

July 2020

Best-Practice Patterns for Biomedical Ontologies: Moving Under the Meta-Modeling Hood

Jana VATAŠČINOVÁ¹, Viet Bach NGUYEN, Vojtěch SVÁTEK, Ondřej ZAMAZAL

*Faculty of Informatics and Statistics, University of Economics, Prague,
W. Churchill sq. 1938/4, 130 67 Prague 3 – Žižkov, Czech Republic*

Abstract. Biomedical ontologies tend to model domain entities as classes, sometimes even such that have the ontological characteristics of instances (particulars) rather than true concepts (universals). Transformation of biomedical ontologies to an alternative representation, with classes meta-modeled using instances, may support their interoperability with linked data vocabularies/datasets and facilitate some reasoning-lean operations on them. We studied a set of best-practice patterns published at the Manchester ontology pattern portal and explored how they would be expressed in terms of instances and what are the implications.

Keywords. ontology pattern, individual, transformation, meta-modeling

1. Introduction

The biomedical domain is probably the first in which description-logic (DL) reasoning had been used at large scale, leveraging on resources such as SNOMED-OWL.² Since traditional DL reasoning tasks put emphasis on the ontology Tbox, the development of biomedical ontologies is naturally prone to modeling domain entities as *classes*, sometimes even for entities that have to some degree the characteristics of ontological *particulars* (e.g., it is unlikely or impossible for them to have instances). On the other hand, *linked data* (LD) *vocabularies*, a large amount of which are cataloged within the LOV portal [5], frequently employ individuals for the meta-modeling of concepts, i.e. ontological *universals*, since it allows them to form statements without recourse to complex Tbox axioms. These individuals may be a part of the ontologies themselves, in the form of embedded code lists, or of standalone datasets (knowledge graphs, KGs) such as DBpedia.³ While instance-based modeling limits the opportunities for DL reasoning, it can make the structures more comprehensible for people beyond the ontology engineering field, and also more accessible to simple querying or visualization applications tailored to consuming instance-based KGs. Some possibilities for model checking still remain at

¹Corresponding author: Jana Vataščinová; e-mail: e-mail: xvajt00@vse.cz

²<http://snomed.org/owl>

³<https://dbpedia.org/>

July 2020

the level of shape (constraint) languages such as SHACL.⁴ Exploring the possibility of transforming class-centric biomedical ontologies to alternative instance-based structures thus appears relevant.

The idea of transforming the ontology Tbox to some kind of Abox representation is not new. Specifically for SKOS, it has been addressed by the OWL2SKOS Converter (and associated guidelines) developed by Abdul-Manaf [1], as well as in several use cases of the PatOMat system [6]. While this effort is aimed at obtaining a simplified representation that allows processing and displaying by common RDF (and SKOS) tools, a different motivation of meta-modeling of classes by instances came from consistency checking. Glimm et al. applied meta-modeling in order to perform the checking of higher-order consistency within OWL DL [3]. A similar approach was used to verify the coherency of background models (in the PURO language) underlying OWL ontologies [4].

Our analysis below is rather in the direction of the former approach. However, it does not primarily aim at taxonomies but at structures involving *non-taxonomic properties*.

Rather than examining particular ontologies randomly chosen from, e.g., those 800+ at the BioPortal,⁵ we focused on a library of *ontology design patterns*, which should, ideally, represent the gist of how biomedical ontologies are designed (or, will be designed in the future by modelers who subscribe to such ‘best-practice’ patterns). A pattern library proposed primarily for the biomedical domain is the ‘Manchester’ one, available at <http://www.gong.manchester.ac.uk/odp/html/>. It contains 17 patterns, roughly categorized to three groups, Extension ODPs, Good Practice ODPs and Domain Modelling ODPs. We examined the structure of the patterns and explored how the same structure (leveraging on the accompanying examples) would be expressed in terms of instance-based meta-modeling. While all patterns are in some way amenable to transformation to an instance-based version, in this paper we focus on a family of three patterns, Entity-Quality, Entity-Property-Quality and Entity-Feature-Values, which have been jointly described in a paper by Egaña et al. [2]. Additionally, for comparison, we also analyze the N-ary DataType Relationship pattern, as a substantially different one from the ontological (universal vs. particular) point of view.

2. Universals vs. Particulars in Biomedical Ontologies

An important question for our research is whether and how we can characterize the entities captured in biomedical ontologies as being inherently universals or particulars. This is a complex topic; in this short paper we only approximate it in pragmatic terms, omitting references to foundational ontology literature where such problems were studied.

Concepts in biomedical ontologies referring to *objects* that have (to some degree) *clear boundaries* and *macroscopic size*, such as organs or medical instruments, usually have a clear universal/particular distinction. This also holds for some *processes*, such as examinations or surgeries, which can be precisely anchored in time. As an example, the concept of heart is clearly different from a particular heart that instantiates it. A concrete surgery (e.g., a valve replacement), instantiating the general concept of such a kind of surgery, is then performed on a concrete heart. In a clinical context, both the universals and the particulars instantiating them may be worth representing in data. Thus, if we

⁴<https://www.w3.org/TR/shacl/>

⁵<https://bioportal.bioontology.org/ontologies>

decide to model the generic concept of heart, or of valve replacement, by an individual, this re-modeling should be viewed as true meta-modeling.

On the other hand, other or even the same ontologies may refer to processes that occur at *microscopic level*, where we do not routinely follow the trajectory of individual objects (such as cells or organelles). The likely purpose of ontologies is then to facilitate the description of *generic biological mechanisms*. In this situation it would be quite natural to represent the respective concepts as individuals. While their instances exist, these are not clinically relevant, and we would not process data on them in most applications (except, say, some sophisticated studies of rare events using high-definition microscopy). Furthermore, we also have biomedical concepts that refer to some kind of *matter*. In this case we do not have a clear notion of instantiation at all; whether a molecule (or some measurable amount) of a compound is an instance of this compound is a tough ontological issue. Again, it would be natural to represent the concept as an individual.

While these differences do not affect the way the classes from pattern-based structures are meta-modeled by individuals in our analysis, they will likely have an impact on the preservation of the ontological faithfulness, from the point of view of a human interpreting the resulting structure.

3. Selected Patterns and Their Instance-Based Meta-Models

Egaña et al. [2] jointly introduced (partly based on existing literature) three patterns that tackle the problem of representing ‘modifiers and values’ in ontologies and are now a part of the Manchester pattern collection: *Entity-Quality* (EQ), *Entity-Property-Quality* (EPQ) and *Entity-Feature-Values* (EFV). The example situation used to illustrate them all is that of cell components having a certain position. With respect to our considerations from Section 2, the example models a piece of *general biological knowledge*: e.g., the apical complex is (always, by its nature) in the apical position in a cell. We do not expect real biomedical datasets to track the lifecycle of individual organelles inside individual, physical, cells. Therefore, instantiation of the respective classes is unlikely to happen, and their meta-modeling by instances appears as a plausible representation option, provided we only want to represent (and then, e.g., search, interlink or visualize) a particular collection of biological knowledge and not perform DL-based inference on it.

The original diagram of the EQ pattern is in Fig. 1. We see that its structure contains three property restrictions, one equivalence axiom with a class union, and one disjointness axiom. All remaining boxes correspond to placeholders for named OWL classes. In the corresponding example diagram,⁶ *entity_n* is substituted with *apical_complex*, *entity_category* with *cell_part*, *has_quality* with *has_position*, *quality_value_n* with *apical*, and *quality* with *position* (only considering entities relevant for the biomedical knowledge statement describing the apical complex position).

The left-hand side of Fig. 2 shows the main part of the proposed meta-modeled version of EQ, namely, its structure of *Abox triples*, with individuals depicted as ovals. The ‘a’ edge corresponds to the *rdf:type* property (as in the Turtle RDF serialization), and entity naming conventions common in LD vocabularies are applied: camel case and upper-initial naming. Finally, placeholder labels are in italics, while concrete entities such as the *hasQuality* property (invariant wrt. pattern instantiation), are in normal font.

⁶See http://odps.sourceforge.net/odp/html/Entity_Quality.html.

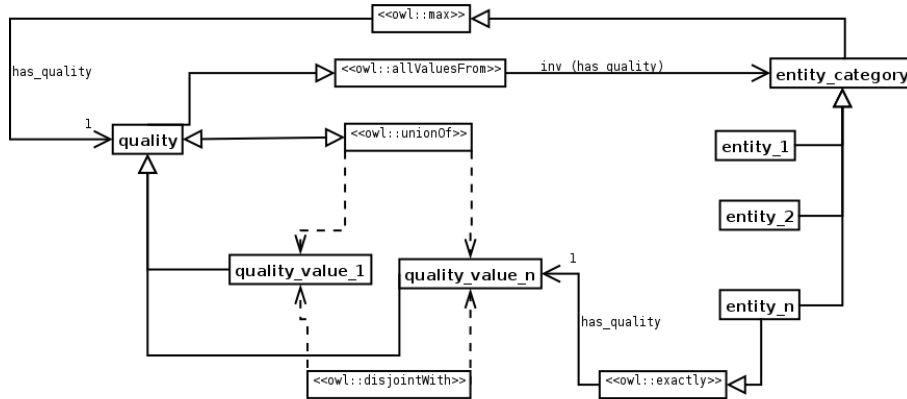


Figure 1. EQ pattern, from http://odps.sourceforge.net/odp/html/Entity_Quality.html

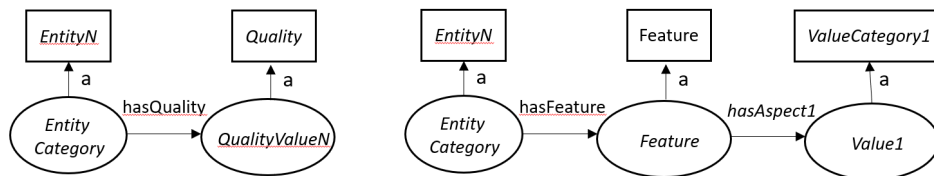


Figure 2. Meta-modeled Abox core of the EQ pattern (left) and EFV pattern (right)

The triple-based structure conveys the given knowledge statement in simple terms; yet, there is no major information loss wrt. the original Tbox version of EQ, since the important axioms can be transformed alongside and kept within a secondary Tbox layer:

- The max-cardinality and universal axioms are left intact, since both *EntityCategory* and *Quality* are preserved as classes.
- The exact-cardinality axiom becomes void: the ‘max 1’ feature is entailed by the max-cardinality defined for the *EntityCategory*, which *EntityN* instantiates, and the ‘min 1’ feature is entailed by the actual *hasQuality* property assertion.
- The equivalence axiom right-hand side is transformed from a concept union to an enumeration of individuals (though, with possible impact on the supported OWL profile); the disjointness axiom is then void, too.

EPQ and EFV⁷ only mildly differ from EQ. Instead of the max cardinality axiom they use domain and range axioms; however, the entities involved are unchanged by the meta-modeling transformation, and thus remain intact. The EPQ pattern meta-model has a *property placeholder* instead of the *hasQuality* property; the property substituted for this placeholder will however be a subproperty of *hasQuality*. The EFV pattern meta-model Abox core is in the right part of Fig. 2; it introduces the specific *Feature* entity allowing to group different aspects of the feature (such as the state and orientation of the position) together. All of this is straightforwardly meta-modeled.

⁷Their diagrams are in http://odps.sourceforge.net/odp/html/Entity_Property_Quality.html and http://odps.sourceforge.net/odp/html/Entity_Feature_Value.html, respectively. We do not include them here due to the space limitations.

July 2020

In contrast to the above patterns, we will also briefly comment on the *N-ary DataType Relationship* pattern,⁸ which describes how a combination of data values can be assigned to a single entity. In the accompanying example, this entity is *standard_water_boiling_point*, which is characterized by the pressure of 1 atmosphere and by the temperature of 100 degrees Celsius. *Standard water boiling point*, although modeled by a class in the original pattern, is however not an ontological universal but a particular. If we represent this entity by means of an individual and replace the restrictions on the (datatype) properties with mere property assertions, we yield a more *ontologically faithful* representation: it is thus *not* meta-modeling from the foundational point of view.

4. Conclusions

We made an early proof of concept in transforming ontology patterns to an alternative representation relying on individuals meta-modeling the original classes. We believe that such a study can be useful both for bridging between the biomedical ontology and LD vocabulary communities and for making some reasoning-lean operations over biomedical ontologies themselves more feasible. As a side effect, we can also obtain insights on the ontological faithfulness of pattern use cases along the universal-particular dimension.

We are currently finalizing the meta-modeling study for the whole Manchester catalog (17 patterns), and plan to extend this analysis, among other, to the lexico-syntactic patterns from the *ontologydesignpatterns.org* portal. We would also like to verify which of the Tbox axioms of the original patterns can be emulated by *SHACL constraints*. Furthermore, we also want to perform a qualitative study of the *likelihood* of biomedical ontology classes to be *populated by instances*, and, consequently, estimate which of them might deserve to be converted to (Abox) code lists easily reusable in the linked data context. Finally, we plan to focus on specific *use cases*, such as that of aligning and comparing fragments of biomedical ontologies, built using the design patterns in a class-intensive style, with pre-existing instance-level code lists or ‘wild’ knowledge graphs.

Acknowledgment: This research is supported by the IGA VSE project no. F4/33/2018.

References

- [1] Abdul Manaf NAB. Transforming Ontologies In the Web Ontology Language (OWL) to Vocabularies in the Simple Knowledge Organization System (SKOS). PhD Thesis, Univ. Manchester, 2015.
- [2] Egaña M, Rector AL, Stevens R, Antezana E. Applying Ontology Design Patterns in Bio-ontologies. Proc. EKAW 2008: 7-16.
- [3] Glimm B, Rudolph S, Völker J. Integrated Metamodeling and Diagnosis in OWL 2. International Semantic Web Conference, 2010.
- [4] Svátek V, Homola M, Kluka J, Vacura M. Metamodeling-Based Coherence Checking of OWL Vocabulary Background Models. OWLED 2013.
- [5] Vandenbussche P-Y, Atemezing G, Poveda-Villalón M, Vatan B. Linked open vocabularies (LOV): A gateway to reusable semantic vocabularies on the web. Semantic Web, (3): 437–452, 2017.
- [6] Zamazal O, Svátek V. PatOMat – Versatile Framework for Pattern-Based Ontology Transformation. Comput. Informatics 34(2): 305-336 (2015)

⁸http://odps.sourceforge.net/odp/html/Nary_DataType_Relationship.html