

# Towards Vagueness-Aware Semantic Data

Panos Alexopoulos<sup>1</sup>, Boris Villazon-Terrazas<sup>1</sup>, and Jeff Z. Pan<sup>2</sup>

<sup>1</sup> iSOCO, Intelligent Software Components S.A., Av. del Partenon, 16-18, 1-7, 28042, Madrid, Spain,

{palexopoulos,bvillazon}@isoco.com

<sup>2</sup> Department of Computing Science, University of Aberdeen, Meston Building, Aberdeen, AB24 3UE, UK.

**Abstract.** The emergence in recent years of initiatives like the Linked Open Data (LOD) has led to a significant increase in the amount of structured semantic data on the Web. In this paper we argue that the shareability and wider reuse of such data can very often be hampered by the existence of vagueness within it, as this makes the data’s meaning less explicit. Moreover, as a way to reduce this problem, we propose a vagueness metaontology that may represent in an explicit way the nature and characteristics of vague elements within semantic data.

## 1 Introduction

Ontologies are formal shareable conceptualizations of domains, describing the meaning of domain aspects in a common, machine-processable form by means of concepts and their interrelations [4], and enabling the production and sharing of data that are commonly understood among human and software agents. Achieving the latter requires ensuring that the meaning of ontology elements is explicit and shareable, namely that all users have an unambiguous and consensual understanding of what each ontological element actually represents. In this paper we examine how **vagueness** affects shareability and reusability of semantic data. Vagueness is a common natural language phenomenon, demonstrated by concepts with blurred boundaries, like *tall*, *expert* etc., for which it is difficult to determine precisely their extensions (e.g. some people are borderline tall: neither clearly “*tall*” nor “*not tall*”) [5].

Our position is threefold. i) That vagueness exists not only within isolated, application-specific, semantic data but also in public datasets that should be shareable and reusable. ii) That vagueness hampers the comprehensibility and shareability of these datasets and cause problems. iii) That the negative effects of vagueness can be partially tackled by making the data **vagueness-aware**, namely by annotating their elements with meta-information about the nature and characteristics of their vagueness. In the next section we explain and support the first two parts of our position with real world examples. In section 3 we describe how semantic data can become vagueness-aware via a vagueness metaontology. Sections 4 and 5 present related work and summarize our own.

## 2 Motivation and Approach Rationale

The possibility of vagueness in ontologies and semantic data has long been recognized in the research literature, especially in the area of Fuzzy Ontologies [3] [2]. An inspec-

tion of well-known ontologies and public semantic data reveals that the possibility is indeed a reality. A characteristic group of such elements are categorization relations where entities are assigned to categories with no clear applicability criteria. An example of such a relation is “*hasFilmGenre*”, found in Linked Data datasets like Linked-MDB (<http://linkedmdb.org>) and DBpedia (<http://dbpedia.org>), that relates films with the genres they belong to. As most genres have no clear applicability criteria there will be films for which it is difficult to decide whether or not they belong to a given genre. A similar argument can be made for the DBpedia relations “*is dbpedia-owl:ideology of*” and “*dbpedia-owl:movement*”. Another group of vague elements comprises specializations of concepts according to some vague property of them. Examples include “*Famous Person*” and “*Big Building*”, in the Cyc Ontology (<http://www.cyc.com/platform/opencyc>), and “*Managerial Role*” and “*Competitor*”, found in the Business Role Ontology (<http://www.ip-super.org>).

The presence of vague terms in semantic data often causes **disagreements** among the people who develop, maintain or use it. Such a situation arose in a real life scenario where we faced significant difficulties in defining concepts like “*Critical System Process*” or “*Strategic Market Participant*” while trying to develop an electricity market ontology. When, for example, we asked our domain experts to provide exemplary instances of critical processes, there was dispute among them about whether certain processes qualified. Not only did different domain experts have different criteria of process criticality, but neither could anyone really decide which of those criteria were sufficient for the classification. In other words, the problem was the vagueness of the predicate “*critical*”. While disagreements may be overcome by consensus, they are inevitable as more users alter, extend, or use semantic data. A worse situation is when a user misinterprets the intended meaning of a vague term and uses it wrongly. Imagine an enterprise ontology where the concept “*Strategic Client*” was initially created and populated by the company’s Financial Manager whose implicit criterion was the amount of revenue the clients generated for the company. Imagine also the new R&D Director querying the instances of this concept when crafting an R&D strategy. If their own applicability criteria for the term “*Strategic*” do not coincide with the Financial Manager’s, using the returned list of clients might lead to poor decisions. The above examples show how the inherent context-dependence and subjectivity that characterizes vagueness may affect shareability in a negative way, due to potential disagreements or misunderstandings. More generally, typical use-case scenarios where this may happen include:

1. **Structuring Data with a Vague Ontology:** When domain experts are asked to define instances of vague concepts and relations, then disagreements may occur on whether particular entities constitute instances of them.
2. **Utilizing Vague Facts in Ontology-Based Systems:** When knowledge-based systems reason with vague facts, their output might not be optimal for those users who disagree with these facts.
3. **Integrating Vague Semantic Information:** When semantic data from several sources need to be merged then the merging of particular vague elements can lead to data that will not be valid for all its users.
4. **Evaluating Vague Semantic Datasets for Reuse:** When data practitioners need to decide whether a particular dataset is suitable for their needs, the existence of vague

elements can make this decision harder. It can be quite difficult for them to assess *a priori* whether the data related to these elements are valid for their application context.

To reduce the negative effects of vagueness, we put forward the notion of **vagueness-aware semantic data**, informally defined as “*semantic data whose vague ontological elements are accompanied by comprehensive metainformation that describes the nature and characteristics of their vagueness*”. For example, a useful piece of metainformation is the set of applicability criteria that the element creator had in mind when defining the element (e.g. the amount of generated revenue as a criterion for a client to be strategic in the previous section’s example). Another is the element creator itself (e.g. the author of a vague fact). In any case, our position is that having such metainformation, explicitly represented and published along with the vague semantic data, can improve the latter’s comprehensibility and shareability, especially in regard to the four scenarios of the previous section. For example, the knowledge of the same vague concept’s intended applicability criteria in two different datasets can i) prevent their merging in case these criteria are different and ii) help a data practitioner decide which of these two concepts’s associated instances are more suitable for his/her application.

### 3 Making Ontologies Vagueness-Aware

#### 3.1 Key Vagueness Aspects

In the literature two kinds of vagueness are identified: *quantitative-* or *degree-vagueness*; and *qualitative-* or *combinatory vagueness* [5]. A predicate has degree-vagueness if the existence of borderline cases stems from the lack of precise boundaries for the predicate along one or more dimensions (e.g. “*bald*” lacks sharp boundaries along the dimension of hair quantity while “*red*” can be vague for both brightness and saturation). A predicate has combinatory vagueness if there are a variety of conditions pertaining to the predicate, but it is not possible to make any crisp identification of those combinations which are sufficient for application. A classical example of this type is “*religion*” as there are certain features that all religions share (e.g. beliefs in supernatural beings, ritual acts) yet it is not clear which are able to classify something as a religion. Based on this typology, we suggest that for a given vague term it is important to represent and share the following explicitly:

- **The type of the term’s vagueness:** Knowing whether a term has quantitative or qualitative vagueness is important as elements with an intended (but not explicitly stated) quantitative vagueness can be considered by others as having qualitative vagueness and vice versa.
- **The dimensions of the term’s quantitative vagueness:** When the term has quantitative vagueness it is important to state explicitly its intended dimensions. E.g., if a CEO does not make explicit that for a client to be classified as strategic, its R&D budget should be the only pertinent factor, it will be rare for other company members to share the same view as the vagueness of the term “*strategic*” is multi-dimensional.

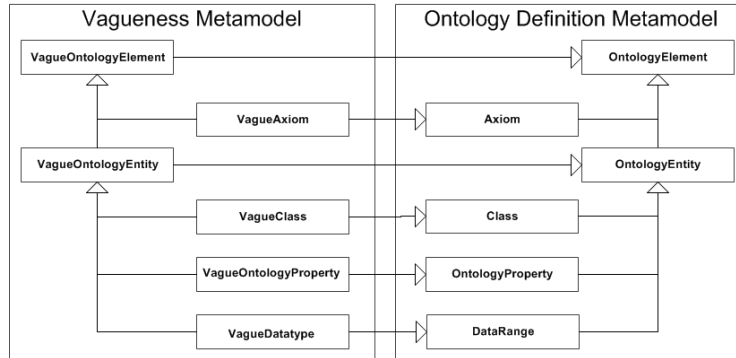
- **The necessary applicability conditions of the term’s qualitative vagueness:** Even though a term with qualitative vagueness lacks a clear definition of sufficient conditions for objects to satisfy it, it can still be useful to define the conditions that are necessary for its applicability. This will not only narrow down the possible interpretations of the term (by including conditions that other people may forget or ignore) but will also provide better grounding on any discussion or debate that might arise about its meaning.

Furthermore, vagueness is **subjective** and **context dependent**. The first has to do with the same vague term being interpreted differently by different users. Two company executives might have different criteria for the term “*strategic client*”. Even if they share an understanding of the type and dimensions of this term’s vagueness, a certain amount of R&D budget (e.g. 1 million euros) makes a client strategic for one but not the other. Similarly, context dependence has to do with the same vague term being interpreted or applied differently in different contexts even by the same user; celebrating an anniversary is different to celebrating a birthday when it comes to judging how expensive a restaurant is. Therefore we additionally suggest that one should explicitly represent the term’s **creator** as well as the **applicability context** for which it is defined or in which it is used.

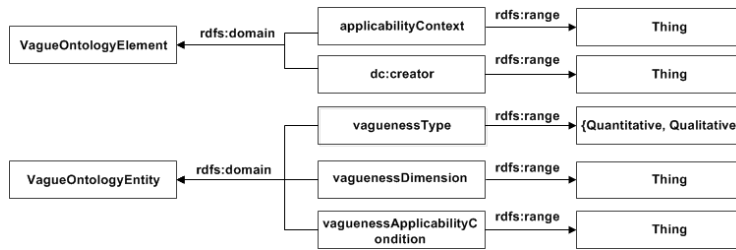
### 3.2 A Metamodel of Vague Ontology Elements

Ontology elements that can be vague are typically concepts, relations, attributes and datatypes [2]. A concept is vague if – in the given domain, context or application scenario – it admits borderline cases; namely if there could be individuals for which it is indeterminate whether they instantiate the concept. Similarly, a relation is vague if there could be pairs of individuals for which it is indeterminate whether they stand in the relation. The same applies for attributes and pairs of individuals and literal values. Finally, a vague datatype consists of a set of vague terms which may be used within the ontology as attribute values (e.g. *performance* may take as values terms like *poor*, *mediocre* and *good*). To formally represent these vague elements by means of a metaontology, we consider the OWL metamodel defined in [6] and extend it by defining each vague element as a subclass of its corresponding element and by defining appropriate metaproperties that reflect the key aspects discussed in the previous sections. Figures 1 and 2 provide an overview of the metamodel while a concrete example of how this may be used to annotate a vague ontology is available at <http://boris.villazon.terrazas.name/data/VagueOntologyExample.ttl>

The metamodel is to be used by producers and consumers of semantic data, the former utilizing it to **annotate** the vague part of their ontologies with relevant metainformation and the latter **querying** this metainformation to better use them. Vagueness annotation is a manual task, meaning that knowledge engineers and domain experts should detect the vague elements, determine the relevant characteristics (type, dimensions, etc.) and populate the metamodel. How this task may be best facilitated is a subject for further research, but a good starting point would be the integration of the process within traditional semantic data production processes. Regarding the consumption of a vagueness-aware ontology, the first benefit it has for its potential users is that



**Fig. 1.** Classes of Vagueness Metamodel



**Fig. 2.** Properties of Vague Elements

it makes them aware of the existence of vagueness in the domain. This is important because vagueness is not always obvious, meaning it can easily be overlooked and cause problems. The second benefit is that the ontology's users may query each of the vague elements' metainformation and use it in order to reduce these problems.

For example, when structuring data with a vague ontology, disagreements may occur on whether particular objects are instances of vague concepts. If, however, information like the applicability conditions and contexts of these elements are known to the people who perform this task, then their possible interpretation spaces will be reduced. Also, when vague elements are used within some end-user application, the availability of vagueness metainformation can help the system's developers in two ways. i) It will make them aware of the fact that the ontology contains vague information and thus some of the system's output might not be considered accurate by the end-users. ii) They may use the vagueness metainformation to try to deal with that. For example, the applicability context of a vague axiom can be used in a recommendation system to explain why a particular item was recommended. Finally, in dataset integration and evaluation scenarios, the vagueness metamodel can be used to compare ontologies' vagueness compatibility. For example, if the same two vague classes have different vagueness dimensions, then the one class's set of instance membership axioms might not be appropriate for the second's as it may have been defined with a different vagueness interpretation in mind. A simple query to the two ontologies' vagueness metamodel could reveal this issue.

## 4 Related Work

Representing semantic data metainformation is common in the community, like the VoID vocabulary for describing Linked datasets [1]. However, no vagueness-related vocabularies are yet available. In a more relevant approach an OWL 2 model for representing fuzzy ontologies is defined [3]. It focuses, however, on enabling the representation of fuzzy degrees and fuzzy membership functions within an ontology, without any information regarding the intended meaning of the fuzzy elements' vagueness or the interpretation of their degrees (e.g. the dimensions a concept membership degree covers). Thus, our approach is complementary to fuzzy ontology related works, in the sense that it may be used to enhance the comprehensibility of fuzzy degrees.

## 5 Conclusions and Future Work

In this paper we considered vagueness in semantic data and we demonstrated the need and potential benefits of making the latter vagueness-aware by annotating their elements with a metaontology that explicitly describes the vagueness's nature and characteristics. The idea is that even though the availability of the metainformation will not eliminate vagueness, it will manage to reduce the high level of disagreement and low level of comprehensibility it may cause. This increased semantic data comprehensibility and shareability we intend to establish in our future work through user-based experiments.

## Acknowledgement

The research has been funded from the People Programme (Marie Curie Actions) of the European Union's 7th Framework Programme P7/2007-2013 under REA grant agreement *n*<sup>o</sup> 286348.

## References

1. Alexander, K., Cyganiak, R., Hausenblas, M., Zhao, J.: Describing Linked Datasets On the Design and Usage of VoID , the "Vocabulary Of Interlinked Datasets", VoID working group, 2009.
2. Alexopoulos, P., Wallace, M., Kafentzis, K., Askounis, D.: IKARUS-Onto: A Methodology to Develop Fuzzy Ontologies from Crisp Ones. Knowledge and Information Systems, 32(3):667-695, September 2012.
3. Bobillo, F., Straccia, U.: Fuzzy ontology representation using OWL 2. International Journal of Approximate Reasoning, 52(7):1073-1094, October 2011.
4. Chandrasekaran, B., Josephson, J.R., Benjamins, V.R.: What are ontologies, and why do we need them. IEEE Intelligent Systems, pp 20-26, 1999.
5. Hyde, D.: Vagueness, Logic and Ontology. Ashgate New Critical Thinking in Philosophy, 2008.
6. Vrandečić, D., Volker, J., Haase, P., Tran, D.T., Cimiano, P.: A Metamodel for Annotations of Ontology Elements. In Proceedings of the 2nd Workshop on Ontologies and Meta-Modeling, 2006.