Terminology representation guidelines for biomedical ontologies in the semantic web notations

Cui Tao*, Jyotishman Pathak, Harold R. Solbrig, Wei-Qi Wei, Christopher G. Chute
Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, United States

ABSTRACT

Terminologies and ontologies are increasingly prevalent in healthcare and biomedicine. However, they suffer from inconsistent renderings, distribution formats, and syntax that make applications through common terminologies services challenging. To address the problem, one could posit a shared representation syntax, associated schema, and tags. We identified a set of commonly-used elements in biomedical ontologies and terminologies based on our experience with the Common Terminology Services 2 (CTS2) Specification as well as the Lexical Grid (LexGrid) project. We propose guidelines for precisely such a shared terminology model, and recommend tags assembled from SKOS, OWL, Dublin Core, RDF Schema, and DCMI meta-terms. We divide these guidelines into lexical information (e.g. synonyms, and definitions) and semantic information (e.g. hierarchies). The latter we distinguish for use by informal terminologies vs. formal ontologies. We then evaluate the guidelines with a spectrum of widely used terminologies and ontologies to examine how the lexical guidelines are implemented, and whether our proposed guidelines would enhance interoperability.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Healthcare and biomedicine, more than most other sciences, has become dependent upon controlled terminologies and ontologies for interoperability, inference, and knowledge integration [6,5,8]. However, many terminologies and ontologies have independently emerged, some with overlapping or conflicting content. Additionally, virtually all of these terminologies and ontologies have their own representation schema, semantic tags, or term-element relationships which for example might indicate that a given text string is a synonym for a particular concept. This paper does not address the content overlap and conflict problem between and among ontologies, terminologies, and vocabularies, which is well characterized and largely understood if not solved [13]. This paper examines terminology syntax, representation, and tagging, which has not received anywhere near the same attention yet is probably equally important to content in practical usage. While many writers might assume the syntax problem largely solved with the introduction of OWL (Web Ontology Language) [1] or certainly now with OWL2 [25], we maintain that while elegant for description logic assertions, flavors of OWL are incomplete for practical distinction of a definition from a usage note. In addition, OWL does not provide modeling constructs to harmonize between ontologies: a well-defined abstraction layer is required for specifying how it is to be used [18]. We propose some guidelines and tags for the representation of terminologies, and separate these into structural or lexical information (e.g. preferred terms, synonyms, definition, and provenance) vs. semantic information (e.g. hierarchies). The latter renderings are divided into elements for informal terminologies, and elements for formal ontologies.

2. Challenge illustration

By examining ontologies hosted in the NCBO BioPortal [7], we found that representation inconsistencies continue to flourish in biomedical ontologies. As an example, one community may publish the definition of a concept as an rdfs:comment, while a second may use the tag DEF, and yet another may use definition. Add to this the fact that the citation for the definition is sometimes found embedded in XML fragments inside a resource, as a secondary data property for a reified resource, or in many other creative and incompatible solutions. Only recently has the Semantic Web community begun to converge on what might be considered a more standard set of tags.1 At the moment, however, these tags are still scattered across a variety of specifications such as Resource Description Framework (RDF) Schema (RDFS), Simple Knowledge Organization System (SKOS), SKOS eXtension for Labels (SKOS-XL), Web Ontology Language (OWL), Ontology Metadata Vocabulary (OMV), and many others [28,29,31,32,1,24]. To compound this problem,

* Corresponding author.
E-mail address: tao.cu@mayo.edu (C. Tao).

1 In this paper, we are using the tag and property interchangeably.
many of the tags are close to identical or having overlapping semantics. Let us use "description" as an example. RDFS provides rdfs:comment – "an instance of rdfs:Property that may be used to provide a human-readable description of a resource." SKOS specifies skos:note – "Notes are used to provide information relating to SKOS concepts … it could be plain text, hypertext, or an image; it could be a definition, information about the scope of a concept, editorial information, or any other type of information." The OMV provides omv:definition – "Free text description of an ontology", (a tag that must be present in any OMV compliant description.) BioPortal's prototype RDF model uses "bioportal:definition" (http://rest.bioontology.org/bioporal/virtual/rdf/1321[NEMOSpatial]/NEMO_0000024), while the Neural ElectroMagnetic Ontologies (NEMO) define their own properties, including "nemo:comment" (http://nemo.nic.uoregon.edu/ontologies/NEMO_annotation_properties.owl). All of these examples are perfectly well-formed RDF and yet, without additional information and transformations, none of the resulting content would be recognized as similar to a software program that was not specifically configured to recognize these as such.

In addition, we also want to ensure that the OWL semantic assertion and definition capacities are used in a semantically correct way. Many thesauri or classification schemes in the biomedical domain were designed to describe information through natural languages and define information in informal means. They mainly define a set of concepts, as well as associations and hierarchies among these concepts. In this case, it is impossible to represent the relations between the concepts using OWL Description Logic (DL) without making further assumptions.

There are many biomedical terminologies and ontologies that are not originally represented in semantic web notations. Just to name a few, the Open Biological and Biomedical Ontologies (OBO) foundry hosts more than 100 ontologies in the OBO format [23], Classification Mark-up Language (ClAML) [34], as another example, is an European norm (CEN/TS 14463) adopted by the WHO to share its classifications such as the International Classification of Diseases (ICD). When transforming ontologies or terminologies from other formats to OWL, we need to understand the potential issues on adding or changing the semantics from the original intents. Many biomedical terminologies or classification schemes do not provide a mechanism for a formal logic-based representation. In order to convert them to a formal logic-based representation, the converter has to re-engineer the information and interpret it to generate formal axioms or facts. Since there are no formal semantics defined in the original representation, it is improbable that all translation would converge upon a single reliable interpretation [30]. Kashyap et al. [17], for example, discussed multiple interpretations of a single relation. A relation Bacteria cause Infection can have five possible interpretations:

AllBacteriacause(each/only/some)Infection
SomeBacteriacause(all/some)Infection

Since the original source only specifies a relation Bacteria cause Infection without any further assertions, it is impossible to know which interpretation faithfully represents the original meaning. Many existing approaches chose to use owl:someValuesFrom as the value constraint to link the restriction to the class description [3,21,22,19]. With this assumption, the converted ontologies would choose one single interpretation: all [ClassA] [relation] some [ClassB]. The Bacteria cause Infection example can be therefore interpreted as: each bacterium must cause some infection or it is not an instance of bacteria, which is not necessarily true. Therefore, we can argue that there cannot be a general solution for choosing one single interpretation when converting relations to OWL. How to correctly interpret the relations and/or the restriction semantics depends on each individual case. SKOS, on the other hand, offers a pragmatic solution for representing associations between two concepts in thesauri or classification schemes [26]. SKOS concepts are first order resources (OWL individuals). We can define predicates between two SKOS concepts directly (e.g., “Bacteria cause Infection”). We can further add qualifiers such as “might” to this relation using annotation properties (see Section 5.5 for details) to specify that Bacteria might cause Infection. In addition, if the object property “cause” is defined as symmetric, then we can further infer that “Disease might be caused by bacteria”. This cannot be achieved by using OWL existential restrictions, which are not symmetric.

It is important for ontology designers to understand when to use OWL or SKOS based on their own applications. SKOS is designed primarily for human users to define or navigate lexical features of terminology resources. SKOS defines concepts as OWL individuals and therefore is able to relate two concepts by using object properties directly. Using SKOS, the problem of ambiguity can be left to human readers who can tolerate ambiguities well. OWL, on the other hand, is designed primarily for automatic machine processing and reasoning. OWL defines concepts on the class level and these classes are usually related by restrictions using object properties. OWL relies on well-defined formal semantics which allows no ambiguities. In summary, it will be helpful to have guidelines according to which users can decide to use OWL or SKOS in their own applications.

These problems and challenges motivated us to propose a set of guidelines that ontology engineers can reference when creating OWL ontologies or converting ontologies from other formats to OWL.

3. Terminology representation guidelines

3.1. Rationale

The guidelines introduced in this paper fall into two broad categories: guidelines for the representation of human readable information such as comments, designations, definitions; and guidelines for the representation of “semantic” information that can be used by automated tooling to classify and reason across ontological contents. We will refer the first category as “Lexical information” and the second as “Semantic information”. Our overall philosophy for designing the guidelines is: (1) using unified and standard tags to represent common lexical information; and (2) using OWL’s semantic assertions only when needed. Ontology designers need to avoid assuming semantic assertions that are not provided by the original sources when transferring terminologies or thesauri from other formats to OWL/RDF.

The first step in the proposal of terminology representation guidelines was to identify a set of commonly-used elements in biomedical ontologies and terminologies. These elements were identified based on our experience with the Common Terminology Services 2 (CTS2) Specification [10] as well as the Lexical Grid (LexGrid) project. CTS2 is an Object Management Group (OMG) Standard for defining the functional requirements of service interfaces that allow the representation, access, and maintenance of terminology contents either locally, or across a federation of terminology service nodes. CTS2 contains a computational model where the common services are specified, as well as an information model where common elements in biomedical terminologies or ontologies are defined. The LexGrid project built upon a set of common tools, data formats, and read/update mechanisms for storing, representing...
and querying biomedical ontologies and vocabularies. The primary
goal of LexGrid was to accommodate multiple vocabulary and ontol-
ogy distribution formats and support of multiple data stores for a
federated vocabulary and ontology access. Other than CTS2 and Lex-
Grid, we also took the OMV (Ontology Metadata Vocabulary) into
consideration when identifying the common elements. The set of
elements for which we propose guidelines essentially cover two
parts (1) metadata about the ontologies and terminologies, their
versions and provenance; and (2) the content of the ontologies
and terminologies, which includes concepts, common annotations
of the concepts, and common relationships between concepts.

After identifying the common elements, we developed canonical
mappings from these elements to a collection of “RDF-centric”
(RDF, RDFS, SKOS, OWL, etc.) tags for representing terminological
information using the appropriate World Wide Web Consortium
(W3C) standards. The standard W3C notations we evaluated and
include in our proposed guidelines are: the Resource Description
Framework (RDF) [28], RDF Schema [29], the Web Ontology Lan-
guage (OWL) [1], OWL 2 [25], and Simple Knowledge Organization
System (SKOS) [31]; all are W3C recommendations. Additionally,
SKOS eXtension for Labels (SKOS-XL) [32] is a W3C candidate rec-

gommendation, where labels are defined as resources to allow
descriptions and associations to be added to these labels. In addi-
tion to the above W3C recommendations, we also include the
Dublin Core Metadata Element Set (dc) [11], which is widely used to
describe digital materials. The Dublin Core Metadata Element Set
includes fifteen properties for use in resource description maintai-
ned by the Dublin Core Metadata Initiative (DCMI). Table 1 lists
the prefixes of these resources and their URIs. Finally, we used the
ISO TC37 Data Category Registry (http://www.isocat.org/) for rep-
resenting lexical tags such as synonyms and acronyms.

3.2. Lexical-information guidelines

Lexical information plays an important role in biomedical ontol-
gies, serving to identify the intent, purpose and meaning of the
elements that constitute an ontology. Table 2 lists a set of guide-
lines for representing lexical information. Each guideline carries
an identifier for easy reference in the subsequent discussion (Col-
mum 1 in Table 2). Column 2 in Table 2 introduces their intended
purposes and the tag or tags that we propose to represent them.
And in Column 3 we give examples of how to represent informa-
tion according to the guidelines. In Table 3, we list sample lexical
representations from a few different ontologies (detailed informa-
tion about the ontologies are listed in Table 4) and if the guidelines
are illustrated in these representations. In the rest of the section,
we use these examples to explain how we propose to represent re-
levant information according to the guidelines.

Guidelines L1 to L5 in Table 2 are proposed for describing meta-
level information of an ontology or an ontology entity. Rows 1–3
in Table 3 show the exemplary evaluation for some sample represen-
tations in different OWL ontologies based on these guidelines. Row
1 in Table 3, for example, uses owl:versionInfo for version, copy-
right, and creators. It is difficult for computer programs and even
for human readers to parse the information and interpret what it
means. This representation does not demonstrate guidelines L1–
L3. Our guidelines propose to represent the information as:

<owl:versionInfo>1.2</owl:versionInfo>
<dc:rights>The University of Manchester</dc:rights>
<dc:creator>Nick Drummond</dc:creator>
<dc:creator>Georgina Moulton</dc:creator>
<dc:creator>Robert Stevens</dc:creator>
<dc:creator>Phil Lord</dc:creator>

The new proposed representation listed version, copyright, and
creators separately using canonical tags recommended by the
guidelines. This way the information can be easily queried and/or
searched by computer programs without tedious configuration
by source.

The examples in Rows 2 and 3 accord with the guidelines L4 and
L5.

Many biomedical terms can be represented in various ways.
SKOS defined a tag skos:prefLabel for representing the preferred la-
bel of each term or concept. Guideline L6 recommends using
skos:prefLabel whenever a preferred label needs to be presented.
Note that SKOS allows only one value of skos:prefLabel per lan-
guage, per context. For example, each concept in an ontology for
clinical terms could have a preferred label for clinicians, and an-
other preferred label for consumers. Rows 4–6 in Table 3 show
examples of preferred labels. Both BIRNLex and NPO use a self-de-
defined OWL property for representing preferred labels. BFO, on the
other hand, uses skos:prefLabel as we recommended. Therefore,
Guideline L6 is illustrated in BFO, but not in BIRNLex nor NPO.

In addition to preferred labels, we may also need to represent
alternative labels for an ontology entity. Guideline L7 specifies
how to use skos:altLabel for alternative labels. Synonyms, acro-
nyms, abbreviations, etc., should also be considered as alternative
labels. We recommend using skos:altLabel to represent this kind of
information in addition to the original properties. W3C has not yet
specified notations for annotation properties such as synonyms,
acronyms, or abbreviations. These labels, however, quite com-
monly exist in the biomedical domain. Therefore, a standard way
to represent these labels are highly desirable. We propose using the
ISO TC37 Data Category Registry (http://www.isocat.org/) to rep-
resent these labels, which could potentially be declared as sub-
properties of skos:altLabel.

For example, here is a way to represent the information in Row
7 in Table 3:

<ISOcat:abbreviation>cislt</ISOcat:abbreviation>
<skos:altLabel>cislt</skos:altLabel>

Note that lexical labels such as acronyms or abbreviation corre-

cpond to a name (or a label) of a concept, not to the concept itself.
For example, a concept might have a preferred label “Food and
Agriculture Organization”, and an alternative label “FAO”. “FAO”
is actually the abbreviation of “Food and Agriculture Organization”,
but not of the concept itself. Therefore, theoretically the abbrevia-
tion should not annotate the concept directly. In SKOS, both the
preferred labels and the alternative labels need to be attached to the
concept. SKOS-XL, on the other hand, provides an approach

<table>
<thead>
<tr>
<th>Prefix</th>
<th>URI</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDF</td>
<td><a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#">http://www.w3.org/1999/02/22-rdf-syntax-ns#</a></td>
</tr>
<tr>
<td>RDFS</td>
<td><a href="http://www.w3.org/2000/01/rdf-schema#">http://www.w3.org/2000/01/rdf-schema#</a></td>
</tr>
<tr>
<td>OWL</td>
<td><a href="http://www.w3.org/2002/07/owl#">http://www.w3.org/2002/07/owl#</a></td>
</tr>
<tr>
<td>SKOS</td>
<td><a href="http://www.w3.org/2004/02/skos/core#">http://www.w3.org/2004/02/skos/core#</a></td>
</tr>
<tr>
<td>SKOS-XL</td>
<td><a href="http://www.w3.org/2008/05/skos-xl#">http://www.w3.org/2008/05/skos-xl#</a></td>
</tr>
<tr>
<td>dc</td>
<td><a href="http://purl.org/dc/elements/1.1/">http://purl.org/dc/elements/1.1/</a></td>
</tr>
</tbody>
</table>
where these representations can be linked to each other. In Sections 5.3 and 5.4, we explore alternative ways to represent such information.

Guideline L8 specifies how to define languages used in an ontology. When specifying the language used on the ontology meta-level, we recommend using dc:language. For example, BFO uses dc:language with the value as the language tag "en" for English as shown below, which fulfils our guideline.

```xml
<owl:Ontology rdf:about=""/>
...
<dc:language>en</dc:language>
...
</owl:Ontology>
```

Guideline L8 also specifies how to use the language tag [27] to identify languages used in a specific annotation property. For example, one can specify two preferred labels for an OWL class, one for English, one for German, like this:

```xml
<skos:prefLabel>
  > Organismstransmittingpathogens@en
</skos:prefLabel>

<skos:prefLabel>
  > ParasitenuebertragendeOrganismen@de
</skos:prefLabel>
```

We also recommend using skos:note and its sub-properties to describe plain definitions and comments as guidelines L9 and L10 specify. SKOS provides a set of tags to define specific types of notes, comments, definitions, and examples. We decided to choose skos:note for representing the definition information of a concept. Therefore, Guideline L9 is not demonstrated in either examples. BioPax (Row 11 in Table 3) uses rdfs:comment for both definition and examples. This introduces ambiguity for both human readers and computer systems. Instead, we recommend using skos:definition for definitions, and skos:example for examples.

### 3.3. Semantic-information guidelines

In this section, we discuss the guidelines for representing semantic information, especially when transforming terminologies and ontologies from other formats to OWL. Table 5 shows the detailed descriptions of these guidelines. We provide two options to fulfil different needs for representing the information in the biomedical domain.

The SKOS route (Guidelines S1a, S2a, and S3) targets thesauri or classification schemes, which “do not assert any axioms or facts, but rather identify and describe information through natural language and define information in informal means” [31]. They mainly define a set of concepts as well as associations and hierarchies among these concepts. We propose using skos:Concept to represent the “concepts” in this kind of thesauri or classification schemes as Guideline S1a specifies. SKOS defines a skos:Concept as “the unit of thought, ideas, meanings, or (categories of) objects and events” without giving any further semantic assertions. This could apply to the named entities in many knowledge organization systems: terminologies, thesauri or classification schemes. For any associations or relations, we propose defining them as an instance of owl:ObjectProperty and a sub-property of skos:semanticRelation as Guideline S2a specifies. SKOS defines relations such as broader, narrower, or related in the same way. Since both the domain and the range of skos:semanticRelation are skos:Concept, any of its sub-property inherits the same domain and range. Therefore, we can use these properties to describe the relations between two skos:Concepts. Here we show an example using a sample term from the adult mouse anatomy OBO ontology.\(^3\)

---

\(^3\) http://www.obofoundry.org/cgi-bin/detail.cgi?id=adult_mouse_anatomy.
We evaluated the guidelines from three aspects: (1) how well the guidelines cover the information in existing ontologies; (2) how well the proposed guidelines can be assessed consistently among different human experts; and (3) how well the existing ontologies already represent corresponding information in accordance to the guidelines. We focus on lexical guidelines only in this evaluation since we believe that the interpretations of the semantic representations require domain knowledge of each ontology, as well as the correct understanding of the ontology designer’s original intentions, which are beyond the scope of this paper.

### 4.1. Ontology selection

We evaluated the guidelines using a set of commonly used ontologies. These ontologies were chosen based on the most-
viewed-ontology list provided by the NCBO BioPortal [7]. The top 15 ontologies in OWL were downloaded from BioPortal. Table 7 shows the list of OWL ontologies we included in the study with their acronyms, names, URIs, versions, and the BioPortal “most viewed” ranks.5

4.2. Evaluation on the guideline coverage

To evaluate the coverage of the guidelines, we studied the annotation properties and data properties defined in the selected ontologies and classified them to four disjoint categories: domain-specific, covered by OWL specifications, covered by the guidelines, and not covered by the guidelines.

Domain specific properties refer to the properties that are specific for the ontology domain. For example, BioPax defined properties such as chemicalFormula, sequence, and molecularWeight, which are specific for describing molecular biology data. Because our guidelines are designed for features that exist commonly in biomedical ontologies, we do not plan to include the definition of domain-specific properties in the guidelines. Properties covered by OWL specifications include those properties that can be represented by OWL synopsis or schema directly. For example, BirnLex has a property called class_or_indiv. Whether an OWL entity belongs to a class or an individual can be defined using rdfs:subClassOf directly without using additional self-defined properties. For those properties that are not domain specific nor covered by OWL specifications, we determined if they can be covered by the guidelines. Table 8 shows the numbers of properties that have been classified to each category.6 As we can see there are three properties that are not covered by the lexical guidelines. The properties external_id_uris and external_ids from BirnLex describe the mappings between a local concept to an external concept. Although we do not cover these in the lexical guidelines, we believe that these mappings can be covered by the semantic guidelines for the OWL route or skos:related for the SKOS route. Note that owl:equivalentClass indicates exact mappings between resources and should be used when the two classes are semantically equivalent to each other. In many cases, the mappings between concepts are not exact [20]. In that

5 Please note that the BioPortal also includes ontologies in formats other than OWL, therefore the ranks of the top 15 ontologies in OWL are not necessary the top 15 overall.

6 RadLex was not included in this evaluation because it is in Protégé XML format, which does not define annotation or data properties.

Table 4
Detailed information for ontologies in Table 3: Column 1 shows the acronym of each ontology; Column 2 shows the full names; Column 3 shows their URIs and Column 4 shows the versions of the ontologies we used to evaluate.

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Name</th>
<th>Source URI</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>Amino acid ontology</td>
<td><a href="http://www.co-ode.org/ontologies/amino-acid/2006/05/18/amino-acid.owl">http://www.co-ode.org/ontologies/amino-acid/2006/05/18/amino-acid.owl</a></td>
<td>2.0</td>
</tr>
<tr>
<td>BFO</td>
<td>Basic formal ontology</td>
<td><a href="http://www.ifomis.org/bfo/1.1">http://www.ifomis.org/bfo/1.1</a></td>
<td>1.1</td>
</tr>
<tr>
<td>BirNLex</td>
<td>BiNLex (Biomedical Informatics Research Network controlled terminology)</td>
<td><a href="http://bioontology.org/projects/ontologies/birlnex">http://bioontology.org/projects/ontologies/birlnex</a></td>
<td>1.3.1</td>
</tr>
<tr>
<td>BRO</td>
<td>Biomedical resource ontology</td>
<td><a href="http://bioontology.org/ontologies/BiomedicalResourceOntology.owl">http://bioontology.org/ontologies/BiomedicalResourceOntology.owl</a></td>
<td>2.7.1</td>
</tr>
<tr>
<td>BioPax</td>
<td>Biological pathways exchange level 3</td>
<td><a href="http://www.biopax.org/release/biopax-level3.owl">http://www.biopax.org/release/biopax-level3.owl</a></td>
<td>0.94</td>
</tr>
<tr>
<td>NPO</td>
<td>Nanoparticle ontology</td>
<td><a href="http://purl.bioontology.org/ontology/npo">http://purl.bioontology.org/ontology/npo</a></td>
<td>1.0</td>
</tr>
<tr>
<td>OBI</td>
<td>The ontology for biomedical investigations</td>
<td><a href="http://purl.obolibrary.org/obo/obi.owl">http://purl.obolibrary.org/obo/obi.owl</a></td>
<td>1.0</td>
</tr>
<tr>
<td>SAO</td>
<td>Subcellular anatomy ontology</td>
<td><a href="http://ccdb.ucsd.edu(SAO)/1.2">http://ccdb.ucsd.edu(SAO)/1.2</a></td>
<td>1.2</td>
</tr>
</tbody>
</table>

Table 5
Guidelines for semantic information: Column 1 shows the guideline numbers; Column 2 shows the guideline descriptions; and Column 3 gives some examples of how to represent information according to the guidelines.

<table>
<thead>
<tr>
<th>No.</th>
<th>Guideline</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>SKOS route</td>
<td>S1a Use skos:Concept for concepts that are not from OWL</td>
<td>Use skos:Concept to represent OBO terms</td>
</tr>
<tr>
<td></td>
<td>S2a Defining any relationship or association between two concepts as an instance of owl:ObjectProperty and a sub-property of skos:semanticRelation</td>
<td>Representing OBO relationships</td>
</tr>
<tr>
<td></td>
<td>S3a Use rdfs:subClassOf for parent–child hierarchical relationships</td>
<td>Use rdfs:subClassOf to represent relationships such as OBO_is_a and UMLS hasSubType</td>
</tr>
<tr>
<td></td>
<td>S3a Use skos:broaden and skos:narrower to assert any kind of direct hierarchical link between two SKOS concepts</td>
<td>Defining OBO part_of as a sub-property of skos:broaden</td>
</tr>
<tr>
<td></td>
<td>S4a Use skos:related to assert an associative link between two SKOS concepts</td>
<td>Use skos:related to represent Mesh see also</td>
</tr>
<tr>
<td>OWL route</td>
<td>S1b Use owl:Class for all the concepts</td>
<td>Use owl:Class to represent OBO terms</td>
</tr>
<tr>
<td></td>
<td>S2b Use owl:ObjectProperty and assertions for relationships between two concepts</td>
<td>Representing OBO relationships</td>
</tr>
<tr>
<td></td>
<td>S3b Use rdfs:subClassOf for parent–child hierarchical relationships</td>
<td>Use rdfs:subClassOf to represent relationships such as OBO_is_a and UMLS hasSubType</td>
</tr>
<tr>
<td></td>
<td>S4b Use owl:equivalentClass for stating the equivalence of two named classes</td>
<td>Use owl:equivalentClass to represent relationships such as OBO_is_a and UMLS hasSubType</td>
</tr>
<tr>
<td></td>
<td>S5b Use owl:disjointWith to assert that the class extensions of the two class descriptions involved have no individuals in common</td>
<td>Use owl:disjointWith to represent relationships such as OBO_disjoint_from</td>
</tr>
<tr>
<td></td>
<td>S6b Use owl:intersectionOf to describe a class for which the class extension contains precisely those individuals that are members of the class extension of all class descriptions in the list.</td>
<td>Use owl:intersectionOf to represent relationships such as OBO intersection_of</td>
</tr>
<tr>
<td></td>
<td>S7b Use owl:unionOf to describe an anonymous class for which the class extension contains those individuals that occur in at least one of the class extensions of the class descriptions in the list.</td>
<td>Use owl:unionOf to represent relationships such as OBO union_of</td>
</tr>
<tr>
<td></td>
<td>S8b Use owl:complementOf to describe two classes that are complement to each other</td>
<td>e.g., owl:AnnotationProperty cannot be inherited by subclasses</td>
</tr>
<tr>
<td></td>
<td>S9b Use owl:ObjectProperty, owl:DataTypeProperty, and owl:AnnotationProperty in a semantically correct way</td>
<td></td>
</tr>
</tbody>
</table>

Table 6
To evaluate the coverage of the guidelines, we studied the lexical guidelines and classified them to three disjoint categories: domain specific, covered by OWL specifications, covered by the guidelines, and not covered by the guidelines.

<table>
<thead>
<tr>
<th>No.</th>
<th>Guideline</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>Amino acid ontology</td>
<td><a href="http://www.co-ode.org/ontologies/amino-acid/2006/05/18/amino-acid.owl">http://www.co-ode.org/ontologies/amino-acid/2006/05/18/amino-acid.owl</a></td>
</tr>
<tr>
<td>BFO</td>
<td>Basic formal ontology</td>
<td><a href="http://www.ifomis.org/bfo/1.1">http://www.ifomis.org/bfo/1.1</a></td>
</tr>
<tr>
<td>BirNLex</td>
<td>BiNLex (Biomedical Informatics Research Network controlled terminology)</td>
<td><a href="http://bioontology.org/projects/ontologies/birlnex">http://bioontology.org/projects/ontologies/birlnex</a></td>
</tr>
<tr>
<td>BRO</td>
<td>Biomedical resource ontology</td>
<td><a href="http://bioontology.org/ontologies/BiomedicalResourceOntology.owl">http://bioontology.org/ontologies/BiomedicalResourceOntology.owl</a></td>
</tr>
<tr>
<td>BioPax</td>
<td>Biological pathways exchange level 3</td>
<td><a href="http://www.biopax.org/release/biopax-level3.owl">http://www.biopax.org/release/biopax-level3.owl</a></td>
</tr>
<tr>
<td>NPO</td>
<td>Nanoparticle ontology</td>
<td><a href="http://purl.bioontology.org/ontology/npo">http://purl.bioontology.org/ontology/npo</a></td>
</tr>
<tr>
<td>OBI</td>
<td>The ontology for biomedical investigations</td>
<td><a href="http://purl.obolibrary.org/obo/obi.owl">http://purl.obolibrary.org/obo/obi.owl</a></td>
</tr>
<tr>
<td>SAO</td>
<td>Subcellular anatomy ontology</td>
<td><a href="http://ccdb.ucsd.edu(SAO)/1.2">http://ccdb.ucsd.edu(SAO)/1.2</a></td>
</tr>
</tbody>
</table>
In the SKOS route, SKOS provides a list of constructs (skos:closeMatch, skos:exactMatch, skos:broadMatch, skos:narrowMatch, and skos:relatedMatch) as sub-properties of skos:related that users can further choose to specify mappings in different situations. The guideline recommends skos:related as the default to define the mappings because the quality of the mappings is often not available. More specific properties should be used when detailed information of the mappings is available. In addition, NCIt defined a property called ALT_DEFINITION, which is similar to skos:altLabel except it is used for defining alternative definitions. In Section 5.2, we proposed a new tag to the SKOS community for representing alternative definitions. The complete list of the properties and their categories can be found in Appendix A.

We also listed the corresponding guideline(s) for each property that is classified as “covered by the guidelines” in Appendix A. Most of the conversions between these properties and their corresponding guidelines are straightforward – i.e. require one to one mappings. Sometimes a property can be covered by multiple guidelines. EFO defined a property called source_definition without giving any further description of how to use this property. From the usage of the property, we saw that EFO either uses the property to describe a textual definition of the entity or uses it to record an external URL that links to more detailed information about the entity. Whether one should use guideline L5 (source) or guideline L9 (definition) depends on each individual usage of the property. EFO also defined two properties definition_citation and definition_editor. The property definition_citation is defined as “a document, ontology class, person or organization from which the definition of the class is derived.” This property can be covered by using the combination of guidelines L9 (definition) and L5 (source). Using OWL 2, we can add annotations to another annotation. Therefore, we can add a source annotation to any definitions to represent a definition_citation. Similarly, we use the combination of guidelines L9 (definition) and L2 (author) to cover the property definition_editor.

4.3. Evaluation on inter-rater reliability

To evaluate the inter-rater reliability for accessing the guidelines, three experts (Tao, Pathak, and Wei) studied the ontologies using the guidelines. Each ontology was first studied by three human experts independently, based on the 10 guidelines for lexical representations. For each ontology with each guideline, each human expert needs to determine whether the ontology covers the information referred by the guideline. If yes, they then need to determine if the ontology represents the contents following the representation guideline or if not, why would following the guideline help semantic interoperability among different ontologies.

The examination results were then compared to identify any conflicts and disagreements among the three experts. Kappa coefficients [12] were measured for the inter-rater agreements of the results. There were 13 conflicts and disagreements in the 450 ratings and the kappa coefficient is 90%. The disagreements fall into the following categories:

- Including imported ontologies. During the examination, one expert considered imported ontologies as part of the main ontology itself, while the other two only considered the main ontology.
- Properties defined but not used. Some ontologies defined properties that were never used to annotate lexical information to the ontology or a concept in the ontology. One expert took all the defined properties into consideration. The other two evaluators did not take the unused properties into consideration. For example, DermLex defined “Definition”, “Synonym_Name”, and “Source” as annotation properties, but these properties were never applied.
- Different interpretations. There were also some debates due to different interpretations of the properties among evaluators. For example, there is a property called “curator” in OCRe2; two experts consider it as contributor or creator, but one did not. Another example is from FMA. The version information was actually embedded in the name space itself. Two experts took it into consideration, but one did not.

The three experts then had study sessions together to resolve the disagreements. We decided to take all the defined properties into consideration, and not to include the imported ontologies. For the disagreements due to different interpretations, we took the result with the most votes. Table 9 shows our findings for the selected OWL ontologies with guidelines for lexical information. Each row in Table 9 presents the result of one ontology with
Table 6
Exemplary illustration of the OWL semantic representations rendered by Protégé 4 for the Example OBO term in Fig. 1. Please note that the lexical information was not included in this table.

<table>
<thead>
<tr>
<th>Example in original format</th>
<th>Source</th>
<th>Illustration (Semantic guidelines)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 &lt;owl:Class rdf:about=&quot;#GO_0010642&quot;/&gt; The GO term in Fig. 1 S1b illustrated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 &lt;rdfs:subClassOf rdf:resource=&quot;#GO_0009968&quot;/&gt; Lines 6–7 S3b illustrated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 <a href="">owl:equivalentClass</a> Lines 8–9 S2b, S6b illustrated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 <a href="">rdfs:subClassOf</a> Line 10 S2a illustrated</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7
Detailed information for the OWL ontologies studied: Column 1 shows the acronyms; Column 2 shows their full names; Column 3 shows their URIs; Column 4 indicates the particular version of the ontology we used; and Column 5 provides ranks in BioPortal most-viewed-ontology list.

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Name</th>
<th>URI</th>
<th>Version</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>Amino acid ontology</td>
<td><a href="http://www.co-ode.org/ontologies/amino-acid/2006/05/18/amino-acid.owl">http://www.co-ode.org/ontologies/amino-acid/2006/05/18/amino-acid.owl</a></td>
<td>1.3</td>
<td>39</td>
</tr>
<tr>
<td>BFO</td>
<td>Basic formal ontology</td>
<td><a href="http://www.ifomis.org/bfo/1.1">http://www.ifomis.org/bfo/1.1</a></td>
<td>1.1</td>
<td>86</td>
</tr>
<tr>
<td>BioPax</td>
<td>Biological pathways exchange level 3</td>
<td><a href="http://www.biopax.org/release/biapax-level3.owl">http://www.biopax.org/release/biapax-level3.owl</a></td>
<td>0.94</td>
<td>13</td>
</tr>
<tr>
<td>BIRNLex</td>
<td>BIRNLex (Biomedical Informatics Research Network controlled terminology)</td>
<td><a href="http://bioontology.org/projects/ontologies/birnlex">http://bioontology.org/projects/ontologies/birnlex</a></td>
<td>1.3.1</td>
<td>26</td>
</tr>
<tr>
<td>BRO</td>
<td>Biomedical resource ontology</td>
<td><a href="http://bioontology.org/ontologies/BiomedicalResourceOntology.owl">http://bioontology.org/ontologies/BiomedicalResourceOntology.owl</a></td>
<td>2.7.1</td>
<td>26</td>
</tr>
<tr>
<td>DermLex</td>
<td>The dermatology lexicon</td>
<td><a href="http://purl.bioontology.org/ontology/-DermLex">http://purl.bioontology.org/ontology/-DermLex</a></td>
<td>1.0</td>
<td>38</td>
</tr>
<tr>
<td>EFO</td>
<td>The experimental factor ontology</td>
<td><a href="http://www.ebi.ac.uk/efo/">http://www.ebi.ac.uk/efo/</a></td>
<td>2.4</td>
<td>18</td>
</tr>
<tr>
<td>FMA</td>
<td>Foundational model of anatomy</td>
<td><a href="http://purl.bioontology.org/ontology/FMA">http://purl.bioontology.org/ontology/FMA</a></td>
<td>3.0</td>
<td>2</td>
</tr>
<tr>
<td>GALEN</td>
<td>The GALEN ontology</td>
<td><a href="http://www.co-ode.org/galen/full-galen.owl">http://www.co-ode.org/galen/full-galen.owl</a></td>
<td>1.1</td>
<td>29</td>
</tr>
<tr>
<td>MGED</td>
<td>The microarray gene expression data ontology</td>
<td><a href="http://mged.sourceforge.net/-ontologies/MGEDOntology.owl">http://mged.sourceforge.net/-ontologies/MGEDOntology.owl</a></td>
<td>1.3.1.1</td>
<td>40</td>
</tr>
<tr>
<td>NCIt</td>
<td>NCI thesaurus</td>
<td><a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl</a></td>
<td>10.03</td>
<td>1</td>
</tr>
<tr>
<td>NIFSTD</td>
<td>The neuroscience informatics framework lexicon</td>
<td><a href="http://ontology.neuinfo.org/-NIF/nif.owl">http://ontology.neuinfo.org/-NIF/nif.owl</a></td>
<td>1.8</td>
<td>9</td>
</tr>
<tr>
<td>OBI</td>
<td>Ontology for biomedical investigations</td>
<td><a href="http://purl.obolibrary.org/-obo/obi.owl">http://purl.obolibrary.org/-obo/obi.owl</a></td>
<td>1.0</td>
<td>5</td>
</tr>
<tr>
<td>OCRE</td>
<td>Ontology of clinical research</td>
<td><a href="http://purl.org/0cre/Ocre-Start-Here">http://purl.org/0cre/Ocre-Start-Here</a></td>
<td>0.95</td>
<td>27</td>
</tr>
<tr>
<td>RadLex</td>
<td>A lexicon for uniform indexing and retrieval of radiology information resources</td>
<td><a href="http://bioontology.org/projects/-ontologies/radlex/radlexOwl">http://bioontology.org/projects/-ontologies/radlex/radlexOwl</a></td>
<td>3.0</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 8
Examination of the guideline coverage with selected ontologies.

<table>
<thead>
<tr>
<th>Ontology</th>
<th>Domain specific</th>
<th>Covered by OWL specification</th>
<th>Covered by the guidelines</th>
<th>Not covered by the guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>BFO</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BioPax</td>
<td>29</td>
<td>5</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>BIRNLex</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>2 (external_id_urls, external_ids)</td>
</tr>
<tr>
<td>BRO</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>DermLex</td>
<td>6</td>
<td>0</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>EFO</td>
<td>7</td>
<td>0</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>FMA</td>
<td>5</td>
<td>0</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>GALEN</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MGED</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>NCIt</td>
<td>61</td>
<td>0</td>
<td>13</td>
<td>1 (ALT_DEFINITION)</td>
</tr>
<tr>
<td>NIFSTD</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OBI</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OCRE</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>
the first cell of each row indicating the ontology acronym. The rest of the columns show results for each ontology on the lexical guidelines. L1–L10 indicates which particular guideline. A “Y” indicates that the ontology contains the information referred by the guideline and uses the solution proposed by the guideline to represent it; an “NA” indicates that the ontology does not contain the information referred by the guideline; and an “N” indicates that the ontology does not use the representation method proposed by the guideline for the corresponding contents, where an “N1” indicates that the ontology uses an W3C recommendation tag to represent the content, but the tag was used either in an ambiguous or improper way; an “N2” indicates that the ontology uses a self-defined property; and an “N3” indicates that the ontology uses a property defined by another ontology. Note that some cells could be marked as “N” for multiple reasons. For example, an ontology defined property; and an “N3” indicates that the ontology uses a self-defined property.

4.4. Ontology evaluation result

As we can see from Table 9, all the ontologies contain some information covered by the guidelines as all of them have at least one “Y” or “N” in the table. Most of the studied ontologies (11 out of 15) cover contents referred by at least 5 guidelines. Therefore, we can see that the contents referred by our proposed guidelines commonly exist in biomedical ontologies. The ontology representation of these contents can be classified into four categories:

Represented following the guideline. In this case, semantic interoperability among different ontologies can be ensured.

Represented using a tag in a W3C recommendation (either in an ambiguous way or improperly). Some ontologies use a single tag for many different purposes. For example, we found that rdfs:comment has been used in many different situations: for representing an example, a definition, an editorial note, etc. In addition, some tags were not used as designed. For example, information such as author, version, and copyright was all represented using owl:versionInfo. These ambiguous representations make it difficult for automatic terminology services to locate the proper information during querying, updating, and integrating ontology elements.

Represented using a self-defined property. Some ontologies defined their own properties for representing contents such as definition, and preferred label. Since these contents are very common, we believe that using a unified way to represent them could ensure better interoperability among ontologies than using arbitrarily defined properties.

Represented using a property defined in another ontology. Similar to above, using a tag from an ontology that is not a W3C recommendation can hardly ensure semantic interoperability since different ontologies could choose to import and use properties arbitrarily.

5. Future direction: additional meta-level information for interoperability

During our evaluation process, we found some meta-level information shared by the biomedical ontologies that is important for terminology interoperability but could not yet be reasonably represented by using W3C recommendations. In this situation, new tags need to be introduced and proposed to W3C. Here we list a few relevant new tags we proposed. We have discussed detailed information about these tags in a semantic web conference [33].

**Table 9**

Examination of a set of commonly-used ontologies for the representation of contents referred by the lexical information guidelines: Column 1 indicates the ontologies by their acronyms; and the rest of columns show the results for each ontology on the lexical guidelines. L1–L10 indicate the particular guideline. A “Y” indicates that the ontology uses the representation method proposed by the guideline; an “NA” indicates that the ontology does not contain the information referred by the guideline; and an “N” indicates that the ontology does not use the representation method proposed by the guideline, where an “N1” indicates that the ontology uses an W3C recommendation tag to represent the content, but the tag was used either in an ambiguous or improper way; an “N2” indicates that the ontology uses a self-defined property; and an “N3” indicates that the ontology uses a property defined by another ontology.

<table>
<thead>
<tr>
<th>Ontology</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>L4</th>
<th>L5</th>
<th>L6</th>
<th>L7</th>
<th>L8</th>
<th>L9</th>
<th>L10</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>N1</td>
<td>N1</td>
<td>N1</td>
<td>N1</td>
<td>NA</td>
<td>N2</td>
<td>N2</td>
<td>Y</td>
<td>Y</td>
<td>N1</td>
</tr>
<tr>
<td>BFO</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N1</td>
<td>Y</td>
<td>N1</td>
<td>N1</td>
<td>N1</td>
</tr>
<tr>
<td>BioPax</td>
<td>N1</td>
<td>N1</td>
<td>N1</td>
<td>N1</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
<td>NA</td>
<td>N1</td>
<td>N1</td>
</tr>
<tr>
<td>BRNLEX</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>N1</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
</tr>
<tr>
<td>BRO</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>N2</td>
<td>Y</td>
<td>N2</td>
<td>N2</td>
<td>N3</td>
</tr>
<tr>
<td>DermLex</td>
<td>N2</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
</tr>
<tr>
<td>EFO</td>
<td>Y</td>
<td>N2</td>
<td>NA</td>
<td>NA</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
</tr>
<tr>
<td>FMA</td>
<td>N1</td>
<td>N2</td>
<td>NA</td>
<td>NA</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
<td>N1</td>
</tr>
<tr>
<td>GALEN</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>MGED</td>
<td>Y</td>
<td>Y</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
</tr>
<tr>
<td>NCIT</td>
<td>Y</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
</tr>
<tr>
<td>NIFSTD</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>NA</td>
<td>N3</td>
<td>Y</td>
<td>N3</td>
<td>N3</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>OBI</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N1</td>
<td>N1</td>
<td>N3</td>
<td>N3</td>
<td>N3</td>
<td>N3</td>
</tr>
<tr>
<td>OCre</td>
<td>NA</td>
<td>N2</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>N1</td>
<td>NA</td>
<td>N2</td>
<td>N1</td>
</tr>
<tr>
<td>RedLex</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
<td>N2</td>
</tr>
</tbody>
</table>

**Fig. 2.** An example of designation type.

5.1. Properties for concept mappings

There are a large number of biomedical ontologies covering overlapping contents [14]. Many research efforts have been focusing on identifying mappings between ontology resources [9,16]. Different approaches or users, however, could have different definitions on mappings. In many cases, these mappings are not exact mappings with semantic equivalence between two resources as owl:equivalentClass indicates, but are rather partial, lexical, or with uncertainties. SKOS provides a list of constructs – (skos:closeMatch, skos:exactMatch, skos:broadMatch, skos:narrowMatch, and skos:relatedMatch) – using which users can further choose to specify mappings in different situations. These properties, however, can only be used for mappings between SKOS concepts. CTS2 allows different mapping resolutions between ontology resources: different versions of mappings in the same pair of resources can be done by different algorithms or users. There is no standard tags from W3C, however, to represent this kind of information. We believe it is necessary to specify standard tags for mapping types and mapping methods for OWL ontology resources.

5.2. Preferred definition

Textual definitions are very common in biomedical ontologies. Many biomedical concepts have more than one textual definitions. SKOS has defined prefLabel and altLabel, but no such constructs are provided for “definitions”. Akin to prefLabel and altLabel, our objective is to propose prefDefinition and altDefinition to the SKOS committee to be included in the future specification.
5.3. Designation type

Typical designation types include acronym, synonym, eponym, abbreviation, etc., which commonly exist in the biomedical domain. Here we propose a new construct called designationType, using which ontology designers can declare the type of a lexical presentation. There is no standard recommendations to define designation types by W3C yet. To make it connect to standards, we propose adopting the ISO TC37 Data Category Registry (http://www.isocat.org/) to be the primary resource for designation types. Fig. 2 shows an example using designationType. Here we use SKOS-XL data model to represent that a concept \(<A1>\) has an alternative label ''FAO''. The first three lines in Fig. 2 entail the expression \(<A1>\text{skos:altLabel} \text{''FAO''}\). Using SKOS-XL, we can further specify that the its designation type is ISOcat:acronymFor. The skos_plus:designa-tionType itself is an OWL annotation property. We can define its range as the collection of a set of pre-defined OWL annotation properties that represent different designation types such as ISO-cat:acronymFor, ISOcat:abbreviationFor, and ISOcat:synonym.

5.4. Relations between lexical properties

Relations between two lexical properties (e.g., definitions, synonyms, etc.) are very common in biomedical ontologies. We propose a new property called noteRelation, which can be used to identify an association between two lexical properties. It can be viewed as a super-property of skosxl:labelRelation. In SKOS-XL, the object property skosxl:labelRelation is designed for representing binary links between instances of the class skosxl:Label. The new proposed noteRelation is designed for representing relations between not only two labels, but also any two lexical properties, such as definitions, notes, and examples. As with designationType, the types of these links could be adopted from the ISO TC37 Data Category Registry. Fig. 3 shows an example for using skosxl:labelRelation to represent the property link between two labels. Similarly, we can represent the relations between any two notes using noteRelation.

5.5. Association qualification

In many cases in the clinical domain, we need to modify a relation between two concepts or instances. For example, one can define an association, \(\text{Poland anomaly} \text{HAS CLINICAL SIGN} \text{Dextrocardia}\), where \(\text{HAS CLINICAL SIGN}\) is the association (relation) name, \(\text{Poland anomaly}\) is the association source, and Dextrocardia is the association target. This association instance also has an association qualification indicates how frequently the disease has the symptom. The association qualification has a name Frequency and a value Very frequent. Fig. 4b shows how we represent this example using an N-ary relation definition proposed by the W3C [36]. We first declare a new node \(\text{HAS CLINICAL SIGN}_{relation}\) for the N-ary relation. For associationQualification, We define a new OWL annotation property, ctm:associationQualification. Every actual association qualifier is defined as a sub-property of ctm:associationQualification, and therefore is also an instance of OWL annotation property.

6. Conclusion

In this research, we propose a set of guidelines using constructs in W3C recommendations such as RDF, OWL, and SKOS for representing common lexical and semantic information in the biomedical domain. The guidelines provide a unified semantic-web compatible model for representing biomedical ontologies and terminologies. Based on them, heterogeneous terminological and ontological information can be translated to or represented in
semantic web notations with a well-defined interoperability. The biomedical informatics community can greatly benefit by applying semantic-web’s combination of formal semantics, rich expressiveness, and shared software base to biomedical and clinical terminologies. We illustrated the benefit of using the guidelines to enhance semantic interoperability with a set of popular ontologies in the biomedical domain. In addition, we have also identified several limitations of the existing W3C specifications and proposed new tags that warrant broader community engagement.

Acknowledgments

This research is partially supported by the National Center for Biomedical Ontologies (NCBO) under the NIH Grant #N01-HG04028 and the NSF under Grant #0937060 to the Computing Research Association for the CIFellows Project. The authors thank Drs. Mark Musen, Natasha Noy, and Nigam Shah for their valuable suggestions during the preparation of the paper.

References