version as we believe most users would resort to this one. Four wrong subclass assignments have been found. For example, `UO→UO_0000325` ("megaHertz") is subclass of `UO→UO_0000106` ("hertz"). We found nine violations of the UO naming convention as applied to all other entities of the same type. E.g. the class `UO→UO_0000219` containing units related to electric charge is named "electric charge" instead of "electric charge unit". Further, UO uses a special kind of annotation properties[27] to categorize classes as units or groups of units. We found 23 cases of missing or wrong assignments of this entity classification. We also found a heterogeneous usage of the property `UO→hasExactSynonym` for alternative names and unit symbols. Though this is compliant to the definition of the property, it might become challenging in applications.

WD Please note that our analysis of WD was not as detailed as the analysis of the other ontologies. This will be part of future work. However we found the following issues:

WD contains links to entities from OM 1 and QUDT. The links to QUDT, however, point to the URLs of an HTML representation of QUDT, but do not refer the entities themselves. The references to OM 1 contain one wrong entry: The unit `WD→Q734439` ("hour angle") points towards `OM→Hour_angle`, which is a kind of quantity, but should point to `OM→hour-hour_angle`. The conversion of "1 E-16 s" to "second" contained a wrong precision. Eight conversions contained a wrong factor like from `WD→Q21014455` ("metre per minute") to `WD→Q182429` ("meter per second") with the factor 0.166666667 instead of 0.0166666667. Further, the conversion offset in the conversion of "degrees Celsius" to "Kelvin" was missing.

5. Limitations

The evaluation of the selected ontologies is based on the detection of deviations between the ontologies. Issues affecting objects that are only present in one ontology can thus not be detected. Similarly, issues that affect all ontologies will remain unnoticed.

The inevitable manual review of the potential issues carries the risk of human error, which might lead to missing out on some issues or false positives. To expose false positives, we requested the ontology authors to comment on the issues. However, for some ontologies we were unable to obtain substantial responses from the respective authors. So we could not discuss some issues in detail with them.

The decision upon the correctness of values is difficult in the case of conversion factors. Though it is usually easy to identify a number that is way off, certain differences in rounded values are harder to spot and verify. Some stem from minor deviations in the definitions for units of the same name, but belonging to different systems of units. Others, however, can be traced back to varying degrees of precision used by the ontology authors. Furthermore, as for some conversions no precise value can be given using floating point arithmetics[28], it remains an open issue whether those values are correct in a strict sense.

As previously noted, the coverage of concepts and individuals by the different ontologies is diverse. Therefore, the overlap for some concepts is limited. As a consequence, some characteristics that might be of interest in the ontology selection process could not be acquired to a meaningful degree and hence are left out for the time being. Examples include *dimensions*, the distribution of units over the various *fields of application* or the number of compound units. In conjunction with the relatively small overlap in units as shown in Table 4, this would not result in any reliable comparison of the different ontologies.

In the development of the implementation, some errors were discovered which did not lead to a separate automatic script. For example, some natural language descriptions were attached to the wrong individual or contained typos and wrong information. Natural language processing would be required to automatically detect this, which is outside the scope of this work.

---

[26]https://github.com/bio-ontology-research-group/unit-ontology/blob/master/uo.owl

[27]The approach, which is also used in the Gene Ontology, is explained on http://geneontology.org/page/go-slim-and-subset-guide.

---

[28]E.g., the conversion factor between *degree Fahrenheit* and *Kelvin*, which equals $\frac{5}{9}$ or between *degree* and *radian*, which is $\frac{\pi}{180}$.

6. Suggestions

The evaluation of the nine ontologies or knowledge bases uncovered several modeling pitfalls in the measurement unit domain. In this section we try to give suggestions on how to avoid them.

Careful design and documentation of conversion properties helps both users and developers to prevent misinterpretations of conversions. Most ontologies employ two properties for the factor and the offset in relation to an implicit or explicit reference unit to define a conversions. Unfortunately, four reasonable interpretations of those values are possible, as shown in Figure 3.

Given that in most cases the offset value is zero, the problem is usually reduced to two reasonable interpretations. However, two ontologies (OBOE and SWEET) used more than one interpretation within their conversion definitions. This highlights the importance of a proper definition and its documentation for error prevention and later usage. Ideally, the modeling approach allows only one reasonable interpretation. The authors of OM 1 employed definition individuals to model conversion that allow only one interpretation:

```
Individual:  :litre
 Facts:
  :definition  :_1.0e-3_cubic_metre
Individual:  :_1.0e-3_cubic_metre
 Facts:
  :unit_of_measure_or_measurement_scale
                            :cubic_metre,
  :numerical_value  "1.0e-3"
```

However, in favor of a leaner structure, the additional definition individuals have been removed in OM 2.

Precision and numerical stability of conversions is also of concern. In some cases, a simple change of the direction of the modeled conversion can improve the accuracy (e.g., 1.8 instead of $0.\bar{5}$ for the conversion between *degree Fahrenheit* and *Kelvin*). If that is not possible, one could think about representing problematic values with fractions given by numerator and denominator instead of a single floating point number. Besides this issue, almost none of the considered ontologies provides information about the precision of conversion values, which again makes it hard for a potential user to evaluate the adequacy for the given use case. Only WD allows to define the precision of values by stating a lower and an upper bound.

Naming conventions, particularly for the URI, can prevent duplicates and simplify usage. This corresponds to [11]. The denominator of a fraction for a derived unit can be named in several ways: Using "per", "reciprocal", "inverse" or "to the power of -1". Furthermore, there might be one modifier per part of the divisor or just one modifier for the whole divisor. Similarly, for exponents the terms "squared" and "cubed" or the general phrase "to the power x" can be used. Moreover the order of base and exponent, as well as the order of all factors or the position of an addition (e.g., the system of units in "U.S. survey mile" or the kind of quantity in "second (angle)") can be ambiguous. This highlights the need for a well defined and enforced naming scheme to prevent multiple individuals for the same unit. We do not suggest to assign only one label per entity, but provide different labels and adhere to a systematic approach. Multiple names should never be assigned using only one (concatenated) label.

7. Conclusion

We presented the first in-depth evaluation of unit ontologies considering individuals and included facts. We evaluated eight major ontologies for measurement units and the relevant parts of the Wikidata corpus. The gathered information is important for projects that intend to use a unit ontology as it accelerates the decision making process in selecting a suitable ontology.

The evaluation is based on an extensible collection of scripts. Due to the modular nature of the implementation, it can easily be extended by adding further scripts for new checks and comparisons or by adding other ontologies to be evaluated. The implementation [23] and the presented results [24] are publicly available. Every interested party is welcome to use, review or improve them. This way, both developers as well as users of unit ontologies now and in the future can use these results to choose or improve their ontologies.

The analysis provided not only statistical information about the ontologies, but also highlighted several issues within. The respective ontology authors were contacted and this led to at least three new releases of the analyzed ontologies which fixed the reported issues. As a byproduct, a mapping between the evaluated ontologies was created that can be used to integrate or migrate between the involved ontologies.

During the analysis we identified several classes of issues concerning the development of ontologies—not limited to the domain of measurement units. We hope

$$value\ stated\ in\ a = value\ stated\ in\ b \times factor + offset$$

$$value\ stated\ in\ a = (value\ stated\ in\ b + offset) \times factor$$

$$value\ stated\ in\ a \times factor + offset = value\ stated\ in\ b$$

$$(value\ stated\ in\ a + offset) \times factor = value\ stated\ in\ b$$

Fig. 3. Reasonable interpretations of conversion factor and offset.

that the characterization of this issue classes will help ontology authors to prevent them in future.

Our future work will focus on a more detailed analysis of Wikidata. Major challenges are the detection of wrong labels in the wealth of more than 300 different languages and the mapping of identical entities with disjoint label language sets.

Acknowledgements

References

[1] Mars Climate Orbiter Mishap Investigation Board Phase I Report, Technical Report, NASA, 1999.

[2] J. Madin, S. Bowers, M. Schildhauer, S. Krivov, D. Pennington and F. Villa, An ontology for describing and synthesizing ecological observation data, *Ecological Informatics* **2**(3) (2007), 279–296. doi:10.1016/j.ecoinf.2007.05.004.

[3] H. Rijgersberg, M. van Assem and J. Top, Ontology of Units of Measure and Related Concepts, *Semantic Web* **4**(1) (2013), 3–13. doi:10.3233/SW-2012-0069.

[4] R. Raskin, Semantic Web for Earth and Environmental Terminology (SWEET), in: *Earth Science Technology Conference*, 2003.

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

[6] D. Vrandečić and M. Krötzsch, Wikidata: a free collaborative knowledgebase, *Communications of the ACM* **57**(10) (2014), 78–85. doi:10.1145/2629489.

[7] G. Schadow and C.J. McDonald, The Unified Code for Units of Measure, 2014, visited on 26 Apr. 2016. http://unitsofmeasure.org/ucum.html.

[8] United Nations Economic Commission for Europe (UNECE), Codes for Units of Measure Used in International Trade, Recommendation, 20, UN/CEFACT Information Content Management Group, 2009, Revision 6.

[9] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P.N. Mendes, S. Hellmann, M. Morsey, P. Van Kleef, S. Auer et al., DBpedia–a large-scale, multilingual knowledge base extracted from Wikipedia, *Semantic Web* **6**(2) (2015), 167–195. doi:10.3233/SW-140134.

[10] Joint Committee for Guides in Metrology, *International Vocabulary of Metrology*, 3rd edn, 2012.

[11] A. Gómez-Pérez, Evaluation of ontologies, *International Journal of Intelligent Systems* **16**(3) (2001), 391–409. doi:10.1002/1098-111X(200103)16:3<391::AID-INT1014>3.0.CO;2-2.

[12] H. Rijgersberg, M. Wigham and J. Top, How semantics can improve engineering processes: A case of units of measure and quantities, *Advanced Engineering Informatics* **25**(2) (2011), 276–287. doi:10.1016/j.aei.2010.07.008.

[13] M.P. Foster, Quantities, Units and computing, *Computer Standards & Interfaces* **35**(5) (2013), 529–535. doi:10.1016/j.csi.2013.02.001.

[14] C. Do and E.J. Pauwels, Using MathML to Represent Units of Measurement for Improved Ontology Alignment, in: *Intelligent Computer Mathematics - MKM, Calculemus, DML, and Systems and Projects 2013, Held as Part of CICM 2013, Bath, UK, July 8-12, 2013. Proceedings*, Lecture Notes in Computer Science, Vol. 7961, Springer, 2013, pp. 310–325. doi:10.1007/978-3-642-39320-4_21.

[15] S. Samadian, B. McManus and M.D. Wilkinson, Automatic detection and resolution of measurement-unit conflicts in aggregated data, *BMC Medical Genomics* **7**(1) (2014), 12. doi:10.1186/1755-8794-7-S1-S12.

[16] M.D. Steinberg, S. Schindler and J.M. Keil, Use Cases and Suitability Metrics for Unit Ontologies, in: *OWL: Experiences and Directions - Reasoner Evaluation - 13th International Workshop, OWLED 2016, and 5th International Workshop, ORE 2016, Bologna, Italy, November 20, 2016, Revised Selected Papers*, M. Dragoni, M. Poveda-Villalón and E. Jiménez-Ruiz, eds, Lecture Notes in Computer Science, Vol. 10161, Springer, 2016, pp. 40–54. ISBN 978-3-319-54626-1. doi:10.1007/978-3-319-54627-8_4.

[17] X. Zhang, K. Li, C. Zhao and D. Pan, A survey on units ontologies: architecture, comparison and reuse, *Program* **51**(2) (2017), 193–213. doi:10.1108/PROG-08-2015-0056.

[18] N. Guarino and C.A. Welty, An Overview of OntoClean, in: *Handbook on Ontologies*, S. Staab and R. Studer, eds, International Handbooks on Information Systems, Springer, 2009,

pp. 201–220. ISBN 978-3-540-70999-2. doi:10.1007/978-3-540-92673-3_9.

[19] M. Poveda-Villalón, A. Gómez-Pérez and M.C. Suárez-Figueroa, OOPS! (OntOlogy Pitfall Scanner!): An On-line Tool for Ontology Evaluation, *International Journal on Semantic Web and Information Systems* **10**(2) (2014), 7–34. doi:10.4018/ijswis.2014040102.

[20] E. Sirin, B. Parsia, B.C. Grau, A. Kalyanpur and Y. Katz, Pellet: A practical OWL-DL reasoner, *J. Web Sem.* **5**(2) (2007), 51–53. doi:10.1016/j.websem.2007.03.004. https://doi.org/10.1016/j.websem.2007.03.004.

[21] A. Thompson and B.N. Taylor, Guide for the Use of the International System of Units (SI), Technical Report, National Institute of Standards and Technology, 2008.

[22] International Bureau of Weights and Measures (BIPM), *The International System of Units (SI)*, 8th edn, 2014.

[23] S. Schindler and J.M. Keil, Unit Ontology Review v1.5.0, Zenodo, 2017. doi:10.5281/zenodo.1171825. https://github.com/fusion-jena/unit-ontology-review.

[24] S. Schindler and J.M. Keil, Unit Ontology Review v1.5.0 Results, Zenodo, 2017. doi:10.5281/zenodo.1171840. https://github.com/fusion-jena/unit-ontology-review-results.

[25] C. Bizer, T. Heath and T. Berners-Lee, Linked Data - The Story So Far, *Int. J. Semantic Web Inf. Syst.* **5**(3) (2009), 1–22. doi:10.4018/jswis.2009081901.

[26] C. Bock, A. Fokoue, P. Haase, R. Hoekstra, I. Horrocks, A. Ruttenberg, U. Sattler and M. Smith, OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax, Recommendation, 27, W3C, 2009. https://www.w3.org/TR/2009/REC-owl2-syntax-20091027/.

[27] I. Johansson, Metrological thinking needs the notions of parametric quantities, units and dimensions, *Metrologia* **47**(3) (2010), 219–230. doi:10.1088/0026-1394/47/3/012.

Appendix A. Glossary

compound unit composition of **measurement units** using the mathematical operations multiplication, division or exponentiation [12] *(e.g., meter per second or kilogram per meter cubed)*

conversion tuple of an offset specified by the respective **quantity-value scales** and a **conversion factor** used to convert between two given units *(e.g., the offset* 459.67 *and the factor* 1.8 *for the conversion between degree Fahrenheit and Kelvin)*

dimension "expression of the dependence of a **quantity** on the **base quantities** of a **system of quantities** as a product of powers of factors corresponding to the base quantities, omitting any numerical factor" [10] *(e.g.,* $\dim diameter = \mathrm{L}$, $\dim width = \mathrm{L}$ *or* $\dim force = \mathrm{LMT}^{-2}$*)*

field of application scientific discipline, domain or area of life, where a **kind of quantity** is used *(e.g., physics or navigation)*

kind of quantity "aspect common to mutually comparable **quantities**" [10] *(e.g., diameter, width or mass)*

measurement unit "real scalar **quantity**, defined and adopted by convention, with which any other quantity of the same **kind** can be compared to express the ratio of the two quantities as a number" [10] *(e.g., meter or gram)*

prefix named factor defined in a **system of units** to obtain the **multiple of a unit** or **submultiple of a unit** [10] *(e.g., kilo or micro)*

quantity "property of a phenomenon, body, or substance, where the property has a magnitude that can be expressed as a number and a reference" [10] *(e.g., width of a particular street or mass of a particular apple)*

system of units "set of **base units** and **derived units**, together with their multiples and submultiples, defined in accordance with given rules, for a given **system of quantities**" [10] *(e.g., International System of Units (SI) or Centimetre–Gram–Second System of Units (CGS))*