Characterising the Gap Between Theory and Practice of Ontology Reuse

Reham Alharbi
Department of Computer Science
University of Liverpool
Liverpool, UK
R.alharbi@liverpool.ac.uk

Valentina Tamma
Department of Computer Science
University of Liverpool
Liverpool, UK
V.Tamma@liverpool.ac.uk

Floriana Grasso
Department of Computer Science
University of Liverpool
Liverpool, UK
F.Grasso@liverpool.ac.uk

Abstract

Ontology reuse is a complex process that requires the support of methodologies and tools to minimise errors and keep the ontologies consistent. Although many efforts have investigated ontology reuse for different tasks and purposes, this body of work does not seem to translate to practice. Indeed, neither sharable good practices within similar communities have been addressed as a code of practice. The goal of this ongoing research is a comprehensive "Ontology Reusability Assessment", that builds on and extends the current state of the art ontology reuse. We carried out a community survey to gain an insight into the gap between theory and practice, in this paper we report part of the survey analysis addressing reuse approaches, ontology types, ontology reporting details and reusability challenges. The results demonstrate that soft reuse is the most widely accepted approach while selecting concepts either from domain ontologies or upper-level ontologies is highly dependent on developers' experience. Also, understanding a project's context is necessary for steering the search process in the right direction. Another important issue detected concerned reusing ontologies that import many other ontologies. A possible approach to lessen these problems could be the creation of sharable good practices recommendations that could be emerged in developing tools.

CCS Concepts: • **Information systems → Ontologies**; **Content analysis and feature selection**.

Keywords: Ontology reuse approaches, Ontology reuse practices, State of the art ontology reuse,Ontology reporting details, Reuse challenges

Woodstock '18, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06...$15.00
https://doi.org/10.1145/1122445.1122456

ACM Reference Format:
Reham Alharbi, Valentina Tamma, and Floriana Grasso. 2018. Characterising the Gap Between Theory and Practice of Ontology Reuse. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/1122445.1122456

1 Introduction

Ontology reuse is regarded as a fundamental step in all ontology design methodologies, and is seen as crucial because it allows ontologies to be developed by saving effort and by reusing ontological fragments that have been independently validated [1, 10, 18, 19, 26]. Despite that, authors of [6] suspected that ontology reuse within a given domain does not seem to be widespread. This intuition had been reinforced by the work carried out by Kamdar and colleagues [11], focused on BioPortal, which states, as one of their conclusions, that reuse among biomedical ontologies is quite limited (less than 5%). In addition, while different approaches to ontology reuse exist, due to different motivations or different implementations, the decisions on such approaches are usually taken by ontology engineers when bootstrapping a project [?]. For example, as presented in [22], ontology reuse is often not performed by means of importing the reused ontologies (hard reuse), but only by referencing the reused ontology elements URIs (soft reuse). Reusing decision methods is claimed to be biased and shallow as they are often taken on a case-by-case basis, so not necessarily meeting the definition of sharable good practice [?]. Nevertheless, reuse seems to pose a significant challenge to the community, owing to a variety of factors that can hamper the effective reuse of existing ontologies; (i) possible deficiencies in ontologies' documentation make it difficult to find all ontologies suitable for reuse [15]; (ii) the lack of a standard way to verify the accuracy of Competency Questions (CQs) against an ontology typically results in misconceptions in determining the reusability of a candidate ontology [31]; (iii) the lack of standardisation in designing an ontology through reuse can introduce future errors, for example, by failing to keep track of changes in reused ontologies [6]; (iv) insufficient information about the requirements that ontology engineers aim to satisfy when assessing candidate ontologies to reuse [6, 15].

From state of the art reviewed, it appears that some ontologies are reused more than others, possibly because they were documented when they were compared with others [21]. One of the aspects affecting the reusability of an ontology is the perception of its quality with respect to some evaluation criteria. However, although much work has addressed the evaluation of the quality of ontologies from the perspective of reuse such as [4, 7, 28–30], there are no definitive principles and practices, and, especially, there is no practical mechanism for providing developers with a qualitative and quantitative assessment of reusability.

In this ongoing research we aim to develop a comprehensive framework for assessing ontologies for reuse, capturing the requirements for reusing ontologies, identifying the modalities of reuse, the ontology features, and any additional information of which ontology engineers need to be aware of when selecting a candidate ontology to reuse.

To achieve this objective, we analysed state of the art reusing ontologies in a top-down and bottom-up manner: we reviewed the current literature on the theory of ontology reuse, whilst we compiled a questionnaire aimed at ontology practitioners and experts to identify mismatches between theory and common practices, condense best practices and establish a mechanism on how to assess an ontology for reuse.

We asked the survey participants (Section 3) to rank several measures, which exemplify different ontology evaluation metrics, from 'most preferred' to 'least preferred' with respect to the reuse of ontologies. Such reuse metrics are "scheme metric" [21, 28, 30], "instance metric" [21, 28, 30], "community and social metric" [15, 29], and "documentation metric" [5, 15, 29]. However, the aim of this paper is to report part of the survey analysis addressing the following research objectives:

- **Objective 1(Obj1) : Identify how people practically reuse**. The questions associated with this objective are the following: how do developers reuse in practice? What are preferred ontology reuses, definitions and functions? Do ontologists prefer to reuse domain specific ontologies or upper level ontologies?
- **Objective 2 (Obj2) : Identify the significant reported information about ontologies that are practically investigated**. This objective is focused on the question: what are the aspects that developers are searching for?
- **Objective 3 (Obj3) : Identify the important empirical obstacles that have hampered reuse**. The essential feature of ontologies that might reduce their reusability.

The main findings of our work are that soft reuse is significantly more popular than hard reuse and reusing ontology design patterns. In addition, experience highly influences the decision to reuse specific domain ontologies or upper-level ontologies. Furthermore, ontologies complexity concerned ontologists and typically hampered reuse, mainly being related to ontology size, conflicts within ontology and importing many ontologies. The remaining sections of this paper are organised as follows: Section 2 presents the research methodology we followed for the study. Section 3 defines the survey participants, Section 4 provides the results obtained during the study while Section 5 offers a discussion of such results. Section 6 focuses on related work and finally, conclusions and future trends are outlined in Section 7.

2 Research methodology

An empirical method through an online questionnaire aimed at ontology developers was followed to fulfil the objectives mentioned earlier (Section 1). The responses were processed in a quantitative way; they were analysed in statistical tests that assess the extent to which these objectives are associated with participants' experience in ontology engineering. The particular steps carried out are presented below:

1. Definition of selective variables.
2. Evaluation of the variables by means of data collection.
3. Data analysis by means of mainly descriptive statistics.

The following (Section 2.1 to Section 2.3) show how each one of these steps were performed.

2.1 Definition of variables

To carry out this step, we reviewed the state of the art reusing ontology to define several aspects of preferences related to each objective. We selected papers from three main standpoints: recognised efforts to define reuse from different perspectives [12, 22, 23]; proposals for meaningful guidelines/ recommendations for ontology selection aiming to increase the reusability of ontologies [4–6, 15, 21, 24, 27, 28, 30]; ultimately, studies that highlighted difficulties and challenges of reuse ontologies [6, 13, 17].

The following variables represent ontology reuse definitions and methods, a question related to these variables is *How would you typically reuse ontologies?*

- Import the whole ontology (Hard reuse or As-is).
- Reference to the ontology elements URIs (Soft reuse).
- Reuse by removing some original axioms (Extraction)
- Reuse by adding new axioms to the ontology (Extension)
- Reuse by merging several ontologies (Combination)
- Reuse by extracting / using Ontology Design Patterns (ODPs)
- Reuse by copying ontology's structure.

The following variables account for features that developers would investigate and might influence developers' decision to reuse. The question related to these variables is: *Rank the following ontology details in order of importance from 1 to 8?*

- Ontology purpose
- Ontology scope
- Target groups and intended uses of ontology
- Ontology developers

- The language used to express the ontology
- Requirement specification document (RSD)
- Competency questions (CQs)
- Ontology design specifications (classes, properties and relationship)

Finally, variables account for aspects that might hamper reuse. The question regarding these variables is as follows: *Rank the following potential obstacles to the reuse in order of significance from 1 to 5 ?*

- Size
- Number of imports (e.g. importing a number of ontologies, that in turn import other ontologies)
- Complexity (large number of classes, axioms and properties)
- Versioning, the existence of different versions of the same ontology.
- Known conflicts within ontology

Let us note that the variables were evaluated in relation to the participants' backgrounds, the latter was determined by the participants at the start of the questionnaire.

2.2 Evaluation of the variables by means of data collection

In the following section, we will describe the different tasks, motivations, and the most decisive features. The survey contains; (i) Popularity tasks for **Obj1** to investigate the most popular definition of reuse that was practically used and how this differs based on the participant's experience. (ii) Ranking tasks for **Obj2** and **Obj3**, each covering different aspects of the engineers' decision-making process. The goal of these tasks was to aggregate and condense the participants' experience and 'gut-feeling' about empirical attitudes.

1. **Popularity Task for Reuse vs. Experience**, this task covers the topic of defining ontology reuse. We provided the participants with seven approaches of ontology reuse implementation along with a brief description.
2. **Popularity Task for Ontology Type vs. Experience**, this task covers the topic of the popularity of upper-level ontologies vs. specific domain ontologies within the community. We provided the participants with examples that illustrate both cases and asked them to select the most preferred choice and give a reason.
3. **Ranking Task for Details vs. Experience**, the first ranking task is about ontology reporting details vs. engineers' background. We provided the participants with eight features, which are common when searching for ontologies and included in most popularity-based metrics.
4. **Ranking Task for Obstacles vs. Experience**, the second ranking task is about ontology reuse obstacles vs. engineers' background. We provided the participants with five difficulties, taken mainly from studies that investigated *why are ontologies not reused?* and studies that highlighted difficulties and challenges of reusing ontologies in the literature.

2.3 Data analysis

The data analysis [1] was carried out by means of descriptive statistics; a number of statistical tests used due to different analysis purposes and variant data types, such as the **Chi-square test** to determine whether there is a statistically significant difference between the expected frequencies and the observed frequencies in one or more categories of a contingency table; the **Friedman test** a non-parametric alternative to the one-way ANOVA with repeated measures used for ranking tasks as the answers (dependent variables) on an ordinal scale, used to detect significant differences between aspects for each objective (with α =.05); the **pairwise Wilcoxon signed-rank tests with Bonferroni correction** to determine if two or more sets of pairs are different from one another in a statistically significant manner.

3 Participants

The target population of this study was the Semantic Web community either in academia or industry. The questionnaire was disseminated during the period March to November 2020 through personal contacts, and relevant mailing lists and conferences (e.g. the 19th International Semantic Web Conference). The geographical distribution of the responses originated primarily from Spain 15%, UK 13% and USA 13%. Overall, N= 78 participants completed the survey, but only N = 54 were eligible to take part, because we excluded respondents who classed themselves as having no ontology engineering experience as well as a subset answered of the questions resulted in the discrimination of that response.

On average the participants have worked for 4 to 5 years with ontology engineering (M =4:75; SD=2:64), and rated their own expertise with ontology engineering quite high (M =3:87; SD =0.933) on a 5-point-Likert scale from 1 (none at all experienced) to 5 (expert). About 65% of the participants consider themselves to be Highly Experienced (HE) or above (4 or 5 on the Likert-scale) and 35% consider themselves to have Moderate Knowledge (MK) or less. They also declared their attitude toward reusing ontologies quite high (M =3:93; SD =0.968) on a 5-point-frequency scale from 1 (rarely) to 5 (always). About 68% of the participants consider reusing ontologies high frequent (often to always on the Frequent-scale) and none of the participants choose never consider reusing. The frequency of reusing ontologies differ depending on

[1] We thank Statistical Department at the University of Liverpool for their contribution to statistical analysis.

the familiarity of the respondents with ontology engineering, as shown by a Mann-Whitney U test (U =200.500, p = 0.012). The highly experienced participants reuse ontologies more frequently than the moderate knowledge ones (add the percentages).

In total, we can say that our participants are quite experienced in the field of ontology engineering and they are willing to reuse ontologies. This makes the results of our survey very promising with respect to their validity for identifying the shareable good practices of reuse.

4 Results

We encode the obtained both popularity and ranking positions with numbers starting at 1,2, and so on, i.e the lower the ranking number the better rank position obtained from the participants. Table ??, Table ??, summarises the results for each group and gives a first insight into how the aspects of each task was ranked presented in the second and fourth columns. Also, including the significant differences between the rankings which are provided in the third and fifth columns. Let's note that all tasks were completed by N=54 respondents, however, we categorised the analysis based on the respondents' experience; (HE) Highly Experienced, and (MK) Moderate Knowledge.

4.1 Popularity Task for Reuse vs. Experience

The analysis showed that the most popular reuse approach was Soft reuse as selected 74.3% HE, and 47.4% MK, while the least favourable was Reuse by copying ontology's structure as selected 22.9% HE, and not selected by MK. The association between different approaches and participants' experience were;

- *Reuse approaches affected by participants' experience* are:(i) hard reuse; a significant association between familiarity level and this method. In other words, this approach is dependent on the engineers' experience, selected 57.1% HE and 26.3% MK. ($x^2(1, 54) = 4.70, p = .030$)
  (ii) soft reuse: which is the participant' best preferred method, the analysis showed a significant association between participants' experience and this approach ($x^2(1, 54) = 3.91, p = .048$)
  (iii) copying ontology's structure; that received the least attention as reported only 22.9% HE. Furthermore, the chi-square result ($x^2(1, 54) = 5.098, p = .024$) highlights that this approach is dependent on engineers' experience.
- *The chi-square results for the rest approaches showed no relationship with the participants' experience* the reported preferences varied between both groups. The following are the approaches in descending order from the most selected approach to the least.

1. For HE; extension 51.4%, using ODP 34.3%, extraction 28.6%, and combination 25.7%.
2. For MK; extension 42.1%, extraction and combination same percentage 31.6%, and using ODP 21.1%.

4.2 Popularity Task for Ontology Type vs. Experience

The analysis showed that reusing specific domain ontologies was dependent on participants' experience mainly preferred by MK as reported HE 31% and MK 78%. In contrast, reusing upper level ontologies was highly recommended by HE reported HE 74% and MK 21%.

4.3 Ranking Task for Details vs. Experience

- *For the HE, N= 35, we found a significant difference between reported information*, $\chi^2(7, 35) = 66.482, p < .001$ Dunn-Bonferroni post hoc tests were carried out and there were significant differences between ontology scope and ontology language (p=.002), ontology scope and target group (intended uses of ontologies) (p=.000), ontology scope and CQs (p=.000), ontology scope and RSD (p=.000), ontology scope and ontology developers (p=.000), ontology design specifications and ontologies developers (p=.000), ontology purpose and ontology developers (p=.000), ontology language and ontology developers (p=.016), target group (intended uses of ontologies) and ontology developers(p=.050), the above mentioned results obtained after Bonferroni adjustments. However, there were no significant differences between any other details.
- For MK, N= 19, we found a significant difference between the eight reported information, ($x^2(7, 19) = 36.243, p < .001$) A post hoc analysis with Wilcoxon signed-rank tests, which conducted with a Bonferroni correction applied and it proved that; there were significant differences between target group (intended uses of ontologies) and RSD (p=.052), target group (intended uses of ontologies) and ontology language (p=.006), target group (intended uses of ontologies) and ontology developers (p=.000), ontology purpose and ontology language (p=.052), ontology purpose and ontology developers (p=.001), ontology scope and ontology developers (p=.010), again the above mentioned results obtained after Bonferroni adjustments. However, there were no significant differences between any other aspects.

4.4 Ranking Task for Obstacles vs. Experience

- For HE, N= 35,there was a significant difference between the five obstacles, ($x^2(4, 35) = 13.305, p < .010$) A further post hoc analysis with Wilcoxon signed-rank tests, which were performed with a Bonferroni correction applied and it proved that there were significant differences between conflicts with other imported ontologies and ontology versioning (p=.032) after Bonferroni adjustments. However, there were no significant differences between any other obstacles.

- Regarding MK, N= 19, again, a significant difference between the five obstacles ($x^2(4, 19) = 28.084, p < .001$) A post hoc analysis with Wilcoxon signed-rank tests, which were conducted with a Bonferroni correction applied and provided final evidence that there were significant differences between number of imports and conflicts with other imported ontologies (p=.021), conflicts with other imported ontologies and ontology versioning (p=.001), conflicts with other imported ontologies and ontology size (p=.000) after Bonferroni adjustments. However, there were no significant differences between any other obstacles.

5 Discussion

The results of analysing the aforementioned objectives (Section 1) to investigate ontology reuse practices show that engineers' practices ranged from practices that are quite different from theory to those that are confirmed by the state of the art ontology reuse. In [?], ontology selection is claimed to be a highly subjective task that is frequently performed by experienced ontology engineers who select ontologies that they intuitively believe are fit for the purpose. Our results could argue this by indicating that even experienced engineers have preferences for measures that influence their decision to reuse ontologies. Nevertheless, investigating the association between these measures and participants' experience showed that some reusability measures were entirely dependent on the experience level, while others confirmed no relatedness to the experience. This could draw the conclusion that there are some reusability aspects that are widely accepted as shareable good practices within the community.

The following are the key findings that will be discussed with respect to which finding contradict the state of the art ontology reuse and which confirmed with it under two headlines; *Engineers' Practical Measures* (for the former), and *Agreed Theoretical Measures* (for the latter).

5.1 Engineers' Practical Measures

1. **Ontology reuse approaches** This set of results is regarding the preferences of reuse approaches, the significant preferred approach for all participants is soft reuse. This is quite an interesting result as it matches our expectations of the flexibility of this approach in terms of modularity, allowing ontologists to decide which parts of an external ontology are reused rather than importing the entire ontology [2]. As a matter of fact, this appeared to be the developers' attitude several years ago. For example, a notable evolution in 2012 of ontology reuse practices from the hard reuse of ontologies to the soft reuse of ontologies' terms showed an increased to approximately half of the total reuses [22]. Two years later, the empirical survey conducted by [24] showed that direct reuse of popular vocabularies is preferred rather than defining new terms and aligning them to other vocabulary terms. Later, in 2020 [?] authors emphasised that ontology reuse practices in domain applications confirm the aforementioned trend. Surprisingly, looking at the quite small engineers preferences for hard reuse in both groups, while in theory most of the ontology development frameworks such as Protege (owl:imports) support it. The reasons are related to; (i) complexity of hard reuse as it does not allow the customisation of the imports of ontologies which results in import concepts that may be irrelevant and incompatible with the requirements of the local ontology [?]. (ii) inaccurate consequences of hard reuse exemplified by firstly, if importing ontologies are not available so their semantics are lost, causing inconsistency of the new ontology, and secondly, if a reused ontology contains contradictions with respect to a reused ontology, these cannot be detected automatically [?]- contradictions may emerge only when both ontologies are imported into a third ontology [6].
   Another significant finding which opposed the state of the art reuse ontologies is the ODP. In theory it was considered the most beneficial component in ontology design as well as to be the reusable modelling solutions which encode modelling best practices [8, 23]. However, the participants showed least interest in using this style of reusing. Also noticed was the limited influence of ODP in published ontologies as stated by different studies. For example, a variety of techniques developed by [14], to detect the prevalence of ODPs in ontologies showed that no evidence could be found of the influence of ODPs, even for the most popular, within biomedical ontologies. Similarly, [16] conclude that there is only scant evidence for influence of ODPs in the biomedical domain. Given the popularity of biomedical ontologies, the conclusion of these studies can be generalised to other domains.
2. **Domain ontologies vs.Upper level ontologies** We noticed here that the results were entirely dependent on the experience level of the participants showing that MK engineers prefer to reuse specific domain ontologies while HE engineers prefer upper- level ontologies. The underlying theories are that; (i) MKs' limited experience may be a barrier, leading them to focus on a clear data structure provided by domain ontologies rather than the entire semantics of the data, which may be missed by these ontologies. (ii) HEs' confidence in knowing how to avoid reusability problems and experience in the field along with networking and

Details	MK Median Rank	Friedman test	HE Median Rank	Friedman test
Ontology purpose	7		6	
Ontology scope	6		8	
Ontology target group	8		4	
Ontology developers	1	$\chi^2(7, 19) = 36.243, p < 0.001$	1	$\chi^2(7, 35) = 66.482, p < 0.001$
Ontology language	2		5	
Requirement specification document	3		2	
Competency questions	5		3	
Ontology design	4		7	

Table 1. Ontology Reporting Details

involvement in long-term projects would be the reasons for preferring upper-level ontologies. Those ontologies are typically owned and maintained by large, well-known organisations so unavailability, versioning, and miscommunications issues are rarely to be faced. To the best of our knowledge, addressing the link between reusing earlier mentioned ontologies and developers' experience has not been investigated in the literature. Regardless of the developers' background, other factors have been suggested to help developers in deciding which vocabularies to reuse, focusing on specific domains such as governmental data[2]. Furthermore, [5], developed methodological guidelines for reusing general ontologies to identify the type of general ontology to be reused and what axioms and definitions of such ontologies to be extracted.

5.2 Agreed Theoretical Measures

1. **Ontology reporting details** The results showed variances in ranking positions based on the participants' experience. Despite that, both groups indicated that the most important aspect is ontology developers, this is also confirmed by some literature such as MIRO guidelines [15] pointed out that Ontology Owner details as "must be included" in ontology documentation. [27] further emphasised that one of the qualities that make ontologies attractive for reuse is when ontologies are owned or published by large organisations, showing that developers' information should be clearly stated. Previously, this detail was not regarded as interesting for example, ontology requirement specification documents RDS by [27] did not include this information as a part of ontologies documentation. The ranking results of purpose, scope, CQs, and intended users are debateable in a two-fold way: Firstly, apparently, those aspects are parts of RSD, widely known by the community, and having them as individual options could be misleading. This is because of the association between them like parent->child or class->subclass relationships, this might justify the lowest rank of these details compared to RSD. Secondly, regarding the different practices and experiences, it is interesting to observe that engineers have different attitudes toward prioritising these aspects which influence ontology' reusability from different perspectives.
2. **Ontology reuse difficulties** There was no difference between the ranking positions between all participants in the first three aspects of this task. That leads us to conclude that practical factors that hampered reuse in order of significance are: conflicts within ontologies, complexity, and number of imports. These factors have been addressed and agree with the state of the art ontology reuse, as ontologists have been advised by [17] to avoid a "Pick and Mix" approach when reusing ontologies, describing as this as potentially confusing. The author [17] suggested drawing terms from the same ontology rather than importing many ontologies, this ensures overall coherence, and allows reasoning to be better leveraged.
Ontology size is another challenge for reuse determined by both groups; this could be related to; (i) the broadening scope issue raised by [17] to which a possible solution is to develop ontologies with fixed scope and core concepts and avoid shadow concepts – having this in mind would increase the overall coherence and connectivity of the information that ontology describes. (ii) size concern could be related to the 'Mismatches and Misunderstandings' and 'Finding Mr. Right Ontology' challenges identified by [13] – the relationship between them could be confusing and ambiguous with respect to the size of reused ontologies. Finally, ontology versioning is an debateable point and could be a double-edged sword. On one hand, having updated versions of an ontology means that it is frequently maintained and up-to-date. On the other hand, versioning typically causes loss of information resulting in an ontology that imports out-of-date or unavailable ontologies, the former highlighted by [6].

6 Related work

[2] https://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook#Step_3_

Details	MK Median Rank	Friedman test	HE Median Rank	Friedman test
Ontology size	5		4	
Number of imports	3		3	
Complexity	2	$\chi^2(4, 19) = 28.084, p < 0.001$	2	$\chi^2(4, 35) = 13.305, p < 0.010$
Ontology versioning	4		5	
Conflicts within ontology	1		1	

Table 2. Ontology Reusing Obstacles

Despite a vast literature on ontology reuse providing definitions, methodologies, implementations, evaluation and verification metrics, we have reported those studies most related to the aim of this paper. This is neither limiting the review process nor neglecting other efforts. The aim of this study was to determine the practical attitudes and preferences toward reuse. The hypothesis underlying this is: The variances of reuse among different repositories would implicitly confirm that reuse is influenced by shareable good practices within the community. One of the first analyses on ontology reuse was by Simperl [25]. The author presented different use cases on ontology reuse as well as methods and tools. The main conclusion was that the reuse approach is a decision-making problem in which the developer's experience plays a crucial role. Nevertheless, shareable good practices within the same repository would increase reusability, for example, authors analysed the reuse in 53 ontologies at three different time points (September 2009, March 2010, and September 2010). The results indicate that reuse among the OBO Foundry candidates increased during the study period [9]. Also, Poveda and colleagues [22] analysed soft and hard reuse in the Linked Data context. This study involved over 73.96% of the 35 ontologies included in the LOV registry. The study showed that the proportion of reused elements was noticeably high (40.53%). In addition, in that time, hard reuse was more frequent than soft reuse, although soft reuse already represented one quarter of the total. Later on, reuse shows an opposed evolution mainly addressed by Fernández-López et al. [6]. More than half of the ontologies on the LOV registry reuse whilst others do not import any entity and the proportion of soft reuse has increased, reaching approximately half of total reuse. Focusing on the BioPortal repository, Kamdar and colleagues [11] analysed the reuse in 377 biomedical ontologies. As a general conclusion, reuse among biomedical ontologies is quite limited (less than 5%). Let us note that this percentage is much lower than that obtained by the above-mentioned studies, [22] and [6]. Consequently, this reflects that the reuse percentage in LOV is a great deal higher than that of BioPortal. Moreover, Ochs and colleagues [20] made an analysis of 355 ontologies in BioPortal. As a result of this analysis, it can be stated that 55% of the analysed ontologies includes elements (classes or properties) from other ontologies. In addition, the work showed that knowledge reuse is especially widespread among the ontologies in the OBO Foundry. Moreover, the knowledge reused belongs to a small set of popular ontologies, which makes popularity one of the key practices in reuse, detected firstly by [11] and addressed later by Schaible and colleagues [24]. They carried out a survey to obtain rankings of diverse modelling examples. Participants in the survey took into account their understanding of the good reuse of vocabularies when ranking the models. This study showed that the trend has been to reuse popular terms from frequently-used ontologies instead of using terms from unpopular ontologies. On the other hand, a statistical study carried by [6], detected that the heterogeneity between the needed conceptualisation and that of available ontologies, as well as the deficiencies in some of such ontologies (concerning for example documentation and licencing), are important obstacles to reusing ontologies of the same domain of the ontology under development.

7 Conclusion and future work

We have argued that despite community efforts promoting reusability as a key step in ontology development and a vast literature of reuse approaches, requirements, methodologies and metrics, reusability practices are not yet widespread. We have presented part of the study that investigates reusability in practice focusing on reuse approaches, ontology types, ontology reporting details and reusability challenges. This has been investigated using a survey consisting of ranking tasks, where the participants were asked to rank various aspects according to their real-life implementation and understanding of good reuse practices. The results demonstrate that soft reuse is the most widely accepted approach, while selecting concepts either from domain ontologies or upper-level ontologies is highly dependent on developers' experience. Also, understanding a project's context is necessary for driving the searching process to the appropriate place [17], raising the importance of non-functional requirements appearing in ontology documentation such as RSD and CQs, which are fundamental for the reuse decision. Another important issue detected concerned reusing ontologies that import many other ontologies. It is recommended that developers maintain an appropriate mix and match, in order to provide a clear data structure and make it easier to be reused [24]. A further significant, related finding to this is that ontology complexity hampered reuse, which could be avoided by developers determining the clear scope, and specific purposes, with the attached documentation of published ontologies. These findings of our survey will also be used for the ongoing research of assessing ontology reusability and will emerge as practical recommendations with tools such as Watson [3] for the NeOn ontology engineering toolkit. An evaluation study including expert ontologists and community members will ensure that the reuse assessment is both practical and relevant.

References

[1] E Blomqvist, K Hammar, and V Presutti. 2016. Engineering Ontologies with Patterns-The eXtreme Design Methodology. In *Ontology Engineering with Ontology Design Patterns: Foundations and Applications*. Studies on the Semantic Web, Vol. 25. IOS Press, 23–50.

[2] V A Carriero, M Daquino, A Gangemi, A G Nuzzolese, S Peroni, V Presutti, and F Tomasi. 2020. The Landscape of Ontology Reuse Approaches. In *Applications and Practices in Ontology Design, Extraction, and Reasoning*. Studies on the Semantic Web, Vol. 49. IOS Press, 21–38.

[3] M d'Aquin, C Baldassarre, L Gridinoc, M Sabou, S Angeletou, and E Motta. 2007. Watson: Supporting next Generation Semantic Web Applications. In *IADIS International Conference on WWW/Internet*. IADIS Press, 363–370.

[4] A Duque-Ramos, J T Fernández-Breis, R Stevens, and N Aussenac-Gilles. 2011. OQuaRE: A SQuaRE-Based Approach for Evaluating the Quality of oOtologies. *Journal of Research and Practice in Information Technology* 43, 2 (2011), 159–176.

[5] M Fernández-López, A Gómez-Pérez, and M C Suárez-Figueroa. 2013. Methodological Guidelines for Reusing General Ontologies. *Data & Knowledge Engineering* 86 (2013), 242–275.

[6] M Fernández-López, M Poveda-Villalón, M C Suárez-Figueroa, and A Gómez-Pérez. 2019. Why are Ontologies not Reused Across the Same Domain? *Journal of Web Semantics* 57 (2019), 100492.

[7] A Gangemi, C Catenacci, M Ciaramita, J Lehmann, R Gil, F Bolici, and O Strignano. 2005. *Ontology Evaluation and Validation: An Integrated Formal Model for the Quality Diagnostic Task*. Technical Report. Laboratory of Applied Ontologies – CNR. http://www.loa.istc.cnr.it/old/Files/OntoEval4OntoDev_Final.pdf Last Accessed 12-09-2021.

[8] A Gangemi and V Presutti. 2009. Ontology Design Patterns. In *Handbook on ontologies*. Springer, 221–243.

[9] A Ghazvinian, N Fridman Noy, and M A Musen. 2011. How Orthogonal are the OBO Foundry Ontologies? *Journal of Biomedical Semantics* 2, 2 (2011), 1–14.

[10] T R Gruber. 1993. A Translation Approach to Portable Ontology Specifications. *Knowledge acquisition* 5, 2 (1993), 199–220.

[11] M R Kamdar, T Tudorache, and M A Musen. 2017. A Systematic Analysis of Term Reuse and Term Overlap across Biomedical Ontologies. *Semantic Web* 8, 6 (2017), 853–871.

[12] M Katsumi and M Grüninger. 2016. What is Ontology Reuse?. In *FOIS 2016 - 9th Conference on Formal Ontology in Information Systems (Frontiers in Artificial Intelligence and Applications, Vol. 283)*. IOS Press, 9–22.

[13] M Katsumi and M Grüninger. 2017. Choosing Ontologies for Reuse. *Applied Ontology* 12, 3-4 (2017), 195–221.

[14] C Kindermann, B Parsia, and U Sattler. 2019. Detecting Influences of Ontology Design Patterns in Biomedical Ontologies. In *ISWC 2019 – 18th International Semantic Web Conference (LNCS, Vol. 11778)*. Springer, 311–328.

[15] N Matentzoglu, J Malone, C Mungall, and R Stevens. 2018. MIRO: Guidelines for Minimum Information for the Reporting of an Ontology. *Journal of biomedical semantics* 9, 1 (2018), 1–13.

[16] J Mortensen, M Horridge, M A Musen, and N Fridman Noy. 2012. Modest Use of Ontology Design Patterns in a Repository of Biomedical Ontologies. In *3rd International Conference on Ontology Patterns (WOP'12)*. CEUR-WS.org, 37–48.

[17] Chris Mungall. 2021. How to Select and Request Terms from Ontologies. Blog: Monkeying around with OWL, July 3, 2021 entry. https://douroucouli.wordpress.com/2021/07/03/how-select-and-request-terms-from-ontologies/

[18] R Neches, R E Fikes, T Finin, T Gruber, R Patil, T Senator, and W R Swartout. 1991. Enabling Technology for Knowledge Sharing. *AI magazine* 12, 3 (1991), 36–56.

[19] N Noy and D L McGuinness. 2001. *Ontology Development 101*. Technical Report KSL-01-05. Knowledge Systems Laboratory, Stanford University.

[20] C Ochs, Y Perl, J Geller, S Arabandi, T Tudorache, and M A Musen. 2017. An Empirical Analysis of Ontology Reuse in BioPortal. *Journal of Biomedical Informatics* 71 (2017), 165–177.

[21] J Park, S Oh, and J Ahn. 2011. Ontology Selection Ranking Model for Knowledge Reuse. *Expert Systems with Applications* 38, 5 (2011), 5133–5144.

[22] M Poveda Villalón, M C Suárez-Figueroa, and A Gómez-Pérez. 2012. The Landscape of Ontology Reuse in Linked Data. In *EKAW 2012 Workshop on Ontology Engineering in a Data-driven World (OEDW)*.

[23] V Presutti, E Blomqvist, E Daga, and A Gangemi. 2012. Pattern-Based Ontology Design. In *Ontology Engineering in a Networked World*. Springer, 35–64.

[24] J Schaible, T Gottron, and A Scherp. 2014. Survey on Common Strategies of Vocabulary Reuse in Linked Open Data Modeling. In *The Semantic Web: Trends and Challenges 11th International Conference, ESWC 2014 (LNCS, Vol. 8465)*. Springer, 457–472.

[25] E Simperl. 2009. Reusing Ontologies on the Semantic Web: A Feasibility Study. *Data & Knowledge Engineering* 68, 10 (2009), 905–925.

[26] R Studer, V R Benjamins, and D Fensel. 1998. Knowledge Engineering: Principles and Methods. *Data & Knowledge Engineering* 25, 1-2 (1998), 161–197.

[27] M C Suárez-Figueroa, A Gómez-Pérez, and M Fernández-López. 2012. The NeOn methodology for ontology engineering. In *Ontology Engineering in a Networked World*. Springer, 9–34.

[28] K Supekar, C Patel, and Y Lee. 2004. Characterizing Quality of Knowledge on Semantic Web. In *17th International Florida Artificial Intelligence Research Society FLAIRS*. AAAI Press, 472–478.

[29] M Talebpour, M D Sykora, and T Jackson. 2017. The Role of Community and Social Metrics in Ontology Evaluation: An Interview Study of Ontology Reuse. In *9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management KEOD 2017*. SCITEPRESS, 119–127.

[30] S Tartir, I B Arpinar, M Moore, A P Sheth, and B Aleman-Meza. 2005. OntoQA: Metric-Based Ontology Quality Analysis. In *IEEE ICDM Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Source*. 45–53.

[31] D Wiśniewski, J Potoniec, A Ławrynowicz, and C M Keet. 2019. Analysis of Ontology Competency Questions and their Formalizations in SPARQL-OWL. *Journal of Web Semantics* 59 (2019), 100534.