The Evaluation of Ontologies: Quality, Reuse and Social Factors

by

Marzieh Talebpour

A Doctoral Thesis

Submitted in partial fulfilment of the requirements

for the award of

Doctor of Philosophy of Loughborough University

3rd July 2019

ABSTRACT

Finding a "good" or the "right" ontology is a growing challenge in the ontology domain, where one of the main aims is to share and reuse existing semantics and knowledge. Before reusing an ontology, knowledge engineers not only have to find a set of appropriate ontologies for their search query, but they should also be able to evaluate those ontologies according to different internal and external criteria. Therefore, ontology evaluation is at the heart of ontology selection and has received a considerable amount of attention in the literature.

Despite the importance of ontology evaluation and selection and the widespread research on these topics, there are still many unanswered questions and challenges when it comes to evaluating and selecting ontologies for reuse. Most of the evaluation metrics and frameworks in the literature are mainly based on a limited set of internal characteristics, e.g., content and structure of ontologies and ignore how they are used and evaluated by communities. This thesis aimed to investigate the notion of quality and reusability in the ontology domain and to explore and identify the set of metrics that can affect the process of ontology evaluation and selection for reuse.

A mixed methods approach was used in this research. The first qualitative phase of this study explored the perspective of ontologists and knowledge engineers and identified the key factors in the ontology evaluation and selection process. A survey questionnaire was then used to confirm the importance of the set of quality metrics identified in the first phase, and to compare them to the ones employed in the literature. Together, the findings of the first two phases were used to propose a new set of quality metrics and a framework for ontology evaluation, which were then validated with the experiments conducted in the final phase of this research.

The findings of this study suggested that the process of ontology evaluation and selection for reuse not only depends on different internal characteristics of ontologies, but it also depends on different metadata and social and community related metrics. This study not only identified a set of metrics that can be used in the evaluation process but also investigated how important each of those metrics was in the evaluation process. It is interesting to note that participants in different phases of this study found many of the metrics proposed by this research more important and also more helpful in the selection process, compared to the ones that are already being used by the existing selection systems. Therefore, the findings of this research can contribute to facilitating and improving the process of selecting an ontology for reuse.

مامان

بابا

مژگان

شما تمام سهم من از عشق و خوشبختی هستید

ACKNOWLEDGEMENTS

Tom and Martin, I still have the first email I sent you to "express my interest in joining your team" and your reply which was "I would encourage you to apply". Now, after exactly four years and one month, I would like to thank you for all the encouragement, kindness and support. I was and still am very lucky to have you two as my supervisors.

Maman and Baba, thank you for all your love, support and for helping me to follow my dreams; you are the BEST. Being far from you has never been easy and will never get easier, but I hope calling me "Khanom Doctor" makes you proud. Baba, thanks for telling me "man pat vaysadam" and "har etefaghi biyofte to barande hasti"; this really means a lot to me! Mozhgan, you are the best sister. I LOVE YOU all.

I would also like to thank my friends and all members of staff at the School of Business and Economics, for their support.

And finally, a note to myself:

You are finishing this very important phase of your life, and you'll probably (and hopefully) move on to start a new life and experience new things. In the life that you have ahead of you, some days will be full of sunshine and happiness and love and laughter. Other days, however, might not be the same. Either way, you need to always remember that none of those days will last forever. So, never lose hope and never forget how strong you are….

TABLE OF CONTENTS

List of Figures

List of Tables

Publications

The research that is presented in this thesis has led to the following publications:

1. Talebpour, M., Sykora, M.D. and Jackson, T. (2017) 'The role of community and social metrics in ontology evaluation: an interview study of ontology reuse', in Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KEOD 2017). Funchal, Madeira, Portugal, 1st-3rd November. pp.119-127.

2. Talebpour, M., Sykora, M.D. and Jackson, T. (2018) 'Ontology selection for reuse: Will it ever get easier?', in Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KEOD 2018). Seville, Spain, 10-20th Sept. Setubal, Portugal: Scitepress, Vol 2, pp. 108-116.

3. Talebpour M., Sykora M., Jackson T. (2019) Social and Community Related Themes in Ontology Evaluation: Findings from an Interview Study. In: Fred A. *et al.* (eds) Knowledge Discovery, Knowledge Engineering and Knowledge Management. IC3K 2017. Communications in Computer and Information Science, vol 976. Springer, Cham.

Work in Progress:

1. The Evaluation of Ontologies for Quality, Suitability for Reuse, and the Significant Role of Social Factors. Submitted (in May 2019) to the CCIS Series book series published by Springer.

Chapter 1: INTRODUCTION TO THE RESEARCH

1.1 Introduction

This chapter aims to provide an introduction to the work presented in this research. It starts by describing and exploring the research background in section 1.2, and evaluation challenges in section 1.2.1. The research questions and aim of the research are then presented in section 1.3 and section 1.4 respectively. A short overview of the research method used to address the aim and objectives is provided in section 1.5. The expected contribution of this research is discussed in section 1.6. Finally, section 1.7 provides an overview of the thesis chapters.

1.2 Background

Ontologies play a significant role in the field of knowledge and information management by furnishing the semantics to the semantic web (Shadbolt, Berners-Lee and Hall, 2006). Ontologies are used in different domains for various purposes and have many benefits. They facilitate communication and knowledge transfer between systems, between humans, and between humans and systems (Bürger and Simperl, 2008; Feilmayr and Wöß, 2016) by uniquely identifying the meaning of different concepts in any domain. They can also avoid the costs associated with new developments of knowledge models (Butt, Haller and Xie, 2016).

Despite the significant role that ontologies play in the semantic web, ontology development has always been a challenging task (Ding and Foo, 2002; Lim, Liu and Chen, 2015; Fernández-López *et al.*, 2019). Ontology development is a costly process (Butt, Haller and Xie, 2016), and some argue that the cost of building and maintaining ontologies in some domains can outweigh the potential benefits gained by using them (Shadbolt *et al.*, 2006). Building ontologies from scratch was the only option in the early days; however, an increase in the number and availability of ontologies has provided developers with the chance of reusing ontologies. Therefore, ontology reuse, using an existing ontology as the basis for building a new one, has been suggested as a solution to address some of the challenges of the ontology development process (Kamdar, Tudorache and Musen, 2017).

Ontology reuse is beneficial to the community of ontologists and knowledge engineers (Fernández-López *et al.*, 2019). From the very early days, it has been suggested that the future of construction of large-scale knowledge-based systems is highly dependent on reusing the components built by others (Uschold, Clark, *et al.*, 1998). Reuse will not only help in achieving one of the primary goals of ontology construction, that is to share and reuse them (Simperl, 2009), but will also reduce redundancy (Zulkarnain, Meziane and Crofts, 2016) and save a significant amount of time and financial resources (Bontas, Mochol and Tolksdorf, 2005; Dramé *et al.*, 2014). Ontology reuse is also helpful in supporting interoperability between different applications and datasets (Kamdar, Tudorache and Musen, 2017).

Different ontology development approaches have emphasized the importance of reusing previously built components while developing new ones. A review of ontology engineering methodologies conducted by Iqbal *et al.*, (2013) highlighted that many of the methodologies advocate ontology reuse (10 out of 14) and have considered the reuse step in their method. Ontology reuse consists of different steps: searching for adequate ontologies, evaluating the quality and fitness of those ontologies for the reuse purpose, selecting an ontology and integrating it in the current project (d'Aquin *et al.,* 2008).

Despite all the advantages of reusing ontologies and the availability of different ontologies, ontology reuse has always been a challenging task (Uschold, *et al.*, 1998; Fernández-López *et al.*, 2019). Guidelines for building ontologies are usually blamed (Zulkarnain, Meziane and Crofts, 2016; Kamdar, Tudorache and Musen, 2017), especially for lack of reuse strategy; some have also argued that these approaches are not explicitly concerned with ontology reuse (Annamalai and Sterling, 2003). Moreover, identification and evaluation of the knowledge sources that can be useful for different application domains (Bontas *et al.,* 2005) are among the other challenging tasks in the reuse process (Jonquet, 2017).

Knowledge engineers not only have to find a set of appropriate ontologies for their search query but should also be able to evaluate those ontologies according to different criteria. Ontology evaluation is at the heart of ontology selection and has received a considerable amount of attention in the literature. Gómez-Pérez (1995) defined the term evaluation as the process of judging different technical aspects of an ontology, namely, its definitions, documentation, and software environment. The term evaluation has also been described as the process of measuring the suitability and the quality of an ontology for a specific goal or in a specific application (Fernández, Cantador, and Castells, 2006). Ontology evaluation can also be concerned with the correctness of an ontology (Gómez-Pérez, 1999).

1.2.1 Evaluation Challenges

The challenge of choosing the right or the best ontology for reuse is what ontologists and knowledge engineers face on a daily basis. Despite the importance of this matter and the widespread research on this topic, there are still many unanswered questions and challenges in ontology evaluation for reuse (Fernández-López *et al.*, 2019). Most of the research work in this field has focused on introducing new sets of metrics, frameworks, and systems; however, no previous studies have explored the users' views, and the criteria they tend to look at while evaluating ontologies. Moreover, while knowledge engineers examine different aspects of ontologies when assessing their suitability for reuse, the main focus of most of the evaluation frameworks introduced in the literature is on a limited set of metrics.

Furthermore, one of the main aims of developing ontologies has been and still is, to use them as a shared conceptualization between different groups of people working in the same domain (Gruber, 1993; Fernández-López *et al.*, 2019). Mcdaniel, Storey and Sugumaran (2016) argued that ontologies should not only be developed by communities that share a common interest but should also be endorsed and evaluated by those communities. However, most of the studies in the field of ontology evaluation are based on internal characteristics of ontologies and have failed to address how social interactions and community recognition affect the quality and reusability of an ontology (Mcdaniel, Storey and Sugumaran, 2016).

1.3 Research Questions

To help formulate the research aim and objectives, the central question that this research asks is how ontologies are evaluated and selected for reuse. The set of metrics that affect this process and whether or not the social and community related metrics can be used to assess the suitability of an ontology for reuse are the areas of particular interest for this research. The following sub-questions help frame the research:

RQ1: How are ontologies evaluated and what are the main quality metrics used in the ontology evaluation and selection process?

RQ2: Do the social interactions among the community of ontologists and knowledge engineers affect how ontologies are evaluated and selected for reuse?

RQ3: How important are social and community aspects of ontologies compared to the well-known internal criteria, such as content and structure?

RQ4: Does the domain, type of organisation and years of experience of the ontology evaluator affect how the quality of an ontology is judged?

1.4 Aim of Research

This study aims to investigate the notion of ontology reuse and to clarify the process of ontology evaluation and selection. It also intends to determine the metrics that can be used to evaluate the suitability of an ontology for reuse and their importance. The core objectives of this research include:

Objective 1: Conduct an extensive critical survey of ontology evaluation techniques and systems. Since 1995 (Gómez-Pérez, 1995), researchers have tried to propose different tools and techniques to facilitate the process of ontology reuse. However, ontology selection for reuse is still known to be a complicated, very time consuming and a manual task (Butt, 2017). Therefore, an extensive critical survey of the available tools and techniques is required to identify both the shortcomings and advantages of current approaches. A review of 36 different tools, algorithm and techniques used in the ontology selection domain is presented in Appendix A-F.

Objective 2: Study the notion of quality in the ontology domain, determine how ontologies are evaluated and selected, and identify and classify the set of metrics that are used in that process. Despite the availability of different tools, techniques, and frameworks, there is very little scientific understanding of how ontologists and knowledge engineers approach the issue of ontology evaluation and selection for reuse. Therefore, there is a need for a study that investigates the selection process from the users' point of view by asking them what they look for while evaluating an ontology for reuse.

Objective 3: Determine whether social and community interactions can affect the reusability of ontologies. Most of the current selection tools and techniques tend to evaluate ontologies based on their internal characteristics and dismiss how the interactions in the community may affect the selection process. Popularity (Martínez-Romero *et al.*, 2017) is among the very few social-related metrics used in the literature to assess ontology quality. This research uses interviews and surveys to investigate if the quality and reusability of an ontology can be assessed using any other social-related metric.

Objective 4: Determine if the choice of metrics used in the evaluation and selection process can be linked to the domain, organisation type and years of experience of ontology users. According to the literature, using ontologies is more popular in some domains than the others. One of the objectives of this study, therefore, is to collect survey data and use them to determine if ontologists and knowledge engineers working in different domains and organisations tend to evaluate ontologies differently.

Objective 5: Construct and test a framework to facilitate the process of ontology selection and recommendation for reuse. Identifying a set of evaluation metrics will lead to a proposal of a new framework for ontology evaluation, selection and recommendation for reuse. The usefulness of the new metrics and framework will then be validated by asking different experts in the domain.

Objective 6: Provide recommendations to the ontology engineering community. The findings of this study will help in proposing a set of steps that can be taken by the community ontologists and knowledge engineers while building, evaluating and selecting ontologies for reuse.

1.5 Research Methodology

A collection of both qualitative and quantitative research designs will be followed to provide a deeper understanding of how ontologists and knowledge engineers evaluate and select ontologies for reuse. Interviews will be used in the first phase of this study to explore the general process of ontology evaluation and selection and the factors that can affect this process. A questionnaire will then be used to clarify, confirm, and generalize the findings of the first phase. Finally, an additional set of interviews with experts in the ontology domain will be conducted to validate the findings of the previous phases and the framework proposed by this study.

1.6 Contributions to Knowledge

This study will make several contributions to the current literature. Firstly, it will clarify the set of criteria that ontologists and knowledge engineers tend to assess while evaluating the quality and reusability of ontologies. Moreover, it will explore the role and importance of metadata and social interactions among the community of ontologists and knowledge engineers in the process of ontology evaluation and selection for reuse. This study also provides an exciting opportunity to advance the understanding of one of the most used, and maybe the only used, social-related metric for ontology evaluation, that is the popularity of an ontology.

1.7 Thesis Outline

This thesis is composed of eight chapters. The second chapter provides a thorough overview of the history of different aspects of ontology evaluation and selection for reuse. The third chapter deals with the methodology used for the research in this thesis. Chapters 4 and 5 will present the findings and analysis of the first two phases of the data collection for this study. The focus of chapter 6 is on validating the findings of the previous phases by reporting the findings of the third phase of this research. The overall findings will then be discussed in chapter 7. Finally, a brief summary is provided in chapter 8, which is the conclusion chapter.

A number of appendices are also included, in particular, a review of different selection systems in the literature (Appendix A-Appendix F). Questions from the different phases of the research are presented in Appendix G-Appendix I. Complete interview transcriptions and survey data are not included in this thesis, as they would add more than 100 pages to the appendices; however, they are available on request.

Chapter 2: LITERATURE REVIEW

2.1 Introduction

Managing and sharing heterogeneous information resources are among the most serious challenges of the current era (Martínez-Romero *et al.*, 2017; Fernández-López *et al.*, 2019; Kolbe *et al.*, 2019). Ontologies, which are commonly defined as a formal specification of a shared conceptualization (Guarino, Oberle and Staab, 2009), are one of the proposed solutions for the problem of information management and overload (Sharp, 2017). With the advancement of technology and knowledge, the number of available ontologies has been growing dramatically (Matentzoglu *et al.*, 2018), and as a result, the process of searching and finding a suitable ontology for reuse has become a challenging task.

Different search and selection systems such as Swoogle[1], Linked Open Vocabularies (LOV)[2], Ontology Lookup Service (OLS)[3] and NCBO BioPortal[4] have been proposed to tackle the problem of searching and finding a reusable ontology. However, despite all the advancements, ontology search, evaluation, and selection are still among some of the very challenging tasks in the ontology domain (Martínez-Romero *et al.*, 2017; Fernández-López *et al.*, 2019; Kolbe *et al.*, 2019). Search and selection systems for ontologies have usually been blamed for lack of automation (Wang, Guo and Fang, 2008), for being mostly keyword-based (Martínez-Romero *et al.*, 2017) and even for dismissing the quality of ontologies in the selection process (Tan and Lambrix, 2009).

This chapter provides a review of different aspects of ontology evaluation and selection for reuse. It starts by providing some definitions for ontology and ontology reuse and then discusses why reusing ontologies matter. It also identifies different search and selections approaches that have been used in the literature to facilitate the process of finding an ontology for reuse. Ontology evaluation and what it depends on is also discussed in this chapter.

2.1.1 Ontology

Ontology is one of the widely used terms in science. According to philosophy, ontology is the study of the kinds of things that exist (Chandrasekaran, Josephson and Benjamins, 1999) and

---

[1] http://swoogle.umbc.edu/2006/
[2] https://lov.linkeddata.es/dataset/lov/
[3] https://www.ebi.ac.uk/ols/index
[4] https://bioportal.bioontology.org/

it tries to answer some questions about the nature of being. The term “ontology” is also widely used in computer science and information science. The most popular definition for ontology, and the one used in the context of this research, comes from one of the very early papers in this field by Gruber (1993) in which ontology was defined as “an explicit specification of a conceptualization”. Ontology has also been defined as the shared understanding or knowledge of a domain that is agreed between some agents (Uschold, 1998).

Numerous other definitions have been proposed over the years in various research communities. In Artificial Intelligence, for example, Chandrasekaran, Josephson and Benjamins (1999) defined ontology as a representation vocabulary that aims to capture conceptualizations. Similarly, and in the information science field, the ontology was described as an explicit axiomatic theory that is designed and used to depict the structure of a specific domain (Zuniga, 2001).

Different types of ontologies have been identified in the literature, and a number of approaches have been proposed to categorize them. One of the earliest classifications was proposed by Mizoguchi, Vanwelkenhuysen and Ikeda (1995), and categorised ontologies into four major types: content ontology, tell & ask ontology, indexing ontology, and meta-ontology. Visser and Bench-Capon (1998) also used the concept of ontological commitments to classify ontologies into three main groups: task ontology, method ontology, and domain ontology.

One of the most common ways of classifying ontologies is to group them into the domain and purposive ontologies. Domain ontologies have been defined as a reusable entity that generically captures an area of a domain, including all the concepts in a domain and also the relationships that exist among them (Annamalai and Sterling, 2003; Ramesh and Iyer, 2016). Some argue that despite the usefulness of these types of ontologies, they are loosely coupled to one another and sometimes only cover the very general upper-level concepts in a domain (Annamalai and Sterling, 2003). “Top-level” or “upper-level” are the two other terms used to refer to the ontologies that describe the general domain and independent concepts (Schmidt *et al.*, 2018). Purposive ontologies, in contrast, are usually built by using different domain ontologies and define terms for supporting a specific purpose or use (Annamalai and Sterling, 2003). This research covers both domain and purposive ontologies.

2.1.2 Ontology Benefits

Research on ontologies is important, as they play a significant role in the semantic web and are used for a broad variety of different purposes. According to the survey conducted by Cardoso in 2007, ontologies are mostly used to share a common understanding of the structure of information among people or software agents (69.9%) in different domains like education, and computer software. Reusing and analyzing domain knowledge and making domain assumptions explicit were mentioned as the other reasons for using ontologies (Cardoso, 2007).

Ontologies have many benefits no matter in which domain they are used. They facilitate communication and knowledge transfer between systems, between humans, and between humans and systems (Bürger and Simperl, 2008; Feilmayr and Wöß, 2016) by uniquely identifying the meaning of different concepts in any domain. Moreover, the costs of new developments of knowledge models can be avoided by reusing the already existing ontologies (Subhashini and Akilandeswari, 2011) and also enriching and updating ontologies with additional knowledge (Cui, Damiani and Leida, 2007). Furthermore, ontologies can be used to find logical inconsistencies in knowledge or enhance the information retrieval process (Bürger and Simperl, 2008).

2.1.3 Ontology Development

Despite the significant role that ontologies play in the semantic web, there is still no clear understanding of the way they should be developed and finding the best development method has always been a challenging task (Frolov *et al.*, 2009). Some have even considered ontology development as an art rather than a science (Jones, Bench-Capon and Visser, 1998; Frolov *et al.*, 2009). Different approaches and design patterns have been proposed in the literature to address this issue and to support the ontology development process. Despite all the similarities, these approaches are different from each other in a number of respects, namely, how they suggest an ontology should be built and also the importance of reuse in the development process.

One of the very first methods for ontology development was proposed by Uschold and King (1995). According to this approach, which was based on the experiences gained while developing an enterprise ontology, a comprehensive methodology for ontology development included a set of steps, namely, purpose identification, ontology development, ontology evaluation and providing documentation. Each of those steps also included a set of techniques, principles, and guidelines. For instance, the ontology development step contained three sub-steps: ontology capture, ontology coding and integrating existing ontologies.

One of the very well cited methodologies for ontology development was proposed by Noy and McGuinness (2001). They argued that there is no correct way to model a domain and the best solution might vary according to the application; however, they suggested that the process of developing an ontology should start with determining domain and scope of an ontology and also considering ontology reuse. The next step according to this approach was to define and enumerate the important terms and classes of the ontology.

De Nicola, Missikoff and Navigli (2009) tried to apply software engineering approaches to ontology development by proposing a methodology called UPON (Unified Process for Ontologies). This methodology took advantage of the Unified process and UML and introduced a novel approach for making large-scale ontology. This approach included different steps: determining the domain of interest and the scope, defining the business purpose (or motivating scenario, with users and their objectives), writing one or more storyboards, identifying the competency questions and use-case identification and prioritization.

Unified Process for ONtology building (UPON) (De Nicola and Missikoff, 2016) lite is one of the recent methodologies for ontology development; it consists of six sequential steps: domain terminology, domain glossary, taxonomy, prediction, parthood (meronymy) and ontology. What made this approach different is that it minimised the role of ontology engineers in the ontology development process by putting users and domain experts with no prior ontology experience in charge of the development process (De Nicola and Missikoff, 2016).

As was seen in the above-mentioned approaches, the process of developing ontologies usually starts by eliciting the requirements and includes ontology development and evaluation. Despite the availability of different methodologies for ontology development, some surveys have indicated that using methodologies in the ontology development process is not that common (Cardoso, 2007) and ontologies are either defined initially or emerge from experiences obtained during ontology development (Iqbal *et al.*, 2013). Moreover, ontology development is known to be a very time consuming (ibid.) and also an error-prone process (Mace, Parkin and van Moorsel, 2010).

This study seeks to obtain data which will help to address some of the challenges faced in the process of ontology development. As will be discussed in the next section, reusing the available ontologies can be considered as one of the solutions to the mentioned problems.

2.2 Ontology Reuse

The recent developments in Semantic Web technology has urged researchers and ontology engineers to develop ontologies in different domains, from biomedical (Jonquet, 2017) to education (Ameen, Khan and Rani, 2012) and tourist (Park, Yoon and Kwon, 2012). An increase in the number of ontologies and the cost of developing them has urged researchers to consider ontology reuse (Bontas, Mochol and Tolksdorf, 2005). Ontology reuse has been defined as the process of using the available ontological knowledge as input to develop new ontologies (Bontas, Mochol and Tolksdorf, 2005; Caldarola, Picariello and Rinaldi, 2015). Building ontologies by reusing the available ones will not only facilitate the development process, but will also help to build the ontologies that have higher quality, and are reusable (Butt, Haller and Xie, 2016; Caldarola and Rinaldi, 2016).

Ontology reuse is beneficial to the community of ontology engineers. It will not only help in achieving one of the main goals of building ontologies, that is to share and reuse them (Simperl, 2009; Fernández-López *et al.*, 2019), but will also save a significant amount of time and financial resources. While reusing ontologies might be an option in some cases, construction of some types of ontologies, such as domain ontologies, without reusing the existing ones will be very costly (Fernández, Cantador and Castells, 2006), time consuming and will also need a great effort (Bontas, Mochol and Tolksdorf, 2005; Xiang *et al.*, 2010).

Merging and integration are two of the well-known approaches for ontology reuse (Caldarola, Picariello and Rinaldi, 2015). Merging refers to unifying two or more different ontologies in one subject domain whereas integration refers to the process of building an ontology in one subject area by reusing one or more ontologies in different subjects (Pinto and Martins, 2001; Caldarola and Rinaldi, 2016). The terms hard reuse and soft reuse have also been proposed to describe two different approaches for ontology reuse, namely, importing an ontology or referring to its element URLs, respectively (Poveda-Villalón, Suárez-Figueroa and Gómez-Pérez, 2012b; Fernández-López *et al.*, 2019).

Reuse has always been a major area of interest in the ontology domain; however, the best way to reuse an ontology is not clear yet. <owl:imports> in Protégé (Musen and Protégé Team, 2015) can be considered as one of the most popular and also easy ways of reusing ontologies. This OWL statement facilitates the process of reuse by allowing ontologies to import other ontologies, without having to import the ontology content into their file (Ong *et al.*, 2016). Despite being very easy and also reasonable for small ontologies, <owl:import> has been blamed for being both unnecessary and impractical for many cases of ontology development (Xiang *et al.*, 2010); it has also been criticized for not supporting partial ontology reuse (Pan, Serafini and Zhao, 2006).

Different methodologies for ontology development have also emphasized the importance of re-using previously built components while developing a new one. In 2013, a review of ontology engineering methodologies conducted by Iqbal et al. (2013) highlighted that many of the methodologies advocate ontology reuse (10 out of 14) and have proposed a reuse step in their guideline. Despite mentioning the importance of the reuse step in different development approaches, they have been usually blamed for lack of reuse strategies and some have even argued that methodologies for ontology development are not explicitly concerned with ontology reuse (Annamalai and Sterling, 2003).

As a part of this study, different steps in the process of ontology reuse were investigated and some of the challenges faced in that process were addressed.

2.2.1 Reuse Challenges

Regardless of all the advantages of reusing ontologies and the availability of different ontologies, ontology reuse has always been a challenging task (Uschold, Healy, *et al.*, 1998; Fernández-López *et al.*, 2019). There are many reasons why ontology reuse is challenging. Firstly, the type of ontologies and how and why they have been built can affect the reuse process. Some have stated that reusable ontologies must be developed independently from the application and context. Annamalai and Sterling (2003), however, argued that this kind of thinking might create fundamental problems, namely, developing overgeneralized ontologies that lack some useful knowledge or having ontologies that are scattered and have a sparse construct. Some have also argued that the general domain and independent ontologies are either not reusable (Bontas, Mochol and Tolksdorf, 2005) or might need a considerable amount of modification and extension before being utilized. However, an interesting investigation conducted by Ochs *et al.* (2017) indicated that ontologies might get reused, even if they have not necessarily been developed to be reused.

The first steps of ontology reuse, that is the identification and selection of the knowledge sources which can be useful for an application domain (Bontas, Mochol and Tolksdorf, 2005; Lewen and d'Aquin, 2010), has also been mentioned as the hardest step in the process of ontology reuse. The lack of appropriate supportive tools and automatic measurement techniques for evaluating and assessing different ontology features can be considered as another barrier for ontology reuse (Fernández, Cantador and Castells, 2006). Moreover, the characteristics of a reusable ontology are not known, and there has been little discussion about the notion of ontology reuse and the factors it depends on.

As was discussed above and despite the very important role of reuse in the ontology domain, there is still no consensus among the community of ontologists and knowledge engineers about how ontologies should be reused or what the characteristics of a reusable ontology are. Moreover, the first step of reuse, which is ontology identification and selection, is considered to be a very complicated task. Different search and selection systems, e.g., Swoogle, NCBO BioPortal Recommender[5], and OLS, have been proposed in the literature to address these challenges. The rest of this chapter continues to discuss ontology selection in more detail, as it is one of the very important factors affecting ontology reuse.

2.3 Ontology Selection

Selection is a fundamental property of the ontology domain. Ontology selection is about finding and choosing the "most suitable" (Suárez-Figueroa and Gómez-Pérez, 2008) or "the best" (Noy *et al.*, 2013) ontologies that are relevant to queries submitted by users (Alani *et al.*, 2007). It has also been defined as "the process that allows identifying one or more ontologies or ontology modules that satisfy certain criteria" (Sabou, Lopez and Motta, 2006). As it is seen in the literature, ontology selection is usually linked to the other processes in the ontology domain, such as ontology evaluation and ontology ranking (Alani *et al.*, 2007). Sabou *et al.*, (2006), for example, argued that ontology selection is "in essence, an ontology evaluation task". Alani *et al.* (2007) also argued that ontology ranking, and ontology selection are "two complementary sides of the problem of finding relevant ontologies".

[5] https://bioportal.bioontology.org/recommender

Ontology selection has been a major area of interest since 1995 (Gómez-Pérez, 1995). There are many different reasons why ontology selection is so important, namely, because the rapid development of the semantic web has led to the great number of available ontologies (Sabou, Lopez and Motta, 2006) and because searching for and identifying semantic web resources is in demand (Martínez-Romero *et al.*, 2017). Moreover, ontology reuse, as one of the very important aspects of ontology engineering, is highly dependent on ontology search and selection (Sabou, Lopez and Motta, 2006; Alani *et al.*, 2007; Cheng, Ge and Qu, 2008; Sridevi and Umarani, 2013). Ontology selection is especially very critical in cases like data annotation, where failing to find an appropriate ontology might lead to reprocessing, re-annotating data or redesigning an application (Jonquet, Musen and Shah, 2010).

Different systems and approaches, namely, selection algorithms (see Appendix A), ontology libraries (see Appendix B), search engines (see Appendix C), evaluation systems (see Appendix D), ranking systems (see Appendix E) and recommender systems (see Appendix F) have been proposed in the literature to support and facilitate the process of ontology selection for reuse. These systems are different from each other in a number of respects, namely, their aims and types and also the selection scenario that they support. These differences are discussed in more detail in sections 2.3.1 and 2.3.2.

2.3.1 Types of Selection Systems

Ontology selection for reuse is supported by different types of systems. The following part of this chapter aims to review some of the most popular types of ontology selection systems. Each of these systems might follow different scenarios, meaning that they get different types of input and conduct different types of evaluation and ranking to come up with a set or combination of ontologies that they believe are the most suitable for users' requirements. A systematic review of different types of systems has also been conducted and is presented in Appendix A-F.

2.3.1.1 Ontology Libraries

Different terms, namely, ontology archive, directory, repository, portal, registries, and ontology library are used in the literature to refer to the group of systems that collect, manage, publish and provide access to ontologies from different resources (Naskar, 2014). Ontology library is the most popular and generic term among the ones mentioned above and is defined as a web-based system that aims to provide access to an extendable collection of ontologies (d'Aquin and Noy, 2012). Ontology library has also been described as a system that supports and fulfills the needs of ontology reuse and selection by storing and maintaining ontologies (Ding and Fensel, 2001; Jonquet, 2017).

As it is seen in the literature, the broad use of the terms “ontology library” is sometimes equated with the use of the term “ontology repository”. Jonquet (2017), for example, argued that ontology libraries are mostly concerned with a listing of ontologies and that ontology search, ontology browsing, ontology mapping, ontology visualization, and metadata management are the kind of services that ontology repositories offer. In this thesis, the term ontology library will be used in its broadest sense to refer to the systems that provide all the discussed functionalities.

Browsing is the minimum function that ontology libraries should provide and includes a quick process of filtering and narrowing down a collection of ontologies according to their different characteristics (d’Aquin and Noy, 2012). The browsing process is not based on an input, and ontologies are usually classified, filtered and shown according to different tags and metadata assigned to them, such as ontology subject, ontology type, ontology publisher, and ontology version (ibid.). Browsing is very helpful, especially in cases where the selection requirements are not clear, e.g., when users are not sure what exactly they are looking for and they prefer to explore and navigate ontologies or to see a list of available options, grouped by different criteria (Naskar, 2014). Ontology management, adaptation, standardization (Ding and Fensel, 2001) and programmatic access to ontologies are amongst some of the other functionalities provided by ontology libraries.

With the growing number of developed ontologies, ontology libraries and repositories have been a major area of interest in the semantic web community. Ontolingua (Farquhar, Fikes and Rice, 1997), WebOnto (Domingue, 1998) and SHOE (Heflin and Hendler, 2000) were among some of the very early systems that provided browsing facility by indexing ontologies by the name of their classes or alphabets (Ding and Fensel, 2001). NCBO BioPortal, OLS, OBO Foundry[6], and ONKI ontology server[7] are some of the most recent and also well-known ontology libraries; they support ontology reuse by offering different functionalities.

A review of some of the ontology libraries is provided in Appendix B. Ontology libraries are of particular interest of this research, due to their importance in the ontology selection domain

[6] http://www.obofoundry.org/
[7] https://onki.fi/

and the set of metrics they use in the ontology evaluation process. Feature variation in these libraries is highly driven by their system type, their purpose, their scope, the users' requirements they support and the maturity of their software (d'Aquin and Noy, 2012). BioPortal, OLS and OBO Foundry, as some of the most well-known examples of ontology libraries, are all developed for biomedical ontologies whereas ONKI is focused on business and geography related ontologies.

2.3.1.2 Search Systems for Ontologies

The considerable amount of increase in the number of ontologies and essentiality of ontology reuse has made ontology search one of the major areas of interest in the ontology domain (Sridevi and Umarani, 2013). Search engines for ontologies aim to facilitate the process of ontology exploration and retrieval (Naskar, 2014) by finding an ontology, a module in an ontology (Suárez-Figueroa and Gómez-Pérez, 2008) or a set of ontologies (Martínez-Romero *et al.*, 2017) that are most relevant to users' queries (Alani *et al.*, 2007). Search engines are important for different reasons, namely, because the evolution of semantic web has led to increase in the number of available ontologies and semantic documents on the web (Buitelaar and Eigner, 2007; Franco *et al.*, 2019) and also because the costs of building ontologies, especially by reusing the available ones, is highly dependent on the ontology search and identification (Arpírez *et al.*, 1998; Alani, Brewster and Shadbolt, 2006).

The process of finding the best or the most suitable ontology on the web consists of different steps, starting from the search requirement specification. Search engines get different keyword(s) as input and search their repository of ontologies to find the ones that best match the users' queries (Alani *et al.*, 2007). The output of these systems is a list of ontologies that are ranked according to their relevance to the input query. Search engines for ontologies cannot function in isolation and use different matching techniques, ranking algorithms and evaluation frameworks (ibid.) to find the most relevant match(es) for users' queries (Sridevi and Umarani, 2013).

A large number of search engines for ontologies have been developed since the early 2000s. Swoogle, as the first, and one of the most known search engines for ontologies was initially introduced in 2004. Swoogle provides keyword(s) based search over the semantic web documents in its repository. Watson (d'Aquin and Motta, 2011), FalconS (Cheng, Ge and Qu, 2008) and Sindice (Tummarello, Delbru and Oren, 2007) are among the other search engines for ontologies.

Despite all the development and the significant increase in the number of ontologies, facilities and functionalities offered by some of the search engines for ontologies are still very similar, if not the same, to what they used to offer in the early days. Moreover and as it is seen in the review conducted by Kolbe *et al.* (2019), the number of search engines developed for ontologies has decreased in recent years and many of them are not available anymore. This might be an indication that the community has lost interest in the search engines or does not find them sufficient for the needs of ontologists and knowledge engineers (Cantador, Fernandez and Castells, 2007).

2.3.1.3 Evaluation Systems for Ontologies

Ontology evaluation, as a feature, is supported by most of the selection systems in the ontology domain; however, some of the systems proposed in this domain are solely concerned with ontology evaluation, and therefore, are called ontology evaluation systems. Broadly speaking, tools, techniques, frameworks, and algorithms for ontology evaluation can be categorised into two main groups: the ones that are concerned with the correctness of an ontology and the ones that focus on the quality and suitability of an ontology, especially for reuse purposes.

Ontology Pitfall Scanner (OOPS)[8] is one of the very well-known, and also available, examples of the first group. As a fault detection tool, OOPS is able to analyse ontologies and identify the ontology elements that are affected by pitfalls (Poveda-Villalón, Suárez-Figueroa and Gómez-Pérez, 2012a). ODEClean (Fernández-López and Gómez-Pérez, 2002), a plugin for WebODE, is another example of ontology validation system which allows its users to assign meta properties to concepts in taxonomies, as well as checking errors automatically. ODEval is also used to evaluate concept taxonomies and is able to detect different flaws in ontologies, namely, redundancy problems and partition errors (Corcho *et al.*, 2004).

The main focus of this study is on ontology selection for reuse and the evaluation approaches that are concerned with the quality and suitability of an ontology, not those that focus on ontology correctness. Some of the very well-known evaluation algorithms and systems in the literature were chosen for a systematic review, which is presented in Appendix D.

2.3.1.4 Ranking Systems for Ontologies

Similar to ontology evaluation, ranking is used as one of the fundamental features of different selection systems in the ontology domain; it has, therefore, been investigated in this study. As

[8] http://oops.linkeddata.es/

it is seen in the literature, some of the systems proposed and implemented in the ontology domain, namely, AKTiveRank (Alani, Brewster and Shadbolt, 2006) and OntoQA (Tartir and Arpinar, 2007) are solely concerned with ontology ranking. These systems usually do not work as a stand-alone system and use search engines like Swoogle to retrieve a set of ontologies. They then re-rank them using their own set of criteria and show them in the output (Alani, Brewster and Shadbolt, 2006). A review of this type of system is provided in Appendix E.

2.3.1.5 Recommendation Systems for Ontologies

Recommender systems have been defined as "software tools and techniques providing suggestions for items to be of use to a user" (Ricci *et al.*, 2010). They are widely adopted in a variety of domains and provide guidance and help for the individuals who lack sufficient personal information or experiences in evaluating the overwhelming number of items available on the web (Ricci, Rokach and Shapira, 2011). In the ontology domain, however, there are only a few investigations about how recommendation algorithms, especially the collaborative filtering ones (Ekstrand, Riedl and Konstan, 2011), can affect the process of ontology selection for reuse.

Cantador, Fernández and Castells (2006) argued that humans tend to make decisions not only based on their own previous experiences, but also by using the experience of the people in their social circle. They blamed the evaluation and selection approaches of their time for neglecting the role of users' collaboration in the evaluation process and stated that some aspects of ontologies can only be evaluated by human's judgement. Therefore, search engines for ontologies were not sufficient (Cantador, Fernandez and Castells, 2007). To address this shortcoming, they proposed Collaborative Ontology Reuse and Evaluation (CORE) and WebCORE, which was an extension of CORE. These two approaches were based on collaborative filtering algorithms and social interactions (Cantador, Fernandez and Castells, 2007). Despite its novelty and advantages, no selection system based on this approach is currently available.

Recently, recommender systems have been gaining attention in the ontology domain and for selection purposes. They have been mostly used in the biomedical field and for finding the most appropriate ontologies for annotation. Biomedical Ontology Selection System (BiOSS) (Martínez-Romero *et al.*, 2014), NCBO Ontology Recommender and AgroPortal

Recommender[9] are some of the most prominent recommender systems for ontologies. Apart from WebCORE, recommender systems in the ontology domain do not apply the recommendation algorithms to the ontologies in their repository. There are very similar to the search engines in this field, as they both retrieve a set of candidate ontologies that are relevant to a search query.

Recommender systems for ontologies are of particular interest of this research, due to their significant role in the selection process. What makes them different is their evaluation component and the fact that they assess all the candidate ontologies according to different predefined evaluation criteria (Martínez-Romero *et al.*, 2017). In other words, recommender systems do not only show a ranked list of ontologies in their output but also assign a score to each of those ontologies to indicate their appropriateness for the input query (Jonquet, Shah and Musen, 2009). Martínez-Romero *et al.* (2017) have argued that evaluation is a fundamental part of any recommendation task and has stated that the main aim of recommender systems for ontologies is to facilitate the problem of selecting an ontology by adding the evaluation element to this process.

2.3.2 Ontology Selection Scenarios

Ontology selection is based on different scenarios that determine the information need, selection criteria and selection output. Sabou *et al.* (2006) have argued that the process of ontology selection mainly consists of four different elements, namely, selection requirements (e.g., query or input), selection criteria, ontology library, and selection results (output). A more detailed account of these elements is discussed in the following.

2.3.2.1 System Repository/Library

Libraries and repositories are important, because the process of ontology selection consists of identifying an ontology or a set of ontologies, that best matches a query, from a collection of ontologies. Different methods are used to collect ontologies and to form libraries or repositories for ontology selection systems. Swoogle, for example, has a discovery component that is responsible for crawling the web and collecting what its developers call semantic web documents (e.g., files with RDF, n3, owl and daml extensions) (Finin *et al.*, 2004). Allowing users to submit URLs of ontologies is another common way of collecting them; this method is used by different selection systems especially ontology repositories, such as OBO Foundry

[9] http://agroportal.lirmm.fr/recommender

(Smith *et al.*, 2007). Some of the systems proposed in the ontology domain, namely, ranking and evaluation systems, only aim to rank a set of previously retrieved ontologies (Subhashini *et al.*, 2011); hence, they do not have a repository. OntoQA, as a sample of this type of system, does not collect the ontologies itself and use the ones in Swoogle to find ontologies (Tartir and Arpinar, 2007).

2.3.2.2 Selection Input (scenario)

One of the main steps in ontology selection is to identify and clarify a set of requirements that the potential ontology should satisfy in order to be selected (Sabou, Lopez and Motta, 2006). According to the literature, there are many different ways of expressing the selection requirements. What follows are the various types of selection input:

- **Keyword(s)-Based.** Having keyword(s) as input is one of the simplest and also the most popular method in the ontology selection process and is supported by most of the tools and frameworks in the ontology domain, for example, Swoogle and NCBO BioPortal. This scenario is particularly popular in the ontology search process. Keyword(s) can refer to the domain of an ontology or a specific concept in it (Alani *et al.*, 2007). Despite its popularity, the keyword(s)-based ontology selection has been blamed for being insufficient, especially for determining the context of the search (Patel *et al.*, 2003).

- **Corpus-Based.** This approach allows users to enter a text corpus, e.g., a part of a scientific paper, as input. This method is mostly used by the ontology recommendation systems like BioPortal Recommender, where the main aim is to find and recommend the most appropriate ontologies for annotating a text corpus.

- **Metadata-Based scenario.** Some of the selection systems allow their users to express their selection requirements by using different types of metadata about ontologies, such as, their format, both the natural and the ontological language that have used in their development process, and the type of the ontology. This scenario is particularly common in ontology browsers, such as AgroPortal browser[10], where different sets of metadata are extracted from each of the ontologies in their repository.

[10] http://agroportal.lirmm.fr/ontologies

- **Query-Based.** Some of the selection systems like AberOwl[11] provide their users with reasoning facilities and allow them to query a set of ontologies using semantic query syntax (Hoehndorf *et al.*, 2015).

It has been argued that the choice of input depends on the selection requirement (Alani *et al.*, 2007). However, users tend to prefer basic keyword(s)-based interfaces over the more complex ones, like the query or semantic-based search (Patel *et al.*, 2003). The choice of input can affect the output quality in different ways. Firstly, the keyword(s) used in search processes usually refer to specific topic or domain. Alani *et al.* (2007) argued that being relevant to a topic does not mean or guarantee that the topic name is mentioned in the name of different classes and properties of an ontology. Moreover, context matters in the search process and some have argued that context-aware search will lead to more relevant and useful results (Cao *et al.*, 2009). However, using keyword(s) as input for ontology search and selection will not help in determining the right query context. Therefore, different pre-processing and query expansion techniques have been proposed in the literature to improve the quality of input for ontology search.

2.3.2.3 Pre-processing

In order to provide the best results, selection systems perform two different types of pre-processing: one on their collection of ontologies and the other one on the input provided by users. Before adding a new ontology to their collection of ontologies, selection systems usually extract different types of metadata and information about it, which can later be used in the selection process (Ding *et al.*, 2005; Côté *et al.*, 2006; Pan, Thomas and Sleeman, 2006).

Besides that, and to provide more relevant results, selection systems perform different types of pre-processing techniques on users' input queries, namely, query expansion and disambiguation. One way to expand a query is to treat it as a domain name and use relevant web pages and Wikipedia[12] to expand it; this can help in collecting a set of terms that best represent a domain (Alani *et al.*, 2007). WordNet (Miller, 1995) has also been widely used to clarify the context and different senses of a query, e.g., synonyms, hypernyms or hyponyms (Patel *et al.*, 2003).

[11] http://aber-owl.net/#/
[12] https://www.wikipedia.org/

Despite the importance of the pre-processing techniques, there is no clear evidence that indicate which one of them is more useful in improving the ontology selection process.

2.3.2.4 Ontology Matching and Identification

Having a collection of ontologies and the users' input, the next step is to make a match between them and to find the ontologies in the repository that are the most relevant to users' selection requirements. There are many ways of doing that, but the most popular method is to determine how well an ontology is covering those terms (Jonquet, Musen and Shah, 2010). To do that, names and labels of ontologies' classes and properties are analyzed to determine the number of times each query term has appeared in those ontologies (Alani *et al.*, 2007). According to Sabou, Lopez and Motta (2006), ontology concepts can match the query terms syntactically, meaning that they are lexically similar or semantically, meaning they are more general, specific or even equivalent to each other.

2.3.2.5 Ontology Evaluation

After identifying a set of ontologies that cover or include the users' query term(s) to some extent, the next step is to measure their adequacy. This is one of the most essential tasks in the ontology selection process, also known as ontology evaluation. Ontology evaluation has been widely defined as the process of determining the quality of an ontology for being used for a specific goal and in a specific context (Fernández, Cantador and Castells, 2006). Ontologies can be evaluated using different sets of criteria, namely, connectedness (Patel *et al.*, 2003) and popularity(Martínez-Romero *et al.*, 2014). In this step, a score will be assigned to different features of ontologies.

2.3.2.6 Ontology Ranking

This step aims to arrange, and rank ontologies based on their evaluation score. Ranking algorithms for ontologies play a crucial role in the selection process by sorting the potentially relevant ontologies based on how well they cover a selection scenario. In the literature, the phrases "ontology evaluation" and "ontology ranking" are sometimes used interchangeably, as they both share a number of key characteristics. Throughout this thesis, however, ontology ranking will refer to the process of sorting ontologies in descending order, and according to their evaluation score.

2.3.2.7 Selection Output

Depending on the selection requirements, criteria and scenario, selection systems for ontologies can offer different types of output. The ideal output for a selection task is a single ontology that best suits the query term(s); however, this is not how selection systems work in the real world (Jonquet, Musen and Shah, 2010). As it is seen in the literature and also the available selection systems, the output of a selection task is usually a set of ontologies or ontology components (e.g., classes) that match, cover or are relevant to input requirements or their combination (Martínez-Romero *et al.*, 2017). To facilitate the selection process further, some of the systems provide different types of metadata (Côté *et al.*, 2008) or scores (Jonquet, Musen and Shah, 2010) for each ontology in the output; this additional information helps users to filter and navigate through ontologies and to select the one that best fits their requirements (Côté *et al.*, 2008).

2.3.3 Selection Challenges

Despite all the advancements in the ontology domain and the availability of different selection systems for ontologies, selecting an ontology for reuse is still a challenging task (Jonquet, Shah and Musen, 2009; Martínez-Romero *et al.*, 2017). How selection systems work might be a reason why ontology selection is challenging. As it is seen in the literature, many of the ontology selection systems and approaches are keyword(s) based (Pan, Thomas and Sleeman, 2006). While being useful, it can sometimes get problematic, especially in cases where an ontology is relevant to a query but does not have the query term in the name of its concept (Alani *et al.*, 2007). Moreover, ontology evaluation and selection is usually a manual task, and the level of automation is not usually enough (Wang, Guo and Fang, 2008), even when selection systems are used.

To select the most appropriate ontology for different requirements, applications and domains, users should not only be able to obtain a set of ontologies that contain their query term(s) but should also be able to evaluate those candidate ontologies (Martínez-Romero *et al.*, 2014). Some have argued that ontology evaluation is one of the main issues that should be addressed, if ontologies are to become widely adopted and reused by the community (Brank, Grobelnik and Mladenic, 2005; Gangemi *et al.*, 2006; Obrst *et al.*, 2007; Maiga and Williams, 2009). Moreover, failure to evaluate ontologies or to choose the right ontology can lead to using the ontologies that are not right or have lower quality (Yu, Thom and Tam, 2009).

Despite its importance, ontology evaluation is known to be a very complicated process (Yu, Thom and Tam, 2009; Mcdaniel, Storey and Sugumaran, 2016). Being a fundamental property of the ontology domain, numerous research has tried to study different aspects of it. This study seeks to investigate the process of ontology selection for reuse and identifies the set of metrics that can be used in the evaluation process. The rest of this section will provide an overview of different aspects of ontology evaluation.

2.4 Ontology Evaluation

Ontology evaluation is one of the major concepts in the ontology domain, especially when it comes to ontology selection. It is mainly important when users need to identify which of the several ontologies best meets their application requirement(s) (Martínez-Romero *et al.*, 2017). Ontology evaluation can be done in different stages and for different purposes (Hartmann *et al.*, 2005); the one that is of particular interest of to this study is the type of evaluation that is done to support the decision making in the selection process and by people other than the ontology development team. A more detailed account of ontology evaluation and the factors it depends on is presented in the following.

2.4.1 Ontology Evaluation: Definition

There are many different ways of defining ontology evaluation. One of the most popular and also the earliest definitions for ontology evaluation was provided by Gómez-Pérez (1995) where the term evaluation was used to refer to the technical judgment of an ontology considering its different aspects, namely, its definitions, documentation, and software environment. According to this definition, evaluation encompasses validation and verification; ontology validation is mainly concerned with the correctness of an ontology whereas ontology verification is more about determining how well an ontology corresponds to what it should represent (Gómez-Pérez, 1999). In other words, ontology validation focuses on building the correct ontology, whereas ontology verification is about building an ontology correctly (Hlomani and Stacey, 2014).

Ontology evaluation has also been widely defined as the process of determining the adequacy and quality of an ontology for being used for a specific goal and in a specific context (Fernández, Cantador and Castells, 2006). This definition is used to link the process of ontology evaluation to ontology selection. Ontology selection aims to identify an ontology, an ontology module or a set of ontologies that satisfy a particular set of criteria or selection requirement(s) (Sabou, Lopez and Motta, 2006). Some consider ontology evaluation as the core to ontology selection and argue that ontology evaluation is influenced by different components of the selection process, e.g., selection criteria, type of output, and the libraries that the selection is based on (ibid.). Ontology assessment is the other term used to refer to this particular definition of ontology evaluation; it is commonly defined as the activity of checking and judging an ontology against different user requirement(s), such as usability and usefulness (Suárez-Figueroa and Gómez-Pérez, 2008). Unlike the first definition of the ontology evaluation, in which the developer team is responsible for validating and verifying an ontology, ontology assessment and evaluation for selection are done by the end users (Gómez-Pérez, 1994).

Ontology evaluation can also refer to a function or an activity that aims to map an ontology or a component of an ontology to a score or a number, e.g., in the range of 0 to 1 (Brank, Mladenic and Grobelnik, 2006). These types of processes aim to measure and assess the quality of an ontology with regards to a set of predefined metrics and requirements (Yu, Thom and Tam, 2009). This definition is somehow similar to what Suárez-Figueroa and Gómez-Pérez (2008) defined as ontology quality assurance, which describes the activity of examining every process carried out and every product built during the ontology development and making sure that the level of their quality is satisfactory.

Throughout this thesis, the term 'evaluation' will refer to the process of determining the quality of an ontology for being used for a specific goal and in a specific context, as was defined by Fernández, Cantador and Castells (2006).

2.4.2 Ontology Evaluation: Aims

There are many different reasons why ontologies are evaluated: fault detection, quality assessment, tracking progress in ontology evolution and ontology ranking. Evaluation for correctness and fault detection refers to the approaches that aim to measure the logical and formal correctness of an ontology content (Jonquet, Musen and Shah, 2010; Duque-Ramos *et al.*, 2013). OOPS is one of the most well-known applications for this purpose, and can automatically detect 40 different bad practices or pitfalls that might happen in the ontology development process (Poveda-Villalón, Gómez-Pérez and Suárez-Figueroa, 2014).

Quality assessment is the other main reason for evaluating ontologies. Ontology assessment is a very challenging task, not only because there is no generic quality evaluation solution, but also because determining the right elements of quality for ontology evaluation is difficult (Burton-Jones *et al.*, 2005). Ontologies can also be evaluated to track how they progress and change in the evolution process (Yang, Zhang and Ye, 2006; Yu, Thom and Tam, 2009); this approach helps in tracking different characteristics and changes that are made to different versions of ontologies over time (Arpinar, Giriloganathan and Aleman-Meza, 2006).· In the ONTO-EVO^{A}L approach, for example, pattern modeling was used to ensure that consistency and quality were maintained in the process of ontology evolution (Djedidi and Aufaure, 2010).

Ranking is the other reason for ontology evaluation. In Swoogle, for example, a PageRank (Page *et al.*, 1999) like algorithm is used to calculate the rank of each ontology based on the number of links from and to those ontologies. The main focus of this research is on the quality assessment approaches.

2.4.3 Ontology Evaluation: Importance

Ontology evaluation is important in the ontology development process, whether it is built from scratch, automatically or by reusing other ontologies (Tartir, Arpinar and Sheth, 2010). While building an ontology from scratch, developers need to evaluate the outcome ontology, to measure its quality (Ning and Shihan, 2006), to check if it meets their application requirements (Tartir, Arpinar and Sheth, 2010), and also to identify the potential refinement steps (Brank, Grobelnik and Mladenic, 2005). Evaluation is also helpful in checking the homogeneity and consistency of an ontology, when it is automatically populated from different resources (Arpinar, Giriloganathan and Aleman-Meza, 2006; Tartir, Arpinar and Sheth, 2010).

Building an ontology from scratch is very costly and time-consuming (Maiga and Williams, 2009; Bandeira *et al.*, 2016); therefore, developers are urged to consider reusing existing ontologies while building the new ones (Brewster *et al.*, 2004). However, the number of ontologies on the web has been increasing rapidly (Tartir, Arpinar and Sheth, 2010), and users usually face multiples of them when they need to choose or use one in their everyday activities (Brank, Grobelnik and Mladenic, 2005; Yu, Thom and Tam, 2007, 2009). Hence, knowledge engineers need to assess the quality and correctness of ontologies and also compare them to the other available ones before selecting them for reuse.

Ontology evaluation is known to be the core to the ontology selection process (Gangemi *et al.*, 2006; Sabou, Lopez and Motta, 2006) and is used to select the best or the most appropriate ontology amongst many other candidates in a domain (Brank, Grobelnik and Mladenic, 2005). Therefore, there has been a variety of research on different aspects of ontology evaluation, including methodologies, tools, frameworks, methods, metrics, and measures since 1995. However, much uncertainty and also disagreement still exists about the best way to evaluate an ontology generally or for a specific tool or application (Yu, Thom and Tam, 2009; Mcdaniel, Storey and Sugumaran, 2016).

The remaining parts of this chapter describe different aspects of ontology evaluation, starting from different approaches that can be used to assess the quality of an ontology.

2.4.4 Methodologies for Ontology Evaluation

Different approaches have been proposed in the literature to support the process of ontology evaluation. Guarino and Welty (2002) argued that having a domain-independent method that can support ontological decisions as well as their evaluation is essential for developing what they called a true ontology engineering practice; OntoClean was proposed to address this need. This approach can be used to validate taxonomies for inappropriate and inconsistent modeling choices. Despite having some limitations, namely, being manual, expensive and requiring the efforts of highly experienced ontology engineers (Völker *et al.*, 2008), OntoClean has been very popular and different tools, such as ODEClean (Fernández-López and Gómez-Pérez, 2002) and OntoEdit (Sure, Angele and Staab, 2002) have been developed based on it.

Yu, Thom and Tam (2009) blamed the ontology evaluation tools and approaches of their time for lacking a suitable method that can be used to obtain a mapping between the ontology requirements for different applications and the relevant measures that should be used to evaluate them. To tackle this shortcoming, they proposed a requirements-oriented methodology for evaluating ontologies (ROMEO) and used the Goal-Question-Metric approach (GQM) (Basili, 1993) to help in the process of specifying and constraining the ontology requirements, using a standard template, as well as identifying the role that an ontology will play in the context of an application.

The similarities between software engineering and ontology engineering urged Duque-Ramos *et al.* (2013) to adopt SQuaRE ISO (2005), which is a standard for software product quality, and to develop a framework for ontology evaluation called OQuaRE. OQuaRE helped in the process of identifying the strength and weaknesses of ontologies by providing an objective, standardize framework for ontology evaluation (ibid.). According to this approach, quality can be evaluated by measuring different characteristics of ontologies, namely, their structure, functional adequacy, reliability, performance efficiency, operability, maintainability, compatibility, transferability, and quality in use.

Bandeira *et al.* (2016) blamed methodologies for ontology evaluation for being limited to different sets of criteria instead of providing guidelines for ontology evaluators. To address this problem, they proposed FOCA, which is based on three main steps, namely, Ontology Type Verification, Question Verification, and Quality Verification (ibid.). The FOCA developers claimed that identifying the type of ontology will be helpful in the process of making a decision about the set of questions that should be asked in the evaluation process.

Besides all the different methods for ontology evaluation, ontology evaluation has also been considered as an essential step in many of the ontology development approaches. Uschold and King (1995), as one of the first methods for ontology development, had proposed an evaluation step in the ontology development process. A similar concept was considered in the approach proposed by Fox, Barbuceanu and Gruninger (1996). One of the very first steps in this methodology was to define and clarify different ontology requirements in the form of what they called competency questions; competency questions were then tested in the last step of the development process.

In summary, the above-mentioned approaches facilitate the process of ontology evaluation and selection for reuse. However, and similarly to what was discussed about ontology development methodologies in section 2.1.3, methodologies for ontology evaluation are not that popular in the ontology domain.

2.4.5 Ontology Evaluation Approaches

What follows is a review of some of the most popular ontology evaluation approaches.

2.4.5.1 User-Based Evaluation

Ontologists and knowledge experts can assess the quality of ontologies (Hlomani and Stacey, 2014) in two different ways: one is the criteria-based evaluation approach, in which the suitability of an ontology for a particular task or requirement is evaluated by being compared against a set of predefined criteria (Maiga, 2008). Peer review based evaluation, as the other type of user-based evaluation approach, allows ontologists and knowledge experts to link subjective information to ontologies by providing metadata and additional qualitative information about different aspects of them (Supekar, 2005). Despite their popularity, user-based ontology evaluation approaches are blamed for being solely based on different characteristics of ontologies and for ignoring the functionality of ontologies in applications (Yu, Thom and Tam, 2009).

2.4.5.2 Golden Standard Based Evaluation

This approach refers to the type of evaluation that is performed by comparing an ontology to another ontology, also known as a "gold standard" ontology and aims to find different types of similarities (e.g., lexical or conceptual) between them. This approach was first proposed by Maedche and Staab (2002) and was then used in many other research projects, such as, the one by Brank, Mladenic and Grobelnik (2006), where a fully automated evaluation approach was proposed by introducing a similarity measure called OntoRand index and comparing ontologies to a gold standard one using that measure. This kind of evaluation is typically applied to the ontologies that are generated semi-automatically, and to measure the effectiveness of the ontology generation process (Yu, Thom and Tam, 2007). A major challenge in this approach is that comparing ontologies is not easy (Sabou, Lopez and Motta, 2006).

2.4.5.3 Data or Corpus Driven Evaluation

This approach is similar to the "gold standard" approach, but instead of comparing an ontology to another one, it compares it to a source of data or a collection of documents (Brank, Grobelnik and Mladenic, 2005). One of the most popular architectures for this type of evaluation was proposed by Brewster *et al.* (2004) and was based on three main steps, namely, extracting keywords from a corpus, applying some query expansion algorithms to the ontology concept, and mapping the terms identified in the corpus to the concepts in an ontology. The final step was to analyse how well the ontology is covering the source of data (Brewster *et al.*, 2004).

2.4.5.4 Task-Based Evaluation

Also known as application-based (Fahad and Qadir, 2008) or black box evaluation (Obrst *et al.*, 2007); this approach aims to evaluate an ontology's performance in the context of an application (Brewster *et al.*, 2004). One of the main assumptions of this approach is that there is a direct link between the quality of an ontology and how well it serves its purpose as a part of a larger application (Netzer *et al.*, 2009). The challenges of performing this type of evaluation include the difficulty of assessing the quality of the performed task as well as making sure that the experimental environment is clean, and that the ontology is the only factor influencing the performance of the application (Sabou, Lopez and Motta, 2006).

2.4.5.5 Rule-Based (Logical) Evaluation

This type of evaluation was proposed by Arpinar, Giriloganathan and Aleman-Meza (2006) and aimed to validate ontologies and detect conflicts in them by using different rules that were either a part of the ontology development language or were identified by users. Rule-based evaluation is more relevant when evaluation aims to detect faults and inconsistencies in an ontology, rather than when the quality assessment or ontology selection is concerned.

2.4.5.6 Other Approaches

Besides the above-mentioned categories, that are very popular in the literature, there are some other ways of classifying ontology evaluation approaches. For example, they can be classified into a glass-box or black-box. Glass-box approaches tend to evaluate the internal content and structure of ontologies (Gangemi *et al.*, 2006) and are blamed for not predicting how ontology might perform in an application. In contrast, black-box approaches do not explicitly use knowledge of the internal structure of ontologies and focus on the quality of an ontology performance and results (Gangemi *et al.*, 2006). Ontologies can also be evaluated as a whole or according to their different layers, e.g., data level, taxonomy level, and application level (Brank, Mladenic and Grobelnik, 2006). Moreover, Bandeira *et al.* (2016) divided the concept of ontology quality into two broad types: “Total Quality” and “Partial Quality”. Some, however, have argued that evaluating an ontology as a whole, especially automatically, is not possible or practical, especially considering the complex structure of ontologies (Brank, Grobelnik and Mladenic, 2005).

Among all the above-mentioned approaches, the main focus of this study is on criteria-based ontology evaluation, as it is one of the most important topics in the ontology evaluation and selection domain. Moreover, investigating the quality of ontologies, and the metrics that it depends on, is a continuing effort in the ontology domain. This approach, a summary of different quality related metrics, and how they can be measured will be discussed in more detail in the next section.

2.5 Criteria-Based Evaluation

Criteria-based evaluation, also known as metric-based, multiple-criteria (Brank, Grobelnik and Mladenic, 2005) or feature-based (Arpinar, Giriloganathan and Aleman-Meza, 2006), is one of the most popular evaluation approaches in the literature. This type of evaluation is mostly based on identifying and selecting multiple attributes or features of ontologies and then evaluating them for ranking and selection purposes (Brank, Grobelnik and Mladenic, 2005). The outcome of this approach is usually an overall or an aggregated score that is computed by adding the scores assigned to each evaluation criterion (Butt, Haller and Xie, 2016; Martínez-Romero *et al.*, 2017). Despite the wide use and popularity of criteria-based evaluation, identifying the right set of metrics for ontology evaluation and measuring them is still a challenge.

Two different methods, namely, inductive and deductive, are proposed for identifying evaluation metrics. In an inductive approach, there is no predefined set of evaluation metrics and ontologies are tested empirically to identify the potential set of characteristics that will lead to a favourable outcome for an application. Deductive approach, in contrast, uses theory and previous research to identify relevant quality metrics (Burton-Jones *et al.*, 2005). It is easier to generalize deductive elements across different contexts; however, finding the best theory or the one with the highest quality that can be used as a benchmark is not that easy. Moreover, elements that have been identified inductively are useful in at least one application context. Instead of being based on a set of predefined metrics, this research seeks to investigate the ontologists and knowledge engineers' views on the topic of quality and to inductively identify the metrics that they find important in the evaluation process.

Despite all the similarities, criteria-based approaches are different from each other in a number of respects. First, the type of metrics used to assess ontologies can be different. Some approaches are based on qualitative metrics and tend to rely on users' judgement and ratings about an ontology or a module in an ontology (Porzel and Malaka, 2004). Qualitative approaches can also be used to evaluate an ontology based on the principles that are/were used in its construction (Brewster *et al.*, 2004). Others, however, are based on quantitative criteria about different aspects of ontologies, such as their structure and content. These approaches, that are also known as formal rational approaches, are usually concerned with technical and economic aspects of ontologies and use different goal-based strategies (Maiga, 2008).

Criteria-based evaluation approaches can also assess internal and/or external attributes of ontologies. Internal attributes are concerned with the ontology itself and its internal organisation, whereas external measures are mostly focused on how ontologies are taken-up or used within the user communities (Kehagias *et al.*, 2008). Burton-Jones *et al.* (2005), for example, followed software engineering measurement traditions to propose a method that aimed to identify what they called key internal attributes of ontologies, including consistency, richness, and clarity; maintainability and application performance were used as examples of external quality attributes.

Moreover, metrics used in the criteria-based evaluation can either be query dependent or query independent. Coverage, for example, can be used to measure how well a candidate ontology matches or covers a set of query term(s) and selection requirement(s) (Buitelaar, Eigner and Declerck, 2004; Buitelaar and Eigner, 2008) and therefore, it is query dependent. Popularity, in contrast, is measured by checking the presence of an ontology in different well-known repositories, as well as looking into the number of visits or page views to an ontology in ontology repositories in a recent specific period (Martínez-Romero *et al.*, 2017); hence, it is query independent and does not depend on selection requirements. This research tends to focus on query independent criteria, e.g., the social aspects of ontologies, rather than the query dependent ones.

For the purpose of this research and according to the previous study conducted by Talebpour, Sykora and Jackson (2017), criteria for evaluating the quality of ontologies have been broadly classified into three main subgroups, including (1) Internal metrics that are based on different internal characteristic of ontologies, such as their content, and structure, (2) Metadata that are used to describe ontologies, and (3) Social metrics that focus on how ontologies are used by communities. The rest of this section moves on to explain different quality metrics for ontology evaluation in more detail.

2.5.1 Internal Criteria and Ontology Evaluation

Internal components of ontologies have always been a mean for their evaluation. Different criteria based on the internal characteristics of ontologies, namely, clarity, correctness, consistency, and completeness are used in the literature to measure how clear an ontology's definitions are, how correctly different entities in an ontology represent the real world, and how consistent and complete an ontology is (Yu, Thom and Tam, 2009). Coverage is yet another very important content related metric; the term coverage is mostly used in the literature to measure how well a candidate ontology match or cover a set of query term(s) and selection requirement(s) (Buitelaar, Eigner and Declerck, 2004; Buitelaar and Eigner, 2008). Domain coverage (Martínez-Romero *et al.*, 2014), input coverage (Martínez-Romero *et al.*, 2017), and topic coverage (Wang, Guo and Fang, 2008) are among some of the other terms used in the literature to refer to a similar concept.

Some of the early ontology evaluation approaches were blamed for being concerned with the ontology syntax and ignoring the other important components or features of ontologies, such as their internal structure (Ning and Shihan, 2006). The structure has been assessed by many of the evaluation tools and frameworks and is usually evaluated by measuring the ratio of the number of properties to the number of classes in ontology (Buitelaar, Eigner and Declerck, 2004). Moreover, the graph structure (Gangemi *et al.*, 2006) of ontologies can be assessed to measure how detailed their knowledge structure is (Fernández *et al.*, 2009) and also to evaluate their richness of knowledge (Sabou, Lopez and Motta, 2006), density (Yu, Thom and Tam, 2007), depth, and breadth (Fernández *et al.*, 2009). It has been suggested that the number of properties in an ontology has a direct relationship with the quality of its structure (Buitelaar and Eigner, 2008). The structure is important not only because a well-structured ontology will facilitate the process of ontology share and reuse, but also because ontologies should satisfy and cover the domain knowledge, which has its own structure (Ning and Shihan, 2006).

Evaluating the internal characteristics of ontologies is very important and also helpful in different cases. However, as the result of all the developments around the semantic web and the increase in the number of ontologies, being assessed using only internal features might not be enough and evaluation approaches have started considering other aspects of ontologies, namely, metadata about them and how they are used by communities. These two aspects are discussed in the following sections.

2.5.2 Metadata and Ontology Evaluation

Some of the frameworks and tools have suggested that providing ontologies with different types of metadata can be helpful in the process of classifying, evaluating and ranking them (Ding *et al.*, 2004). Metadata or "data about data" is widely used on the web for different reasons, such as, to help in the process of resource discovery (Gill, 2008). Sowa (2001) argued that the primary connection between different elements of an ontology is in the mind of the people who interpret it; so, tagging an ontology with more data will help to make those mental connections explicit. Considering the ontology selection process, metadata is mostly used in the ontology libraries and for classification and organisation purposes. However, as seen in the rest of this section, using metadata can be helpful in mapping systems and to help interoperability between different applications.

Ontologies can be tagged and described according to their different characteristics, namely, type, version and the language that they are built and implemented with (Lozano-Tello and Gómez-Pérez, 2004). As it is seen in the literature, Swoogle was one of the very first selection systems to introduce the concept of metadata in the ontology domain (Ding *et al.*, 2004). In this system's architecture, a component is responsible for creating and storing three different types of objective metadata about each discovered ontology by this search engine, including basic, relation, and analytical metadata (Ding *et al.*, 2004).

Supekar (2005) also proposed two sets of metadata that can be used to evaluate ontologies: source metadata and third-party metadata. The domain of the ontology, version number, verification tools used and development methodology, naming policy and peer reviews are among these two sets of metadata, that can either be provided by the author of an ontology or by the users of an ontology (ibid.).

Metadata is also a very popular concept when it comes to ontology libraries. Ontology libraries are not only interested in collecting ontologies, but they also need to collect different sets of information, such as, the ontology name and domain, information about who created an ontology and when it was created, licensing and versioning policies to facilitate ontology search and reuse (d'Aquin and Noy, 2012). There are different ways of collecting metadata about ontologies; some of the metadata such as name and creator are attached to an ontology itself as properties (Schober *et al.*, 2012). Moreover, those libraries that allow users or authors to submit ontologies, usually provide them with some facilities, e.g., a form-based interface that allows them to submit different types of provenance information about the ontologies they want to submit (d'Aquin and Noy, 2012).

Using metadata is also very common in ontology mapping. Ontology mapping is defined as the task of finding and making relationships between different concepts of ontologies in the same domain (Kalfoglou and Schorlemmer, 2003). Noy, Griffith and Musen (2008) argued that providing metadata for ontology mapping is critical, especially in cases where mapping can come from many different sources. According to them, different types of metadata or additional information can be provided for each mapping, e.g., general comments on mapping, user discussions, and comments on a mapping, application or context of the mapping, the date that mapping was created, and the user who performed the mapping.

Moreover, metadata is created and used to help interoperability between different applications and ontologies. Ontology Metadata Vocabulary (OMV), for example, is one of the most popular sets of metadata for ontologies (Hartmann, Palma and Sure, 2005); it allows ontologists and knowledge engineers to describe their work using a wide range of metadata about different aspects of ontologies, such as, ontology type, ontology language, ontology syntax, ontology base, and license model. OMV is not directly concerned with ontology evaluation or ranking, and its main aim is to support and facilitate ontology reuse. Moreover, Matentzoglu *et al.* (2018) proposed a guideline for minimum information for the reporting of an ontology (MIRO); MIRO aims to help ontologists and knowledge engineers in the process of reporting ontology description and providing documentation. MIRO developers claim that it can improve the quality and consistency of the information in ontology descriptions and documentation.

Providing and using metadata for discovery purposes is very popular, especially among repositories and libraries for ontologies (Jonquet *et al.*, 2018). However, too little attention has been given to the role of the metadata in ontology evaluation process. In other words, the quality of an ontology has not been measured based on the metadata that is available for it. For example, while information about the language that an ontology has been built with is available, it is not clear if the choice of language can affect the quality and reusability of an ontology. This study addresses this research gap by investigating the potential role of metadata in the evaluation process.

2.5.3 Community and Ontology Evaluation

One of the interesting questions in the ontology domain is how social interactions among ontologists and knowledge engineers can affect different activities in this domain, from ontology development to ontology evaluation and selection for reuse. As it is seen in the literature, the community plays an important role in ontology development processes. Collaborative Protégé[13], as a well-known example, supports collaborative ontology development by allowing users to comment on different components of ontologies, e.g., classes

[13] https://protegewiki.stanford.edu/wiki/Collaborative_Protege

and properties as well as making changes in an ontology. This system also enables users to communicate and interact with each other live and allows them to propose or vote for changes (Tudorache and Noy, 2007). BioPortal and AgroPortal also try to facilitate interactions between ontology users, by enabling them to add notes to ontologies (Jonquet *et al.*, 2018).

Mailing lists and ontology-related forums are also used to support social and community interactions between ontology users and ontology developers. Gene Ontology (GO)[14], for example, supports its user community through different mailing lists, namely, GO mailing list, GO friends mailing list, GO discussion, GO helpdesk, and GOC Ontology Editors list. These mailing lists have different access levels; some of them are open to the public and provide an environment to discuss different aspects of GO, such as, content, and structure. Some others, however, are only open to GO consortium members or GO editors. LOV also supports community interactions through a google group where authors, publishers, and users of LOV can share and discuss different matters concerning ontology engineering and linked data.

Besides how ontologies are built and what they are covering or even not covering, some have suggested that how they are used by communities can be can be an effective factor in their evaluation process. Mcdaniel, Storey and Sugumaran (2016), for example, argued that there is a link between the quality of an ontology and community approval and participation in its evolution. The term "Social Quality" has also been used to reflect the existence of ontologies, as well as agents and users in ontology communities (Burton-Jones *et al.*, 2005) and refers to the level of agreement about an ontology and among different participants or members of a community (Mcdaniel, Storey and Sugumaran, 2016).

A study conducted by Lewen and d'Aquin (2010) suggested that relying on the experiences of other users for evaluating ontologies will lessen the efforts needed to assess an ontology and reduce the problems that users face while selecting an ontology. Mcdaniel, Storey and Sugumaran (2016) also highlighted the importance of relying on the wisdom of the crowd in ontology evaluation and argued that improving the overall quality of ontological content on the web is a shared responsibility within a community.

As it is seen in the literature, social and community features of ontologies have not been the main focus of many of the evaluation frameworks until recently. One of the main questions of this research is how communities and social interactions amongst them affect the process of

[14] http://geneontology.org/

evaluating the quality of ontologies, before selecting them for reuse. The rest of this section will review the existing literature on social based ontology evaluation; it starts with reviewing different metrics and criteria and moves on to describe some of the frameworks and approaches.

2.5.3.1 Popularity

Popularity is probably the most mentioned social metric in the literature and many of the research in this field have either identified a metric called popularity or have somehow referred to a similar concept, but by using a different term. In the approach proposed by Burton-Jones *et al.* (2005), that was influenced by software engineering and the semiotic framework proposed by Stamper *et al.* (2000), ontology quality was known to be a multidimensional formative construct formed by different sets of metrics, namely, syntactic quality, semantic quality, pragmatic quality, and social quality. Social quality focused on how community affect the quality of ontologies and depended on two main factors: (1) authority, or the number of other ontologies that refer or link to an ontology and (2) history, that is about the number of times an ontology is accessed (Burton-Jones *et al.*, 2005).

These factors were then used by Mcdaniel, Storey and Sugumaran (2016) to propose an approach that aimed to facilitate the process of ontology selection by assessing community recognition of an ontology. Their proposed system would take keywords and desired weights for the selection metrics as input, and recommend a list of potentially relevant ontologies, that were ranked based on their social quality, in the output. The effectiveness of this approach was tested by applying the social quality metrics to about four hundred of the ontologies in NCBO BioPortal. The results of this experiment suggested a reasonable link between the social quality of an ontology with the number of site visits and the linkage between the ontologies. OntoKeeper, as one of the most recent evaluation tools, also used the metrics proposed by Burton-Jones *et al.* (2005); however, it is not clear if and how they evaluate the social aspects of the ontologies (Amith *et al.*, 2018).

ONTOMETRIC as one of the most well-known criteria-based evaluation frameworks for ontologies had a dimension called methodology, that depended on a factor called Maturity. The maturity of a methodology for an ontology could be evaluated by counting the number of ontologies that have been developed using that method, the importance of those ontologies and also the number of different domains that those ontologies belonged to (Lozano-Tello and Gómez-Pérez, 2004). Mcdaniel, Storey and Sugumaran (2016) argued that these factors are linked to the social acceptance of a methodology, thus they are social based evaluation metrics.

Popularity and acceptance tend to be mostly used to refer to the number of times an ontology has been viewed or used in a specific repository. In BiOSS, for example, the popularity of an ontology in a community was measured by checking if the ontology or some other information about it exists in different resources, namely, PubMed[15], NCBO BioPortal, Wikipedia and Twitter[16] (Martínez-Romero *et al.*, 2014). NCBO BioPortal Recommender also calculates the popularity of an ontology by checking its presence in different well-known repositories, as well as looking into the number of visits or page views to an ontology in ontology library in a recent specific period (Martínez-Romero *et al.*, 2017).

The other definition of popularity is based on applying the PageRank algorithm (Page *et al.*, 1999) to the ontology domain and focuses on the import feature of ontologies. Fernández et al. (2009), for example, defined the term "direct popularity" as the number of ontologies importing a given ontology. Wang, Guo and Fang (2008) used the same definition to describe what they called popularity, which for them, was measured by considering the number of times an ontology has been referenced by others. Moreover, as a part of the authority metric in Supekar, Patel and Lee (2004), a metric called "citation" was introduced and was defined as the number of occurrence of "daml:sameClassAs", "rdfs:seeAlso", and "owl:imports" in a given ontology.

Connectedness is yet another term used to refer to a similar concept. In Buitelaar and Eigner (2008) for example, connectedness was used to measure the number of ontologies that were included in a candidate ontology, as well as measuring how well established those ontologies were. To measure connectedness, they did not only consider the number of ontologies that are imported by a candidate ontology but also considered the quality of those ontologies (Buitelaar, Eigner and Declerck, 2004).

Besides using the import feature to measure the connectedness between ontologies, two other types of linkage and relationships, namely, extension and use-term have been identified and considered in Swoogle (Ding *et al.*, 2005). Extension refers to the situation when a semantic web document defines a term using the terms defined in another semantic web document; the use-term happens when one semantic web document uses a term defined by another (Ding *et al.*, 2005).

[15] https://www.ncbi.nlm.nih.gov/pubmed/
[16] https://twitter.com/

Despite the widespread use of the terms popularity and acceptance in the literature, there is still no consensus on the meaning of the term "popularity" in the ontology evaluation context, what it depends on and how it should be measured. This study aims to investigate the social aspects and popularity of ontologies in detail and determine their role and importance in the evaluation process.

2.5.3.2 Reviews and Ratings

The process of evaluating ontologies is known to be very time-consuming and challenging; some argue that it can be facilitated by allowing developers and users of ontologies to annotate and review them. There are different ways of annotating ontologies; for example, ontology users can describe them using different sets of metadata, and also comment on a wide variety of their dimensions, namely, by providing different qualitative information (e.g., reviews) or quantitative information (e.g., ratings and rankings) (Noy, Guha and Musen, 2005).

Using ratings and reviews is not that common in the ontology domain. Two of the most known examples of considering reviews in the evaluation process are proposed by Noy, Musen and Guha (2005) and Smith (2008). Noy, Musen and Guha (2005) argued that ontology evaluation is a subjective task and argued that while it is easy to agree that an ontology is "bad", it is not easy to find or call an ontology "good". They also claimed that users evaluate ontologies differently, based on their expertise and their choice of evaluation metrics (Noy, Guha and Musen, 2005). Therefore, they proposed what they called an open rating system or democratic ranking; this type of system would not only allow everyone to publish content, ontologies in this case, but would also allow them to provide reviews and rankings for ontologies.

The notions of the open rating system and democratic ranking was also used in a social-based evaluation approach proposed by Lewen *et al.* (2006). This approach consisted of some ontologies, some users or agents, some values that can be used to evaluate ontologies (e.g. 1 star, 2 stars, etc.), some values that can be used to evaluate other users' reviews about an ontology (e.g. thumbs up or thumbs down), and also two different functions, one to store ontology ratings and the other one to store the ratings of the reviews. They had also identified a new concept called "meta-rating", which would allow users to not only review an ontology itself but to also review the content or the reviews of other users about an ontology. Meta-rating or how users review each other's reviews had a direct link with the concept of trust in the open rating systems and corresponded to the feeling that a review or information provided by an individual will be correct or useful.

In contrast to open rating systems, ontologies in a closed rating system are only evaluated by two different groups of editors, namely, coordinating editors and associated editors (Smith, 2008). In OBO Foundry, for example, associate editors are responsible for editing the ontologies and ensure that they follow the set of principles proposed by the Foundry method (Smith, 2008). Coordinator editors, however, are responsible for harmonizing interactions between different ontology development projects. Moreover, ad hoc reviewers with special expertise are asked to provide reviews, if the occasion demands. This approach can only be applied to domains in which there are some known accepted quality criteria by the community and also, a group of trusted reviewers who comply with those criteria to assess the subject of evaluation, that are ontologies in this case. The editorial process of OBO Foundry is only applied to the ontologies in the biomedical and biomedicine domain (Smith *et al.*, 2007).

There are some known advantages and disadvantages for both of the above-mentioned approaches and many different ways of comparing these two schools of thoughts. The first and maybe the most notable difference is about the evaluators and the providers of reviews. The open rating systems follow the web motto that is "anyone can say anything about anything" and provide the opportunity of evaluating and ranking ontologies for anyone who is willing to do so (Noy, Guha and Musen, 2005). In contrast, ontology evaluation on the closed rating systems like OBO Foundry are generated by a small group of editors (Smith, 2008).

Being open to the public or only open to a specific group of experts will dramatically affect the number of ratings and reviews generated for each ontology. The huge amount of information and reviews generated in the open rating systems forces them to deal with a number of challenges, including finding a method to aggregate the rankings, ratings and reviews. Having no restriction on the people who create reviews, it can also be argued that the quality of a significant portion of the reviews and annotation generated in the open rating systems is lower than the ones generated by editorial reviews (Noy, Guha and Musen, 2005; Smith, 2008).

Who and which reviewer to trust is the other challenge that should be addressed in an open rating system. To solve this problem, some have suggested providing a mechanism that allows rating the rankings and ratings and filter out those with a lower quality (Noy, Guha and Musen, 2005). Furthermore, Lewen and D'aquin (2010) blamed open rating systems for not allowing what they called "multi-faceted" review. They argued that different aspects or features of ontologies might have a different level of quality and therefore, assigning an overall review or score to an ontology might not be helpful in some cases. They also stated that agents and users trust each other differently based on different aspects of an object, that is an ontology in this case.

To address these issues, they proposed personalized rankings for ontologies that are not only based on different components of open rating systems, such as users' ratings and the trust relations between them but also considers what they defined as a meta-trust statement (Lewen and d'Aquin, 2010). According to this statement, a user might trust another user to review a specific aspect (e.g. coverage) of a specific object (e.g., an ontology), a specific aspect of all objects, all aspects of a specific object or all aspects of all objects.

It has been argued that the quality of reviews provided in a closed rating system is higher as the reviews are created by a group of recognised experts with a willingness to contribute to ontology development and evaluation (Smith, 2008). However, this approach faces different limitations and challenges, such as, scalability. The evaluation process in the closed rating systems has been criticized for being very time-consuming, non-quantifiable and non-periodic. Therefore, closed rating systems might not be able to deal with the dramatic increase in the amount of content, that is created every day (Noy, Guha and Musen, 2005).

Moreover, some have argued that ontologies should be used before being evaluated and have blamed the evaluation process in the closed rating systems (Noy, Guha and Musen, 2005) for being based on inspection, rather than assessing the run time performance (Smith, 2008). In a closed rating system, the editorial review is open, meaning that people would know who has reviewed an ontology. It can be argued that how people behave or evaluate an ontology in a public forum is influenced by different social aspects, such as status, personal relationships and also political considerations (Smith, 2008).

In summary, open rating systems have to deal with challenges like the huge number of reviews, quality of reviews and trust related problems. Closed rating systems, however, have to deal with challenges like the great number of ontologies developed daily, ontology evolution, maintenance and management, and the fact that their evaluation is mostly based on inspection rather than actually using an ontology in an application. This research investigates the role and importance of the social interactions and reviews in the ontology evaluation process.

2.5.4 Ontology Evaluation Measures

In evaluation, it is not only important to identify different sets of metrics that can be used to assess the quality of an ontology, but it is also important to find a method to measure those metrics (Hlomani and Stacey, 2014). Pressman (2005, p. 434) defined measures as "a quantitative indication of the extent, amount, dimension, capacity, or size of some attribute of a product or process". Having measurements in mind, metrics for ontology evaluation can be categorised into three main groups: metrics like consistency and popularity that can be successfully determined and measured using the available ontology evaluation tools or formulas (Yu, Thom and Tam, 2009), metrics like correctness that are difficult to measure automatically and experts are needed to manually inspect and measure them (Yu, Thom and Tam, 2009) and metrics like completeness that are difficult to evaluate, as there is no means to determine them (Yu, Thom and Tam, 2007).

Unlike the definition provided for measuring in software engineering by Pressman (2005, p. 434), ontology quality measures do not only accept quantitative numeric values, but different relative, categorical, and qualitative values can also be assigned to them. As an example of assigning numeric values to ontology features, Cantador, Fernandez and Castells (2007) measured correctness, readability and flexibility of ontologies using discrete numeric values between 0 to 5, where 0 meant that an ontology does not fulfil the criterion and 5 meant that an ontology completely satisfies the criterion. Burton-Jones *et al.* (2005) also normalized the score assigned to different evaluation metrics and came up with a total score with the absolute value between zero and one. Popularity in the BiOSS system can be used as an example of a metric with a relative value (Martínez-Romero *et al.*, 2014); the values of relative metrics are not precise and depend on an external benchmark (Burton-Jones *et al.*, 2005).

In ONTOMETRIC, it was argued that humans usually make non-numeric judgments and therefore a linguistic scale (very_low, low, medium, high and very_high) was suggested to measure different quality features of ontologies (Lozano-Tello and Gómez-Pérez, 2004). Similar categorical values were used by Cantador, Fernandez and Castells (2007) to measure the type of the model (e.g. domain, upper level or application ontologies) and the level of formality (highly informal, semi-informal, and rigorously-formal) of an ontology. Finally, metrics related to social and community aspects of ontologies can be measured using different qualitative reviews (Noy *et al.*, 2009).

Measurable quality aspects of ontologies can either be determined in isolation or in the context of an application. Yu, Thom and Tam (2007) blamed the criteria-based evaluation approaches for measuring the quality of ontologies in isolation and out of the context of an application or benchmark. They used coverage as an example and argued that measuring how well an ontology covers a set of requirements or application needs will make more meaningful results compared to measuring how well an ontology is covering a domain (ibid.). What they have argued is true for some metrics like coverage; however, many of the evaluation criteria, namely, the social and community related ones, can be measured out of the context of an application.

Besides the type of the metrics used in the evaluation process, how they are measured and the values that are assigned to them, the final quality score of an ontology depends on the weight that is assigned to the value of each feature. Burton-Jones *et al.* (2005) suggested that ontology quality is a formative construct, meaning that it is formed by different measures, all of which can equally be important but need not to be and might have a different level of importance assigned to them. In NCBO BioPortal Recommender, for example, the weight assigned to ontology coverage is much higher than the weights assigned to the other quality features, namely, acceptance, details and specialization, 0.55, 0.15, 0.15 and 0.15 respectively (Martínez-Romero *et al.*, 2017).

2.6 Challenges, Limitations, and the Research Gap

The notion of ontology quality and the process of evaluating it is one of the most significant and also complicated challenges in the ontology domain. Recent evolution in the semantic web has led to an incredible increase in the number of available ontologies and therefore, the main challenge today is not finding ontologies, but it is to find the right ones, or as Kolbe *et al.* (2019) called it, the "well-fitting and requirements-meeting" ones. Despite the importance of this matter and the widespread research on this topic, there are still many unanswered questions about ontology evaluation and selection for reuse.

Diversity of tools, languages, and methodologies, lack of a generic framework for building ontologies and also identifying the relevant quality elements that should be used for ontology assessment and evaluation are among some of the main challenges in the ontology domain (Burton-Jones *et al.*, 2005; Hlomani and Stacey, 2014). A considerable number of tools and systems have been proposed in the literature to support the process of ontology selection and reuse; however, many of them are not available anymore. Moreover, most of the researches in the ontology domain are concerned with constructing a new approach, instead of coming up with a generic one or something like a common benchmark for ontology development.

Identification of the set of metrics that should be used in the evaluation process has known to be a central issue of this domain (Burton-Jones *et al.*, 2005). As was discussed in section 2.5, a wide range of metrics have been proposed in the literature to help ontologists and knowledge engineers assess the quality of ontologies before selecting them for reuse. However, as seen in the literature, some of those proposed metrics are very similar and different names have been used to refer to the same criterion. Moreover, previous studies have failed to identify the importance of each of those metrics in the ontology evaluation process. Metrics proposed for ontology evaluation usually tend to be useful in a specific context or application and it is hardly possible to find a general quality metric that will work for different use cases and scenarios (Burton-Jones *et al.*, 2005).

Knowledge captured by ontologies might change regularly and not everyone in a domain might agree on what is presented as facts in an ontology or their meaning and relationships (Bard and Rhee, 2004). Hence, subjectivity has been considered as one of the issues that the research in ontology evaluation domain is concerned with. According to the literature, the main focus of many of the current approaches is on identifying and measuring different objective characteristics of ontologies, such as consistency, semantic validation, and hierarchy. In the real world, however, ontologies are usually evaluated subjectively, meaning that ontologists and knowledge engineers are usually looking for a "well-defined" ontology that "best" fits their application requirements.

Hlomani and Stacey (2014) have identified three different types of subjectivity in ontology evaluation: (1) subjectivity in the selection of the criteria for ontology evaluation, (2) subjectivity in the thresholds for the measurements of each criterion, and (3) influences of subjectivity on the overall quality evaluation. As a part of the ontology evaluation process, it is important not only to come up with a set of right criteria that can be used to evaluate ontologies, but it is also very important to establish who the right ontology evaluators are (Hlomani and Stacey, 2014).

Supekar (2005) blamed the evaluation approaches of their time for neglecting the importance of subjective qualities of ontologies and for not providing helpful subjective information, such as peer reviews and ratings for ontologies. By contrast, Gangemi *et al.* (2006) argued that automatic or semi-automatic techniques should be applied to ontology evaluation and make the evaluation process less subjective. Despite all the endeavour in this field, there is still no automatic method or approach that can be used to assess the quality of an ontology (Amith *et al.*, 2018) and ontology evaluation and selection has always been based on some kind of human experts' judgment (Lewen and d'Aquin, 2010). Hence, subjectivity and bias are some of the inevitable parts of these types of evaluation processes; this is against the idea of the good science, that is to exclude subjectivity from scientific experiments (Hlomani and Stacey, 2014).

Community is one of the most significant notions in the ontology domain. One of the main aims of developing ontologies was and still is to use them as a shared conceptualization between different groups of people working in the same domain (Gruber, 1993). However, most studies in the field of ontology evaluation have only focused on syntactic and structural aspects of ontologies and have failed to address how social aspects and community recognition affect the quality of an ontology (Mcdaniel, Storey and Sugumaran, 2016).

As was mentioned before, some of the evaluation approaches like OntoMetric (Lozano-Tello and Gómez-Pérez, 2004), Semiotic Metric Suite (Burton-Jones *et al.*, 2005) and NCBO BioPortal Recommender (Jonquet, Musen and Shah, 2010) have identified social based quality metrics; however, Mcdaniel, Storey and Sugumaran (2016) argued that they are not able to fully evaluate the level of ontology acceptance in the community. It can be argued that social quality of an ontology might depend on a wide variety of metrics, rather than being only based on acceptance or popularity.

Besides different dimensions of ontologies that have been studied in the literature and different metrics that have been identified to assess those dimensions, how evaluation metrics are measured is the other very important aspect of ontology evaluation. Mcdaniel, Storey and Sugumaran (2016) blamed some of the measurement approaches for reducing the assessment accuracy. Acceptance in BioPortal, for example, is based on the existence of or the number of visits to an ontology in a specific website; therefore, this metric is not applicable to the ontologies that are in other libraries (Mcdaniel, Storey and Sugumaran, 2016).

To address some of the above-mentioned shortcomings, this study investigates the ontologists and knowledge engineers' opinions about the notion of quality and the factors it depends on. Moreover, this research aims to examine the role of community in ontology evaluation and determine how social interactions can help in the evaluation process. The importance of different quality metrics will also be assessed. Finally, it will be investigated if there exists a generic set of quality metrics that people tend to consider for ontology evaluation, no matter in what domain they work in.

Chapter 3: METHODOLOGY and METHODS USED

Before collecting any data, it is important to choose a research methodology that supports the aims and objectives outlined in the previous chapters. The literature review has shown the importance of ontology evaluation and selection in the reuse process. Reuse is a fundamental property of the ontology domain, not only because ontology development is very costly and time-consuming, but also because the choice of ontology directly affects the performance of the systems that use those ontologies. Although this research is unique within the field of ontology, as it investigates the notion of ontology quality from the knowledge engineers' perspective, the methods used to collect and analyse the research data have a number of commonalities with that reported in the published literature in different related domains, such as ontology evaluation, ontology selection, and ontology recommendation.

This chapter covers the research methodology and explains the rationale behind the final chosen path. It starts by discussing different philosophical worldviews and moves on to describe the purpose and approach of this study. Different strategies of inquiry and the methods used in this research are then discussed in detail. Sampling strategies and data collection, analysis, and integration methods are discussed in section 3.9, 3.10, 3.11, 3.12 respectively. The last part of this chapter will discuss the ethical considerations of this research.

3.1 Introduction

Research design starts by identifying and clarifying the research philosophy (Saunders, Lewis and Thornhill, 2012), which is also known as philosophical worldviews (Creswell, 2014), paradigm or epistemology and ontology (Crotty, 1998; Gray, 2013). Saunders, Lewis and Thornhill (2009, p. 108) defined research philosophy as the "assumptions about the way in which you view the world". Creswell (2014, p. 6) has used the term "worldview" to refer to a comparable concept; he stated that worldview is a "general philosophical orientation about the world and the nature of the research that a researcher brings to a study". Similarly, paradigm has been defined as "a person's worldview, complete with the assumptions that are associated with that view" (Tashakkori and Teddlie, 2003, p. 139). Crotty (1998, p.3) has also used the term epistemology to refer to the "theory of knowledge embedded in the theoretical perspective and thereby in the methodology".

Before choosing a philosophical position for research, four main perspectives, namely, ontology, epistemology, axiology and methodology (Creswell, 1994) should be taken into the consideration (Bryman, 2016). Ontology is concerned with the nature of "reality" (Saunders, Lewis and Thornhill, 2012, p. 130) and "social entities" (Bryman, 2016). There are two different aspects of ontology: (1) objectivism (also known as realism), and (2) subjectivism (also known as relativist and constructionism) (Bryman, 2016). According to objectivism, the reality is objective and singular (Creswell, 1994, p. 5) and social entities exist in a reality that is external to and independent from the social actors (Saunders, Lewis and Thornhill, 2012, p. 131). Subjectivism, however, claims that reality and social phenomena are subjective, multiple (Creswell, 1994, p. 5) and are "created from the perceptions and consequent actions of social actors" (Saunders, Lewis and Thornhill, 2009, p.111).

Epistemology is concerned with the relationship between the researcher to that researched (Creswell, 1994, p. 5), or the relation between the knower to the known (Guba and Lincoln, 1989). It tries to find an answer to "what is (or should be) regarded as acceptable knowledge in a discipline" (Bryman, 2016) or to "how can we be sure that we know what we know?" (Guba and Lincoln, 1989). There are two main ways of answering these questions. One is by taking the dualist objectivist epistemology view, also known as etic or outside perspective (Morris *et al.*, 1999), and believing that the researcher should remain independent (Creswell, 1994, p. 5), detached and distant from the phenomenon studied. The second way is to take the subjective epistemology view, also known as emic or inside perspective (Morris *et al.*, 1999), and believing that the researcher and the phenomenon are interlocked, and that the findings are the "literal creation of the inquiry process" (Guba and Lincoln, 1989, p.84).

Axiology is concerned with the role of values and ethics in a study (Killam, 2013, p.6). Researchers can either keep their own values out of the study and only report the facts and evidence that they gather in a study or they can actively report and mix their values and biases with the information gathered from the field. Quantitative studies are known to be value-free and unbiased, whereas qualitative research is value-laden and biased. Finally, methodology is concerned with the entire process of a study. Based on the choices made in the previous steps, methodology can aim to explain nature as it really is and as it really works and establish a cause-effect relationship, or it can shape patterns, theories, and factors and lead to the emergence of a joint construction of a case (Guba and Lincoln, 1989; Creswell, 1994).

There are many different philosophies identified in the literature; four of the widely discussed and used ones are described in the following:

- **Postpositivism.** A deterministic philosophy that claims "data, evidence, and rational considerations shape knowledge" (Creswell, 2009). According to this worldview, there exists a single, but not an absolute (Creswell and Creswell, 2018, p.7), reality and researchers can either reject or fail to reject hypothesis related to that reality by objectively collecting data and deductively analysing them (Creswell and Clark, 2017, p.24).
- **Constructivism.** Also known as, social constructivism; believes in multiple realities and perspectives and states that "meanings are constructed by human beings as they engage with the world they are interpreting" (Creswell, 2009, p.8). According to this viewpoint, meaning or truth is not simply objective or subjective and it cannot be created, but it is constructed (Crotty, 1998, p.43). Research based on this type of philosophy will try to use many open-ended questions to explore people's ideas and views about the subject of the research. Those views will then be analysed inductively and are biased toward the researcher beliefs and interpretations (Creswell and Plano Clark, 2007, p. 24).
- **Advocacy and Participatory.** This perspective aims to address an important issue or problem in society and tries to bring a change in practices, by actively involving, engaging and collaborating with participants in different stages of research, such as, research design, data collection, and data analysis (Creswell and Plano Clark, 2007; Creswell, 2009).
- **Pragmatism.** This viewpoint is widely associated with mixed methods. Pragmatism is not committed to any of the discussed philosophies (Creswell, 2009, p.10) and gives researchers the freedom of choice by enabling them to use the method that best answer their research questions, to analyse the data both inductively and deductively, to have both objective or subjective view (epistemology), depending on the phase of research, and to have diverse viewpoints regarding the nature of reality (ontology) (Teddlie and Tashakkori, 2009).

Pragmatism is the philosophical approach used in this study. The purpose of this research is to study, understand, explore and clarify the notion of quality in the ontology domain. It also wants to identify what ontology quality depends on and determines the criteria used by ontologists and knowledge engineers when evaluating an ontology for selection and reuse purposes. This research is asking different kinds of questions, and therefore, needs to use different approaches and different types of data to find answers to those questions.

3.2 Research Purpose

Saunders, Lewis and Thornhill (2012) have identified four main objectives of conducting research: exploratory, descriptive, explanatory or some combination of them. Exploratory studies are used when the precise nature of the problem is unknown, and the researcher aims to understand what is happening. Descriptive research, alternatively, aims to “gain an accurate profile of events, persons or situations” and can be used as a part of both exploratory and explanatory research. Explanatory research aims to identify and establish casual relationships between different variables in research.

The first phase of this study took an exploratory research approach, aiming to understand the process of ontology evaluation and selection for reuse and to identify and explore the main used criteria in that process. The second phase of this study was an explanatory research, built upon the findings of the first stage, and aimed to identify the importance of different metrics used in the evaluation process and to find out if there are any existing relationships amongst the criteria used in the selection process and other independent variables, such as the ontology domain. The final stage of this study aimed to validate and confirm the findings of the first two phases by interviewing experts in the ontology domain.

3.3 Research Approach

There are two main research approaches identified in the literature: inductive and deductive. Research purpose dictates the used approach. The deductive approach, for example, is used where a conceptual framework exists, and the role of the researcher is to clarify the relationship between a particular proposition and previous research (Rudestam and Newton, 2014). The inductive approach, in contrast, is proposed for those areas of research where “there is not enough former knowledge about the phenomenon or if this knowledge is fragmented” (Elo and Kyngäs, 2008). Moreover, the inductive process is usually associated with qualitative research methods whereas the deductive approach is used more with quantitative data.

Different cycles and frameworks have been proposed in the literature to explain the inductive and deductive approaches. In the research wheel (Figure 3-1), for example, research is considered as a recursive cycle of steps rather than a linear process, and it usually starts from some form of empirical observation (Rudestam and Newton, 2007). Teddlie and Tashakkori (2009, p.26) also proposed an inductive-deductive research cycle for mixed methods studies and claimed that any research question or hypothesis at any point in time can be explained using their proposed cycle. According to this cycle, inductive reasoning can be applied to observations and lead to some general inferences e.g., a theory which can then be deductively tested to predict some particular hypothesis (Teddlie and Tashakkori, 2009, p. 26).

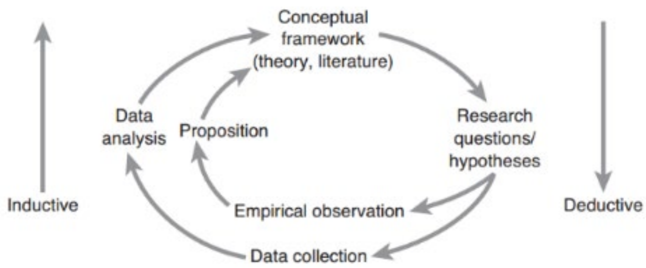

Figure 3-1 Research Wheel Adapted from (Rudestam and Newton, 2007, p.5)

The first phase of this study inductively explored the ontologists' and knowledge engineers' perceptions of ontology quality and the factors it depends on. The deductive approach, however, was mostly used in the second phase to find answers to different research questions and to test the hypotheses. The findings of the first two phases led to proposing a framework for ontology evaluation. The usefulness of this framework was deductively validated in the third phase.

3.4 Method and Research Design

Crotty (1998, p. 3) defines methods as the "techniques or procedures used to gather and analyse data related to some research question or hypothesis". Research method aims to clarify how different parts of research, namely, its design, sampling, data collection and analysis and also the interpretation of the findings are implemented and conducted (Teddlie and Tashakkori, 2009, p. 21). There are different ways of classifying and explaining methodological choices and methods that are used in research. According to Saunders, Lewis and Thornhill (2012), researchers might choose to use a single data collection technique and its corresponding analytical procedure (mono method) or may apply more than one method to address their research questions (multiple methods). When using multiple methods, researchers may decide to use multiples of either quantitative or qualitative research strategies, known as multimethod (Creswell and Plano Clark, 2007, p.12), or may decide to mix and integrate qualitative and quantitative methods, which is known as mixed methods.

Either way, all the methodological choices are based on one or the combination of qualitative and quantitative methods. Quantitative methods were originally developed in natural science and aim to use numerical or statistical data and mathematical modelling to study natural phenomena (Myers, 1997). Creswell (2014) defined quantitative research as an approach for testing objective theories by examining the relationship among measurable variables. The quantitative method has also been defined as the "techniques associated with the gathering, analysis, interpretation, and presentation of numerical information" (Teddlie and Tashakkori, 2009, p. 343). Quantitative research usually follows positivism/postpositivism philosophy and data gathered by this method is usually analysed using statistical methods (Teddlie and Tashakkori, 2009, p. 5).

The qualitative method was originally developed in social science and enabled researchers to study social and cultural phenomena. For Creswell (2014), qualitative research referred to an approach for exploring and understanding the meaning individuals or groups assigned to a social or human problem. In contrast to the quantitative method, which was concerned with numerical data, the qualitative method has been defined as "the techniques associated with the gathering, analysis, interpretation, and presentation of narrative information" (Teddlie and Tashakkori, 2009, p. 343). Qualitative research is usually based on constructivism and the data gathered by this method is usually analysed by inductive and iteration techniques (Teddlie and Tashakkori, 2009, p. 6).

Some argue that combining quantitative and qualitative approaches provides a more complete understanding of a research problem compared with using either of them alone (Creswell, 2013). Therefore, mixed methods research has been proposed as the third major research paradigm and has been defined as "research in which the investigator collects and analyses data, integrates the findings, and draws inferences using both qualitative and quantitative approaches or methods in a single study or programme of inquiry" (Tashakkori and Creswell, 2007, p.4). Mixed methods research usually follows pragmatism viewpoints, meaning that different types of data are collected to understand the research problem in the best way and truth is what works at the time (Creswell, 2009).

There are many different reasons for using and mixing more than one method in a study, such as, facilitation, complementarity, interpretation, generalisability, diversity, problem-solving, focus, triangulation, and confidence (Saunders, Lewis and Thornhill, 2004, p.169). Creswell and Plano Clark (2007, p.34) suggested that mixed methods design can be used when qualitative exploration is not enough, or quantitative results cannot explain the outcome and when a study can be enhanced by using a second source of data. Moreover, using a mixed methods strategy enables researchers to elaborate or develop analysis and also to expand the scope and breadth of the study (Miles and Huberman, 1994).

Mixed methods should also be applied to cases where there is more than one type of method or design that suits the questions and problems that a study tries to address. Creswell and Plano Clark (2007) have provided a list of research problems and the type of method or design that match those problems. Survey design, for example, is suggested where the researcher wants to identify broad trends in a population and correlation design is suggested where the researcher needs to find out what factors influence an outcome.

The questions of this study were answered by using more than one method. Therefore, both qualitative and quantitative research methods were adopted. Qualitative data was initially collected to explore the process of ontology selection for reuse and to clarify the notion of quality in the ontology domain. The findings of the first round were very helpful and led to the identification of some metrics that were not previously mentioned in the literature. Thus, the researcher had to use a second source of data to elaborate, enhance, clarify, confirm and generalize them. Thus, a survey questionnaire was designed, and quantitative data was collected and helped to explain and to elaborate on the qualitative result obtained in the first phase. To finish, the third phase was used to validate and finalise the findings of the previous phases.

3.5 Strategies of Inquiry

After choosing the type of study that best addresses the questions or hypotheses of research, the next step is to choose how the data should be collected. Creswell (2009, p. 11) defined strategies of inquiry as the “types of qualitative, quantitative, and mixed methods designs or models that provide specific direction for procedures in a research". The convergent parallel mixed methods, explanatory sequential mixed methods, and exploratory sequential mixed methods are among some of the very well-known designs proposed in the mixed methods fields (Creswell, 2013, p.15). Convergent parallel procedure, also known as parallel mixed design (Teddlie and Tashakkori, 2009, p.26), refers to cases where both qualitative and quantitative data are collected at roughly the same time and are then integrated in the interpretation phase.

Exploratory sequential mixed methods was used for the purpose of this study. Creswell (2009, p. 14) stated that in a sequential procedure, "the researcher seeks to elaborate on or expand on the finding of one method with another method". Sequential mixed methods research consists of more than one phase of data collection and analysis and one method is followed by the other (Saunders, Lewis and Thornhill, 2012, p. 166). This study started by collecting qualitative data via interviews for exploratory purposes and was followed with a quantitative survey, to confirm and generalize the results of the qualitative interviews, and another set of interviews to validate the findings of both previous phases.

The alternative approach would start with testing hypotheses quantitatively and then exploring and following it up by collecting qualitative data (Creswell, 2014, p.15). This approach was inappropriate for this study because unlike most of the previous studies in the ontology evaluation domain, that start by proposing a set of predefined metrics for ontology evaluation and selection, this research was to inductively identify the set of features that ontologists and knowledge engineers tend to assess while selecting an ontology for reuse.

Apart from the strategies of inquiry in mixed methods and according to the strategies layer of the "research onion" (Saunders, Lewis and Thornhill, 2004), different data collection methods, e.g., experiment, survey, case study, action research, ground theory, ethnography, and archival research can be used in qualitative and quantitative phases. Some of the above-mentioned strategies, such action research, archival research, narrative research and grounded theory were rejected, as they were not suitable for investigating the ontology evaluation and selection process. Ethnography was first considered as one of the options but was then rejected, as it would involve the very time consuming (if not impractical) task of observing ontologists and knowledge engineers while building and selecting ontologies for reuse.

Interviews were used as the strategy of inquiry in the first and the final phase of this mixed methods research. The interviewing method is very powerful in understanding the human situation and is considered as “a key venue for exploring the ways in which subjects experience and understand their world” (Kvale, 2008). Interviews can broadly be classified into two main groups, namely, close-ended and open-ended interviews. Open-ended interviews are used more frequently and lead to the generation of a large amount of information. An alternative approach could have been to use focus groups. This strategy was rejected, because participants in this research were based in different geographical locations and bringing them together would have not been possible.

Patton (1990) has identified three basic open-ended interview approaches, including informal conversational interview, general interview guide approach and the standardized open-ended interview. Standardized open-ended interviews were used in the first and the final phase of this study, meaning that all the interviewees were asked an identical set of questions; however, questions were worded in an open-ended format, which urged the respondents to share as much details as they wanted (Teddlie and Tashakkori, 2009; Turner III, 2010).

The second phase of the mixed methods used in this study was based on a survey questionnaire. The survey provides “a quantitative or numeric description of trends, attitudes, or opinions of a population by studying a sample of that population” (Creswell, 2009). Survey strategy aims to explore a research problem and is usually associated with the use of a questionnaire (Saunders, Lewis and Thornhill, 2012, p.177). A mixed methods questionnaire, including both closed-ended questions, e.g. Likert scales and open-ended qualitative questions, was used as the technique in the second phase of this research (Teddlie and Tashakkori, 2009, p.235).

3.6 Characteristics of Mixed Methods in this Research

In the previous sections, the reasons for choosing mixed methods and inquiry strategies used in this research were explained. The following part of this section aims to clarify the type of the mixed methods strategy. Teddlie and Tashakkori (2009) have presented different criteria that should be considered when deciding about the type of mixed methods strategy for research. Some of those criteria are as follows:

1. **Type of Implementation Process.** Also known as timing or sequence (Creswell and Plano Clark, 2007); this criterion clarifies if qualitative and quantitative data are collected sequentially or in parallel. In this study, the real meaning of ontology quality and the metrics that it depends on was not that clear and therefore, the researcher had

to first learn what ontology quality meant to different ontologists and knowledge engineers and how it is evaluated before being able to test different hypotheses. So, different phases of this research could not be conducted concurrently, and the sequential implementation process was applied.

2. **The Priority of Methodological Approaches.** This criterion identifies the importance of different approaches and whether the qualitative or quantitative component has any priority or importance. In other words, this criterion determines which of the methods used in a study is more important in understanding the phenomenon of the research and is often decided after the completion of a study. In this research, all of the methods used in the data collection helped equality in understanding the notion of ontology quality and what it depends on. Therefore, they all had the same priority.
3. **Stage of Integration of Approaches.** This criterion identifies in what stage(s) (when) will the study be mixed and also, how will mixing occur. Different data sets can be connected, merged or embedded (Creswell, 2014). This is one of the most important decisions that should be made while designing mixed methods research. In this study, the qualitative and quantitative phases were connected to each other as one of them led to the other. Some qualitative questions were also embedded in the quantitative phase of this research.

3.7 Choosing a Time Horizon

One of the inner circles of the research onion is time horizon and is concerned with the timeframe and duration of the research study. Saunders, Lewis and Thornhill (2012, p.190) suggested that researchers should identify if they want their research to be a "snapshot taken at a particular time", also known as cross-sectional study, or if they want it to be more "akin to a diary or a series of snapshots" over a given period, also known as longitudinal study. The data collection process in this study was limited to a specific timeframe and aimed to identify what quality meant to different ontologists and knowledge engineers, not how it changes over time. Hence, this research used a cross-sectional time horizon.

3.8 Scale (Measure) Development Strategy

One of the earliest procedures for measurement development was proposed by Churchill (1979) and suggested different steps, namely, specifying the domain of construct, generating a sample of items, collecting data, purifying measure, collecting data, assessing reliability, assessing the validity and developing norms in the process of scale development. Gerbing and Anderson (1988) suggested confirmatory factor analysis as an update to the procedure proposed by Churchill (1979). This study has followed one of the most recent guidelines for scale development, which is the one proposed by DeVellis (2012). This section moves on to provide an overview of the design of the questionnaire used in the second stage of this research.

Step 1. The first step in the scale development is to clarify what exactly the scale aims to measure. In this study, the survey questionnaire aimed to identify the quality metrics that are mostly used by those involved in the process of ontology evaluation and selection for reuse. The other goal was to find if there is any correlation between the years of experience someone has got and the domain and organisation they work in, with the quality metrics they use in the evaluation and selection process.

Step 2. After identifying the purpose of a scale, the next step is to generate what DeVellis (2012, p. 76) called "item pool". In this step, the items that can be included in the survey in order to be able to measure what the survey aims to measure are identified. Items can be generated either inductively or deductively. In an inductive approach, domain experts are usually asked to provide their thoughts on the subject, while in a deductive approach, theoretical definitions are used in the generation process (Hinkin, Tracey and Enz, 1997). Interviews and/or focus groups and literature review are the most widely used methods in the inductive and deductive approach respectively (Morgado *et al.*, 2017). In this study, the quality metrics identified in the interviews and also in the literature of ontology evaluation were used as the items of the questionnaire.

To ensure that the proposed items are properly constructed, Hinkin, Tracey and Enz (1997) suggested some basic guidelines that should be followed. First and to prevent confusion on the respondents' part, they argued that each item should only address a single issue. Second, researchers should ensure that all the items are kept consistent and are assessing the same type of behaviour or information. Third, it was suggested that statements should be kept short and simple and use a language that is familiar to the target audience. Finally, to get meaningful responses, it was argued that researchers need to make sure that respondents understand the items as intended by them. DeVellis (2012, p. 77) has also stated that good items are consistent, clear, unambiguous, and not lengthy.

Each item in this study covered one aspect of the ontology quality and assessed the same type of information. One of the challenges in developing the questionnaire was the choice of words. Ontologies are utilized by a large group of people in different domains and with different levels of expertise, who tend to use different terminology to refer to similar concepts (Suárez-Figueroa and Gómez-Pérez, 2008). To bring clarity to this survey, a brief description or examples of the intended meaning was provided for different items in the questionnaire. The adequacy of content and classification of items were assessed both by the previous literature and the experts in the ontology domain.

Step 3. The third step, that can also happen simultaneously with the second step, is to identify the formats for measurements. A Likert scale was widely used in the survey questionnaire of this study to measure how important the respondents thought different quality metrics were in the evaluation process. They were also used to measure the level of agreement amongst respondents about different community related statements.

Step 4. After identifying and developing a scale, the next step is to assess its validity by seeking opinions of expert judges or target population judges (Morgado *et al.*, 2017). It includes asking them how relevant and appropriate they think each item is to what the research wants to measure, how clear and concise they think each item is, or asking them if they think the researcher has failed to include an item (DeVellis, 2012). To do that, a pilot study was conducted, and the survey was reviewed by some experts in the ontology domain.

Step 5. This step aims to add and include some items that can be used for validation. It includes questions from previous constructs in the same field or the questions that can be used to identify response biases.

Step 6. The aim of this step is to identify a large and representative sample for the research. The sample size is one of the most important and also debated issues that will affect the validity of a construct. Larger samples are highly recommended in the literature but there is also uncertainty about "how large is large enough?" (DeVellis, 2012). Hinkin, Tracey and Enz (1997) argued that the number of sample size is dictated by the number of variables or items that an instrument is assessing. According to the literature, item-to-response ratio can range from 1:4 to 1:10; however, Hinkin, Tracey and Enz (1997) suggested that a sample size of 150 and 100 should be sufficient for Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA), respectively. A total number of 157 complete responses were collected in this phase of the study.

Step 7. The last step in scale development is to assess if "the new scale has constructed validity and reliability" (Morgado *et al.*, 2017). DeVellis (2012, p. 59) defined a reliable instrument as the one which performs in a consistent and predictable way and argued that "reliability concerns how much a variable influences a set of items" whereas "validity concerns whether the variable is the underlying cause of item covariation". According to the literature, different tests, namely, internal consistency and inter-observer reliability can be used to measure the reliability of an instrument (Morgado *et al.*, 2017). EFA and CFA are usually used to assess the validity of an instrument.

3.9 Sampling Strategies

Identifying a sampling strategy is one of the most important stages in any research, or as Tashakkori and Teddlie (2003, p. 275) argued, "sampling is destiny". Sampling is more complicated in mixed methods research, where the aim is to collect different types of data. There are different issues that one should consider before choosing the sampling strategy, such as making sure that the chosen strategy is ethical and matches the research questions and the conceptual framework, and also making sure that the sample will allow "for credible explanation" and "the possibility of drawing clear inferences from the data" (Tashakkori and Teddlie, 2003, p. 276). Moreover, the plan for sampling should be feasible and let the research team generalize the research findings to other populations (Tashakkori and Teddlie, 2003, p. 276). Budget and time are amongst the other factors that should be considered when selecting a strategy (Saunders, Lewis and Thornhill, 2012, p. 260).

Different sampling strategies, e.g., probability sampling, purposive sampling, convenience sampling, and mixed methods sampling have been proposed in the literature (Teddlie and Tashakkori, 2009, p. 170). Probability and non-probability sampling are the two broad categories of sampling strategies (Saunders, Lewis and Thornhill, 2012, p. 261). From all the above-mentioned strategies, probability and purposive sampling are more popular and are primarily used in quantitative-oriented and qualitative-oriented studies, respectively. Despite their differences, these sampling strategies share two main characteristics: (1) both aim to find an answer to the research questions and (2) both are concerned with generalizability (Teddlie and Tashakkori, 2009, p. 178). The rest of this section describes different sampling strategies.

3.9.1 Purposive Sampling

Purposive sampling, also known as non-probability or judgement sampling (Patton, 2015), is one of the main strategies associated with qualitative data collection. Purposive sampling aims to strategically identify some cases or participants that are relevant to the research questions (Bryman, 2016). Patton (1990) argued that "the logic and power of purposeful sampling lies in selecting in information-rich cases for study in depth" and suggested 15 different strategies that can be used for purposive sampling, including extreme or deviant case sampling, snowball or chain sampling, and criterion sampling. Techniques used in purposive sampling usually have two main aims: to generate representative cases or to produce contrasting cases (Teddlie and Tashakkori, 2009, p. 176).

3.9.2 Probability Sampling

This strategy is usually used in quantitative research and aims to identify and randomly choose individuals that are representative of a population (Creswell and Plano Clark, 2007,p.112). The end goal of a research based on this strategy is to be able to generalize the finding from the population of the study to a larger population (Tashakkori and Teddlie, 2003, p. 277). The process of probability sampling consists of four stages, namely, identifying a sampling frame, deciding on a suitable sample size, selecting sampling techniques, and checking if the sample is representative of the population (Saunders, Lewis and Thornhill, 2012, p. 277). Sampling techniques suggested for this strategy include simple random, systematic random, stratified random, cluster, multiphase or multi-stage (Acharya *et al.*, 2013).

3.9.3 Mixed Methods Sampling

Sampling in mixed methods research may involve the use of more than one sampling strategy. Similar to both of the above-mentioned strategies, the mixed methods sampling aims to generate a sample that will find answers to the questions under study and is also concerned with the issue of generalizability (Teddlie and Tashakkori, 2009, p. 181). Mixed methods sampling may simultaneously use purposive sampling techniques to increase inference quality and probability sampling to increase transferability and generalizability (Tashakkori and Teddlie, 2003, p. 284). Teddlie and Tashakkori (2009, p. 170) have suggested five main techniques for mixed methods sampling, including basic mixed methods sampling, parallel mixed methods sampling, sequential mixed methods sampling, multilevel mixed methods sampling, and combination of the above-mentioned techniques.

Before selecting the sample for this study, the researcher had to first come up with a set of characteristics or inclusion criteria that were important to be represented in the sample and then to identify the sample that would meet and satisfy those characteristics. Questions of this research could best be answered by ontologists and knowledge engineers that had not only been involved in the process of ontology selection but those who had also considered ontology reuse and had evaluated different ontologies before selecting them for reuse. Moreover, to identify characteristics of the reusable ontologies, the researcher wanted to identify the developer(s) of those ontologies that had already been reused and find the set of steps or principles they had followed to develop a reusable ontology.

3.9.4 Sampling in this Study

This research followed a sequential mixed methods design. Purposive sampling was the only strategy used in this research. Sampling in this study started by applying homogenous purposive sampling and aimed to identify a group of ontologists and knowledge engineers that were or had been involved in the process of evaluating, selecting and reusing ontologies. To do that, different ontology repositories, like BioPortal, were explored and a set of ontologies that have previously been reused was identified; people who had developed and/or reused those ontologies were then contacted.

Mixed methods research is usually associated with using both probability and non-probability sampling strategies. However, according to Teddlie and Yu (2007), one of the techniques alone, either probability or non-probability, is appropriate for some research. While non-probability sampling is often linked with qualitative research, Bernard (2017, p. 145) argued that non-probability samples are also appropriate for large surveys, when the aim is to collect data from expert informants. In other word, non-probability samples can be used when the aim is to conduct research by collecting data from informed informants and not just responsive respondents (ibid.). Thus, based on the research aim, research questions, and inclusion criteria, purposive sampling was used for the second phase of this research, with the aim of finding a larger population of experts in the ontology domain.

The survey was sent to the community of ontologists and knowledge engineers in different domains; similar inclusion criteria were used in the second phase. Besides going through ontology repositories and libraries, different research groups in universities and organisations were explored to identify the experts that were involved in the process of ontology development and reuse. The survey was also forwarded to different active mailing lists in the field of ontology engineering. Some of the mailing lists used are as follows:

- The UK Ontology Network (ontology-uk@googlegroups.com)
- GO-Discuss (go-discuss@lists.stanford.edu)
- DBpedia-discussion (dbpedia-discussion@lists.sourceforge.net)
- The Protégé User (protege-user-bounces@lists.stanford.edu)
- FGED-discuss (fged-discuss-bounces@fged.org)
- Linked Data for Language Technology Community Group (public-ld4lt@w3.org)
- Best Practices for Multilingual Linked Open Data Community Group (public-bpmlod@w3.org)
- Ontology-Lexica Community Group (public-ontolex@w3.org)
- Linking Open Data project (public-lod@w3.org)
- Ontology Lookup Service announce (ols-announce@ebi.ac.uk)
- Technical discussion of the OWL Working Group (public-owl-wg@w3.org)
- This is the mailing list for the Semantic Web Health Care and Life Sciences Community Group (public-semweb-lifesci@w3.org)

The aim of the final phase of this study was to validate the findings of the previous phases. Expert sampling was used in this phase (Etikan, Musa and Alkassim, 2016) and led to the identification of some of the key informants in the ontology domain. This type of sampling was very helpful because most of the findings of the previous phases, especially the social-related features, were novel and had not been discussed previously. Therefore, it was important to know what the experts in the domain think about them.

3.9.5 Sampling Size

Deciding on a suitable sample size is one of the most important issues to address when selecting the sampling strategy of research. There has been a lot of discussions on sample size. As it is seen in the literature, the sample size is influenced by the type of research and the sampling strategy used in it. Teddlie and Tashakkori (2009, p. 179) argued that sample size in probability sampling needs to be large enough, at least 50 units, so that it can be used to establish representativeness. Sampling size in purposive sampling, however, is typically small and usually less than 30 units (ibid.). Creswell and Clark (2017, p.123) also argued that the sample size in a qualitative study is much smaller than a quantitative data collection and stated that sequential research designs usually have unequal sample sizes.

Sampling size is more complicated in purposive sampling, where there exist no rules about the right number of participants (Saunders, Lewis and Thornhill, 2012, p. 283). In purposive sampling, it is suggested to continue collecting data until "data saturation is reached" (Guest, Bunce and Johnson, 2006; Saunders, Lewis and Thornhill, 2012, p. 283). Guest, Bunce and Johnson (2006) argued that in research with the main aim of understating commonalities within a homogenous group, the saturation occurs within the first twelve interviews. Symon and Cassell (2012, p. 45) also claimed that the minimum non-probability sample size can range from 4 to 36, depending on the nature of the study.

The first phase of this research consisted of two parts. Initially, five pilot interviews were conducted to test the wording of the interview questions and to detect any potential ambiguities as well as the flow of them. Convenience sampling was used in the pilot phase, and the five participants were chosen from the ontologists working in the School of Business and Economics, Loughborough University. They all had previous experiences of developing and reusing ontologies. The pilot phase gave the researcher a good chance to improve her interview skills and time management. Afterwards, an invitation email was sent to 34 ontologists and knowledge engineers who had previous experience of developing and selecting ontologies for reuse; 15 of them accepted the request and participated in the interview study.

The sample size of the first phase was sufficient for different reasons. Firstly, interviews were conducted until no new information or theme was found (Guest, Bunce and Johnson, 2006), and the conceptual saturation was reached. Secondly, and according to Symon and Cassell (2012), anything between 4 to 36 is considered as an acceptable sample size in a non-purposive sampling strategy. Thirdly, some of the well-known studies in this domain, like the survey conducted by Lozano Tello (2002), had fewer responses (only 10). Finally, this was not the only phase of data collection in this study and the findings of this research are based on data collected from the largest group of experts in the ontology domain.

Different sampling strategies were used in the second phase to identify a larger population of ontologists and knowledge engineers. A link to the survey was sent to more than 500 people, including the participants in the first round, as well as 12 mailing lists. A total number of 314 people clicked on the link to the survey, and 157 of them completed the survey. Like the first phase, before sending the survey out, the researcher discussed the wording of survey questions, questions' types, and survey designs with a group of experts in the ontology domain and made some adjustments accordingly.

Eight ontology experts were interviewed in the third phase. The sample size is small but is normal in expert sampling (Trotter, 2012). This phase was very helpful in clarifying and validating the newly identified quality metrics and the framework.

3.9.6 Sampling Issues with an Online Survey

Bryman (2016, p. 191) have identified different problems that one might face while conducting an online survey, like the fact that many people have more than one email address. While contacting people in academia, the other main issue was that people changed their workplaces and organisations and many of the email addresses found online were invalid. Finding participants might be more difficult in some domains, like ontology engineering, compared to the others. However, this research was very successful in identifying participants; one of the unique characteristics of this study is that it has the largest population compared to the previous studies in this domain (Lozano-Tello, 2002; Matentzoglu *et al.*, 2018).

3.10 Data Collection

As was mentioned earlier, sequential data collection was used for the purpose of this research. Data can be collected using different primary and secondary data collection methods. Hox and Boeije (2005) defined primary data as "original data collected for a specific goal" and secondary data as "data originally collected for a different purpose and reused for another research question". Observation, interviews, and surveys are among the most well-known primary data collection methods. Literature review and official data archives can be considered as the main sources of secondary data (ibid.).

3.10.1 Literature Review

Reviewing prior and relevant literature is one of the first and most important steps in conducting any research project. A literature review builds an understanding of theoretical concepts; it supports the identification of a research topic, gap, and the areas that are beneficial to research (Rowley and Slack, 2004). Ontology evaluation and selection for reuse is a very complicated task and depends on many different factors. Thus, the following topics were reviewed:

- The general notion of ontology, ontology selection, and ontology reuse.
- Different selection systems in the ontology domain.
- The evaluation methods, metrics, and frameworks used in the literature to address different challenges in the ontology selection domain.

3.10.2 First Phase: Qualitative Data Collection

A group of ontologists and knowledge engineers with different levels of expertise and backgrounds in building and reusing ontologies were contacted, and 15 of them accepted to participate in the first phase interviews. Participants in this study were working in domains; four of them had only worked on biomedical ontologies, five had some biomedical experience but had also worked in other fields, such as computer science; the rest of the interviewees were mostly involved in developing ontologies in manufacturing, smart cities, oil, and other non-biomedical domains.

The semi-structured interview protocol focused on how each individual (i) built, (ii) searched for, (iii) evaluated and (iv) reused ontologies. Interviews ranged from 20 to 60 minutes and all, except one of them, were done over Skype as the interviewees were based in different geographical locations. One face-to-face interview was also conducted with an expert in the UK. Interviewees were first informed about the purpose of the study and were asked if they could be recorded. After obtaining consent, interviews were recorded. The interviewer also took field notes during the interviews. Procedures suggested by Creswell and Plano Clark (2007) were followed to analyse the data, starting from data preparation (e.g., transcribing), data exploration, using (QSR International, 2015) for coding, and finally representing and discussing the identified themes or categories.

3.10.3 Second Phase: Quantitative Data Collection

A questionnaire was designed with the total number of 31 questions, broadly divided into four different sections. Each section consisted of different number of questions and aimed to explore and discover the opinion of ontologists and knowledge engineers regarding (1) the process of ontology development, (2) ontology reuse, (3) ontology evaluation and the quality metrics used in that process, and (4) the role of community in ontology development, evaluation and selection for reuse. Screening questions were used throughout the survey to ensure that respondents are presented with the right set of questions and the answers are valid. None of the participants were discarded from the survey based on their answers to the screening questions, but they were presented with a different set of questions that best suited their previous experiences.

The first screening question was used to discover how often survey respondents build ontologies. This question aimed to make sure that all the respondents were involved in the process of ontology development. The second screening question checked how often respondents consider reusing ontologies. If the respondents had never reused an ontology, they would be presented with a question that would ask them whether they had ever evaluated an ontology. At the end of the survey, five demographic questions were asked to learn about the respondents' job title, the type of organisation they worked for, how experienced they were, the main domains they had built or reused ontologies in and the primary language they used for ontology development.

The second part of the survey focused on ontology reuse and started by asking respondents about how often (never to always) they consider reusing ontologies. If they chose anything other than never, meaning that they had some experiences of reusing ontologies, they were presented with a list of search and selection systems for ontologies and a 5-point Likert scale, ranging from never to always; they were asked how often they use the suggested search and selection systems to find an ontology for reuse. A comment field was also provided for mentioning the other search engines or repositories participants would use. Respondents were also asked open-ended questions about (1) the main challenges they had faced while searching for a reusable ontology and (2) the best ontology they had reused and why they called it the best? If respondents selected never in answer to the question "How often do you consider reusing existing ontologies?", they would be asked to share the reasons why.

The most important part of this survey aimed to investigate the process of ontology evaluation and the set of criteria that can be used in that process. In this section, respondents were first asked about the approaches and metrics they tend to consider while evaluating ontologies; this part aimed to explore the respondents' views and opinions. They were then presented with four different sets of quality metrics, gathered both from the literature and the first phase of the data collection, and were asked how important they thought each of those metrics was, by offering a 5-point Likert scale, ranging from “Not important” to “Very important”. Those four categories were:

1. Internal aspects of ontologies that can be used in the evaluation process, including their scope, content, and structure.
2. Metadata or different pieces of additional information that can be used for ontology evaluation like documentation, language, accessibility, and frequency of updates.
3. Community related metrics, such as community activeness and responsiveness, and reputation of the ontology developer team and/or organisation.
4. Popularity related metrics, namely, the number of times an ontology has been reused and the popularity of ontology in the community.

The quality metrics presented in this section were defined and clarified using a brief description or some examples. The reason for proving these descriptions was because the researcher wanted to make sure that respondents knew what each of the metrics exactly meant and referred to. For instance, the metric “language that ontology is built in” had a short example “e.g., OWL”, which would mean that the question is asking about the programming language, and not the natural language that an ontology is built in.

The last section of the survey presented the respondents with some of the interesting statements mentioned in the first phase about the role of community and how it can affect the process of ontology evaluation and selection for reuse. A 5-point Likert scale ranging from “Strongly Disagree” to “Strongly Agree” was used in this section to collect the opinions of ontologists and knowledge engineers on this matter and to measure the level of agreement/disagreement amongst them.

There are many different ways of conducting an online survey, including embedding the survey questions in an email, sending a questionnaire or a survey program as an email attachment or emailing a link to the survey to the respondents (Gunter *et al.*, 2002). In this research, a link to the survey was emailed to different potential respondents. The survey was designed and tested using three different online tools, but the one designed in “Qualtrics”[17] was chosen at the end, as it would provide a better layout and would support different types of questions. There are many different advantages for conducting online surveys, such as, global reach, low cost, convenience, flexibility and ease of analysis (Evans and Mathur, 2005; McPeake, Bateson and

[17] https://www.qualtrics.com

O'Neill, 2014). Global reach and convenience were two of the main reasons for conducting an online survey in this study.

3.10.4 Third Phase: Qualitative Data Collection

The third phase of the data collection aimed to validate the findings of the previous phases. It involved two different experiments. In the first one, the identified metrics were applied to a set of ontologies in the NCBO BioPortal to test if they can predict the number of times ontologies get (re)used. The second one was a user centred experiment (Mandran and Dupuy-Chessa, 2018); eight experts in the ontology domain were interviewed using Skype. This experiment aimed to determine how well the findings of this study can predict the ontology that knowledge engineers would likely select for reuse. This experiment also investigated whether ontologists and knowledge engineers find having information about the factors proposed in this study helpful in the selection process.

3.11 Data Analysis Methods

This study followed a sequential mixed method design and therefore, the data collected in each phase of this study was analysed before conducting the next phase. The analysis process started with identifying themes in the qualitative data. Those themes and findings were then used to create the survey instrument. Finally, the survey results were used in the ranking process of the validation phase. Due to the different nature of the collected data, different methods were used in this study. The following part of this section presents the data analysis process in detail.

3.11.1 First Phase: Qualitative Data Analysis

NVivo (QSR International, 2015) was the main application used for the primary analysis of the qualitative data collected in the first phase of this research. The first set of codes was developed before analysing the data; it was structured according to the main research themes and to find answers to different research questions that were about:

- Ontology development process
- Ontology search and selection for reuse
- Ontology evaluation and quality metrics

Other codes were then added to cover the frequently mentioned themes, e.g., the role of the community in the ontology evaluation and selection process.

As it was seen in the first round of analysis, the most interesting and mentioned part of the interviews' transcriptions was about the way ontologies are evaluated and selected for reuse. So, the second round of analysis focused on summarization of all the information available about ontology quality and selection metrics. During this round, the researcher extracted what each of the interviewees had mentioned about the quality of an ontology, selecting an ontology, and different characteristics of a reusable ontology. This information was then categorised in the following groups and was used as the basis for the second round of the data collection.

- Building a reusable ontology
- Characteristics of a reusable ontology
- Finding a reusable ontology
- Evaluating/trusting /selecting ontologies
- The importance of community

3.11.2 Second Phase: Quantitative Data Analysis

The survey was conducted over a five-month period from September 2017 to early January 2018. The data analysis started by removing the incomplete responses. A total number of 314 respondents had clicked on the link to the survey, but only 162 of them had completed (100% progress) and submitted the survey. After analysing the data more closely, four of the responses were eliminated as they had not provided valid information for some of the very important questions in the survey (e.g., had used invalid characters). Survey data was then analysed using both R (R Core Team, 2013) and SPSS (IBM, 2016).

3.11.3 Third Phase: Qualitative Data Analysis

Similar to the first phase, NVivo (QSR International, 2015) was used to analyse the qualitative data collected in the second round of the interviews. The research questions of this phase of the study were mainly about the usefulness of the metrics identified in the previous phases. Ranking comparison and metric usefulness codes were used to analyse the interviews' transcriptions. Another code was then added to cover and categorise what participants had to say about the characteristics and usefulness of the current selection systems for ontologies.

3.12 Data integration

Fetters, Curry and Creswell (2013) have identified three different levels of integration in mixed methods research, including integration at the design level, integration at the methods level and integration at the interpretation and reporting level. At the method level, this study followed the building approach, meaning that the items included in the survey used in the second phase of this study were built based on the interview findings. Different approaches were used at the interpretation and reporting level. The qualitative and quantitative findings are presented in different chapters (staged approach); however, in some chapters, the qualitative and quantitative findings are reported together (weaving approach).

3.13 Ethics

Different ethical issues need to be considered when doing research. According to Saunders, Lewis and Thornhill (2012, p.277), ethics refers to "the standards of behaviour that guide your conduct in relation to the rights of those who become the subject of your work, or are affected by it". Informed consent is one of the main issues to consider when doing research. All the data for this study was collected online. Saunders, Lewis and Thornhill (2012, p.234) suggests that in an online setting, consent should be obtained by specifically asking the participants. Participants in different phases of this study were fully briefed on the purpose and process of the research and were asked if they would like to participate.

As was mentioned before, the interviews for this study were recorded for transcription and analysis purposes; before starting each interview, the researcher would inform the participants about the recording and ask for their consent. Moreover, participation in this study was voluntary; the researcher sent email requests to the potential participants, and they could choose to participate or not, or even withdraw after participating. In the second phase, for example, many people clicked on the link to the survey, but did not complete it. The researcher removed all the incomplete responses and only kept the ones with 100% progress.

The data collected in different phases of this study does not contain any sensitive information of the participants, and therefore, there should be no risk involved in different phases of this research. All personal information of the participants in different phases of this study were kept confidential and will remain confidential. Moreover, acronyms were used throughout this thesis to protect the privacy of the participants in different phases. For example, Bi, SBi, NBi were used throughout the report to refer to the participants working in different domains.

The survey questionnaire used in the second phase of this study was very long (31 questions) and its completion hence time-consuming. Therefore, and as a thank you to participants for offering their time, they were given the option to enter a prize draw to win one of three £50 gift vouchers. Entry to the prize draw was optional, and separate from the survey, in order to secure participants' anonymity and to keep their responses confidential. Not every participant decided to enter the draw. Three of those who chose to enter were selected at the end of the second phase and their prizes were emailed to them by the finance team of the School of Business and Economics at Loughborough University.

3.14 Summary

This chapter explains the research methodology used in this study and clarifies the rationale behind the chosen path. Data collection in this study included a literature review, two phases of qualitative interviews and a quantitative survey. This research was very successful in reaching a very large group of ontology experts and knowledge engineers. The findings of different phases of this research addressed all of the research questions and clarified the process of ontology search, evaluation, and selection for reuse. The rest of this thesis presents different phases of data collection in more detail.

Chapter 4: INTERVIEW FINDINGS

4.1 Introduction

The second objective of this research is to address some of the challenges faced in the first steps of the general process of reusing ontologies, which is to evaluate and then select the right, or the most appropriate ontology for reuse. To do that, ontologists and knowledge engineers in different domains were interviewed and asked about the challenges they face while searching, evaluating, and selecting an ontology. This phase of the study aimed to qualitatively understand the process and reasoning behind the decision-making in ontology selection, with a particular focus on the under-researched social and community aspects of ontology quality.

As it is seen in Table 4-1, 15 ontologists and knowledge engineers with different levels of expertise and backgrounds were interviewed. Four of them had only worked in the biomedical domain. Five had some biomedical experience but had also worked in other fields, such as computer science. The remaining six interviewees were mostly involved in developing manufacturing, smart cities, and non-biomedical ontologies. The Bi, SBi, NBi acronyms are used throughout this chapter to refer to the interviewees with biomedical, some biomedical and no biomedical experiences, respectively. This categorisation was applied because ontologies are very popular, very successful and widely used in the biomedical domain.

Table 4-1 Domain Expertise of Participants in the Interviews

Interviewee Code	Role | Domain
Bi1	Academic Staff | Bioinformatics, Gene Ontology
Bi2	Software Engineer and Developer | Biomedical Ontologies
Bi3	Ontology Developer | Bioinformatics, Gene Ontology
Bi4	Researcher | Biomedical Informatics
SBi1	Ontology Developer | Industrial Ontologies, W3C, NHS
SBi2	Researcher | Biomedical and Non-biomedical Ontologies
SBi3	Ontology Developer | Biomedical and Non-biomedical Ontologies
SBi4	Academic Staff | Computer Science and Biology
SBi5	Research Scientist | Protégé, Biomedical and Non-biomedical Ontologies
NBi1	Ontologist | Smarter Planet Project
NBi2	Academic Staff | Manufacturing Informatic
NBi3	Ontology Engineer | Semantic Application Developer
NBi4	Researcher | Applied Ontologies
NBi5	Researcher | Smart Cities, Geo Ontologies
NBi6	Researcher | Industrial ontologies

Semi-structured interviews were used to collect data. Designing the interview questions was one of the most important and also time-consuming parts of this phase of research. As it is seen in Appendix G, different types of questions were asked in the interviews, starting from a descriptive ice-breaking question asking *"could you tell me a bit about yourself and your experiences in ontology domain"* to more structural questions like *"what has been your experience with searching for ontologies?"*, *"Do you reuse ontologies? if yes, how do you find ontologies for reuse?",* and *"what are the main properties that you tend to look at when judging the general quality or suitability of an ontology?"*. Each question also included some detailed sub-questions to make sure that the same path is followed, and the same data is obtained in every interview (see Appendix H).

Every interview was transcribed and then analysed using NVivo (QSR International, 2015). The initial set of codes and themes developed for analysis included "ontology search", "ontology reuse", "ontology evaluation", and "ontology quality". Those codes were evolved during the analysis to cover the frequently mentioned themes by the interviewees, namely, the role of the community in the ontology evaluation. The second round of analysis was conducted and focused on summarization of all the information available about ontology quality and selection criteria. During this round, interviewees' statements about ontology quality, ontology selection, and different characteristics of a reusable ontology were extracted and categorised.

4.2 Findings

The findings from the interview were categorised into two main groups: (1) ontology development process and (2) ontology selection and evaluation for reuse. The first category included what interviewees had to say about the ontology development process and the role of methods and ontology reuse in that process. The second category covered the evaluation and selection process and the set of metrics that are considered while assessing the quality of an ontology and before selecting it for reuse. These categories are discussed in the following section.

4.2.1 Ontology Development Process

One of the very first questions of the interviews was about how ontologies are built. This question was asked to clarify the process of ontology development and the role of ontology reuse in that process. It also aimed to understand how developers approach the notion of "quality" while building an ontology and how they make sure that the ontology they are developing is "good" and "reusable". The answers to this question mainly included different steps that are taken and the information that is provided in the development process, and to make sure that ontologies with good quality are developed.

Firstly, interviewees were asked if they follow any methodology or principle while developing an ontology. Different methods for ontology development are proposed in the literature (Iqbal *et al.*, 2013); therefore, the researcher wanted to determine the role and importance of those methods in the development process. Interestingly, when asked about methods, some of the interviewees had no informed view on what methodology, if any, existed, and asked the interviewer to provide some examples. When provided with examples like METHONTOLOGY (Fernabdez, Gomez-Perez and Juristo, 1997) and NeOn (Suárez-Figueroa, Gómez-Pérez and Fernández-López, 2012), most of the respondents stated that they do not use any methodology while developing ontologies. Some of the interviewees like SBi1, however, argued that following a specific strategy and design principles are essential, especially in cases where different partners are involved, and it is important to ensure that "all are working from the same page".

When asked about quality, internal characteristics of ontologies, such as their scope, consistency and syntax were among the very first aspects to be assessed by the ontology developers. Providing documentation and making the ontology available, were the other steps knowledge engineers would take to make sure that their ontology is reusable. NBi3 also stated that building ontologies by collaborating with their community instead of building them on their own is what they do to ensure that the ontology they are developing is reusable. Moreover, two of the respondents, NBi2 and SBi1, emphasized the importance of reusing other ontologies, especially while defining basic concepts and relationships.

4.2.2 Ontology Quality, Evaluation and Selection

This study aimed to understand how ontologies are evaluated and to identify the metrics that are used in the evaluation and selection process. The interviews' findings suggested that metrics for evaluating the quality of ontologies for reuse can be classified into the following categories:

- Metrics based on different internal components of ontologies, including content, structure, and coverage.
- Metrics related to different metadata about ontologies, such as the methodology used, availability of documentation, language, and size.
- Metrics based on the social aspects of ontologies like community, popularity, ontology developer team and organization.

Table 4-2 provides an overview of the discussed metrics and the interviewees who found them important. The rest of this chapter describes each of the above-mentioned categories in more detail.

Table 4-2 Interviews Summary

Category	Metric	Participants that Supported the Metric
Internal	Scope, coverage and fit	NBi1, NBi6, SBi1, SBi4, SBi5
	Clarity	NBi3, NBi6, SBi4, SBi5
	Consistency	SBi1, SBi5, Bi3
	Structure	SBi1, SBi5, Bi3
	Correctness	SBi3, SBi4
Metadata	Documentation	NBi3, NBi4, NBi6, SBi1, SBi4, SBi5, Bi1, Bi3
	Following standard approaches and principles	NBi1, NBi2, NBi4, SBi1, SBi3, SBi5, Bi3
	Availability (being online) and accessibility	NBi1, NBi3, NBi6, SBi3, Bi3
	Frequency of updates and maintenance	NBi3, NBi4, SBi1, SBi3, SBi4, SBi5, Bi1, Bi3, Bi4
	Availability of publications about an ontology	NBi4, NBi6, SBi3, SBi5, Bi4
	Ontology size and specialization	NBi5, SBi2, Bi2, Bi4
	Reusing other ontologies	NBi3, NBi4, NBi5, NBi6, Bi3
	Ontology development language	NBi4, Bi2
Social	Community judgment and reviews	NBi3, NBi4, NBi5, NBi6, SBi5, Bi1, Bi2
	Reuse related data (e.g., who else has reused the ontology?)	NBi6, SBi3, SBi5, Bi4
	Reputation and responsiveness of the ontology developer team	NBi1, NBi3, SBi3, SBi4, SBi5, Bi1, Bi3
	Popularity and acceptance	NBi3, SBi2, SBi3, SBi5, Bi2, Bi3, Bi4

4.2.2.1 Evaluation based on Internal Aspects of Ontologies

As it is seen in the literature, many of the evaluation methods are based on different internal components of ontologies, such as their content and structure. When asked about evaluating the quality of an ontology, internal aspects were among the first features assessed. SBi3 and SBi4, for example, pointed out that they check the correctness of an ontology's content before selecting it for reuse. Consistency and coverage were amongst the other mentioned aspects of ontologies. NBi1 and SBi5, for example, stated coverage as one of the metrics they assess while evaluating and selecting an ontology.

Being "well-structured" was also mentioned by some of the interviewees as one of the characteristics of a reusable ontology. According to them, a well-structured ontology can be defined as an ontology that has some rich (NBi4) and correct (SBi1) connections between its categories. An ontology's syntax, definitions and scope were amongst the other internal characteristics that could be examined in the evaluation and selection process. Hearing about internal characteristics of ontologies was expected, as many of the previous studies in this domain are based on assessing these features. It is easy to examine some of these internal features, e.g., syntax; however, assessing some factors like correctness and completeness, is very challenging (Yu, Thom and Tam, 2007).

4.2.2.2 Evaluation based on Metadata about Ontologies

Metadata about ontologies are widely used in the ontology repositories and for classification purposes (Jonquet *et al.*, 2018). So far, however, there has been a very little discussion about the role of metadata in the evaluation process. Many of the respondents found having access to additional information about ontologies, both in the form of labels and comments for/on different ontology components or as external documentation, to be very helpful. Some of the interviewees mentioned that before selecting an ontology for reuse, they would like to know if the ontology has followed or has been endorsed by any standard or common framework. Respondents also emphasized the importance of reusing other ontologies while building a new one and stated that they would check if an ontology has reused other ontologies and if it is based on an upper-level ontology. The language that an ontology has been built in and its size were amongst the other pieces of information that were at times used in the ontology evaluation process.

4.2.2.3 Evaluation based on Community and Social Aspects

A considerable amount of literature has been published on ontology evaluation since 1995 (Gómez-Pérez, 1995); however, far too little attention has been paid to the role of community in the evaluation and selection process. Popularity is one of the very few metrics introduced and used in the literature to address how communities might affect the quality and reusability of an ontology. This research was interested in determining if there is any social-related criterion, other than popularity (Martínez-Romero *et al.*, 2017) or the number of times an ontology has been reused before, that can affect how the quality and reusability of an ontology is assessed.

To address this research question, interviewees were asked if, and how, they interact with their community while searching for ontologies and evaluating them for reuse. They were also asked how important they thought social interactions and community were in the ontology development and selection process. Regarding popularity, interviewees were asked if they find having information about the number of times an ontology has been (re)used helpful in the selection process. The responses to the above-mentioned questions are discussed in the following sections.

Community and Ontology Search. The question "how do you find the ontology you want to reuse?" was asked to determine how ontologists and knowledge engineers search for and identify reusable ontologies. While the researcher was expecting to hear about some of the popular search engines in the ontology domain like Swoogle and BioPortal, literature and published papers were mentioned by many of the interviewees as one of their primary sources of finding ontologies.

Interviewee NBi4, for example, blamed his domain for lack of good and well-established repositories for ontologies and stated: *"I go to the literature"*. Another interviewee, SBi3, also emphasized the significant role of literature in the process of searching for ontologies and claimed that *"reading publications around the ontology"* is a very good method to help find ontologies, especially if someone is new to the field.

Besides helping to find a reusable ontology, some of the other interviewees stated that they use the literature and research papers as a tool to evaluate the quality of an ontology. Respondent NBi4, for example, argued that *"If an ontology is good and is used, you find a cite in the literature"*. Bi4 also linked the availability of research papers about an ontology with its popularity and stated: *"Popular ontologies are better ontologies; people are just familiar with popular ontologies so whenever you go to any ontology related conference, you will always have a workshop or a paper that talks about the ontology"*.

Community and Ontology Evaluation. Interviewees were asked if they interact with their community and how it affects the way they evaluate and select ontologies for reuse. According to the interviews, participants not only interact with their community to evaluate ontologies before selecting them for reuse, but some of them also evaluate the ontologies they are building by the feedback they receive from the community. Social aspects affecting ontology quality and evaluation are discussed below.

Build Related Information. Several respondents discussed the importance of different types of build related information, such as who and which organization has built the ontology, what the ontology has been built for (e.g., the use case), who are the different stakeholders of the ontology, and how the ontology was built (e.g., in collaboration) in the evaluation and selection process.

Interestingly, some of the interviewees claimed that to evaluate an ontology, they will ask themselves if they know the developer of the ontology. Interviewee Bi3, for example, stated: *"I have to say, in reusing things, there is often politics, and connections are as important as anything else. So, it is not always the best one that wins"*. They also argued that the quality of an ontology might sometimes come second: *"I may not like a particular ontology, but because a bunch of other people are using it, and I want to standardize with them, I might use it anyway"*.

Respondent SBi4 also discussed the issue of trusting an ontology's developer team and stated:

"Science is a social enterprise, I mean this is how everything works in science; you know, if you look at a paper, do you trust the paper? You look at the authors first and then you read the paper, and then you pick about what they have done. But, yes, I mean it is a major criteria, major quality criteria, it may or may not be right. It is a bit of old boys club, but yes, that is how people make decision. I normally read the definitions and then go to other things; do I trust the people who are making it?"

Besides the information about the developer team or organization, some of the respondents would consider the reasons that an ontology was built and used for, before selecting it for reuse. They were also interested in having some information about the stakeholders and the other users of ontologies. Interviewee SBi3, for example, stated:

"Completely separated from the people developing it, are there other people who use this ontology beyond just that group? That tells you something about it. I think, also, finding out how they are using it, is also important; you know, what data is being annotated with those ontologies is also an important question. But, I have some data, and I know I want to integrate

with something done in another institute; what is the ontology there they are using? That is also important, so I think there is a list of the things you want to check".

Regular Updates and Maintenance. In the interviews, there were numerous examples of linking the quality of an ontology to how regularly it is updated and maintained. For some participants like NBi3, the regularity of updates was the first thing that they would look at when evaluating ontologies: *"Somebody built an ontology during his research in 1998, and he stored it on the web, and then he left it; it is available but not updated. Things will get obsolete very soon. So, we make sure to use the ontologies which are regularly updated; it is the first thing".*

SBi3 also compared maintenance with some of the very popular quality metrics in the literature, like coverage, and elaborated:

"Does it have my terms? I think is important but there are many others that you need to consider when you are picking an ontology, beyond just does it have the words in ontology. About maintenance, do they update regularly? Do they release regularly? Do they have a record of doing that? How responsive they are to updates when you need new terms? All that sort of stuff. If they are publishing it once every two years, it is probably not a good ontology".

Bi1 emphasised the significant role of updates and maintenance in their domain and stated: *"No way that an ontology is keeping on in biology, not getting updated. Biology is changing too fast; so, all the relevant ontologies in biology are getting updated".* Interviewee NBi2 also made a link between the nature of the domain they were working in and the necessity of regular updates: *"It is about flexibility; in manufacturing business things are changing all the time. So, you need solutions that are easy and flexible to stay in, to stay relevant to what you are doing tomorrow, as well as what you are trying to do today".*

Interviewee Bi3 compared ontology engineering with software engineering and argued: *"If you are going to reuse a piece of open source software, you will do the same thing; you will open the GitHub*[18] *website, and say you know, if you looked in it and nobody updated it or anything in three years, you might think no; whereas if it looks like there is an active ongoing community, you will think yes. If I have problems, I can ask people, and I can get bugs fixed".*

Bi4 suggested that there is a link between the popularity of an ontology and the regularity of updating it and stated: *"It might be useful to use popular ones because there are the ones that*

[18] https://github.com/

are mostly updated; so, Gene Ontology has a release, I think, every day or every 12 hours; so, the popular ontologies are the ones that are most updated". Apart from the regularity of updates, how people deal with it is also an important issue. Respondent SBi3 talked about the importance of having an update mechanism and argued:

"I think in the field that I am working, there are other challenges; one of which is how you deal with update mechanism of ontologies. If you annotate data to ontology which is typically use case for how you keep up-to-date with the fact that ontologies change reasonably often; you might have a big database of data, that you used the data in, new ontologies come along; that affect the way the data has been represented in your database, gotta have a update mechanisms for dealing with that and that can be tricky actually; it is not as simple often as swapping things out when something gets made obsolete, it is replaced with other things, you have to deal with".

Responsiveness. Responsiveness of the ontology developer team was among one of the other widely mentioned criteria for ontology evaluation and selection for reuse. Some of the respondents argued that, not only knowing the developer team and organization is essential, but also having an active ongoing community and their willingness to collaborate, evolve and develop the ontology further is important.

This is what Interviewee Bi3 had to say about the importance of responsiveness: *"I would say, it is definitely high up; I mean having someone at the other end of the line that you feel that you can trust is definitely very important. If it looks like there is an active ongoing community, you will think yes, if I have problems, I can ask people and I can get bugs fixed".* SBi5 used one of the popular ontologies in their field as an example and added: *"For example, the fact that the Gene Ontology has a huge community behind it is important, because it means that they have a curation process in place and quality assurance and so on; so, that kind of gives more confidence that the ontology is as good as it can be. It is not perfect for sure, but I mean that it is vetted by the community".*

Respondent Bi1 chose responsiveness as the first quality metric they would consider for evaluating an ontology and compared it with other very popular metrics, such as, the availability of documentation:

"I would say the responsiveness of the team, obviously, is the top-quality metric for me because nothing is perfect, but if something gets improved, then it will get good. Like, if you have a question, you need to add a term, something does not make sense, you contact them, they

answer, and they answer in a constructive way. This is good because all the ontologies are work in progress; there is no finished ontology in my domain".

Popularity. Respondents' thoughts and comments on popularity can be classified into the following groups: (1) those who were against this metric, (2) those who liked popularity as a quality metric for ontology evaluation, but did not agree with the way it was being computed and (3) those who found this metric useful.

The first group of respondents argued that the popularity of an ontology or the number of times it has been used is not that important. As interviewee Bi1 would put it: *"To me, it would not be very important, except if two ontologies are really very equal in everything else, I will take the most used ones, but I do not think; It is not really relevant to me. If it is the right tool for the job, it is the right tool!".*

It was also argued that the number of times an ontology is used depends on different factors, such as its size, its level of specialization and the domain that it is built in; therefore, popularity cannot be considered as a metric to measure quality. According to interviewee Bi1, for example, *"some ontologies are more specialised, so less people use them, because it corresponds to a very special need; but maybe these people, are the right people and are using it well".*

Interviewee SBi3 also linked the number of times an ontology has been used to its size and added: *"If there is a small ontology, but really focused on representing an area that has not been done before, but it is correct, it is absolutely correct. I think that is perfectly reasonable, even if it is not widely used"*. NBi5 found popularity a helpful metric, but argued that it is highly dependent on the domain that an ontology is used in: *"It depends on the domain that it has been reused in; if it is just medical domain, it is difficult to say that it is a reusable ontology".*

The second group agreed on the necessity of having such a metric to identify the more popular ontologies in different domains but were not sure about the usefulness of the current method that is used to measure popularity. As interviewee NBi3 would put it: *"How many times an ontology is viewed will not help you. I may click just for exploration, and I will say, it is not my thing, and I don't want it. It shows how catchy the term is or how important, how regularly, how often this term is chosen, but it does not mean the use of the ontology. So, I think there should be some other way".*

Bi4 used a very interesting personal experience while discussing the inaccuracy of the current techniques of measuring popularity:

"We found that Gene Ontology is not accessed that much using BioPortal, and I thought that it was very surprising, because the Gene Ontology is very famous, and then I found out because there is a Gene Ontology browser called AmiGo[19]*, and their visualizer tool is much better than BioPortal visualization of Gene Ontology, so people generally go to Gene Ontology website and lunch the AmiGo browser and go to Gene Ontology there. So, you can say that Gene Ontology is much more accepted but if you just look at the clicks (in BioPortal) and you might say that Gene Ontology is not that much famous".*

Interviewee SBi3 thought that having a quality metric, like popularity, is a step in the right direction but argued that it might be misleading by causing a snowball effect. According to them:

"I can see that you can also put in a little metric for usage or browsing or how many people read these things; that is a kind of useful, but it does not tell you the whole picture. You know, you can end up with a false signal there. You recommended an ontology because it is useful, because someone uses it and then you recommend it; so, someone else uses it and so on and so on. What I mean, so you are getting in that cycle of, it grows and grows".

The last and also the minority group were those who claimed it is worth having a metric like popularity and highlighted the importance of community acceptance. According to interviewee NBi4:

"If a community is using the ontology and is happy with it, I take things to account. So, I try to reuse or to do something to extend it or maybe very careful on changing it. I need to have motivations because after all, ontologies should have people working in the domain and so, if they are happy with that one, and I see things that are no good, I point it out and I may suggest an extension, whatever, but I try to reuse what I have".

The other definition of popularity, that focuses on the link between popularity and the number of imported ontologies, was also brought up by some of the respondents. NBi5, for example,

[19] http://amigo.geneontology.org/amigo

made a link between the quality of an ontology and the fact that the ontology has reused other ontologies and said:

"The quality of an ontology depends on the relation between the ontology to upper-level ontologies; the more 'same-as', 'equivalent-as' links I can find in an ontology. It also can be seen as a sign or a feature of the ontology that can be reused because, if it is 'same -as' a concept that we already know, then it can be replaced".

NBi6 also argued that reusing some of the ontologies are inevitable and not reusing them will leave a bad impression:

"Whenever I have an ontology, where there is a person, I will never ever create my own person class; I will always reuse FOAF [ontology]. I think it would be ridiculous to create my own class and some of those are very very strong class definition. So, it will always worth reusing and I think it will be even mistake by ontology engineer to develop their own class and for me, if I see an ontology doing that, I will get a negative impression".

4.3 Summary

The main goal of this chapter was to determine and explore the set of steps that ontologists and knowledge engineers tend to take in different phases of the ontology selection process, from ontology search and discovery to ontology evaluation and selection for reuse. The interview study confirmed the findings of the previous studies, that have mainly focused on the importance of the internal aspects of ontologies, such as, their content, structure and consistency. Moreover, it contributed to existing knowledge by providing a new understanding of ontology evaluation and the factors it depends on.

As the results suggested, the quality and reusability of an ontology can be assessed by exploring different metadata that is available for it, or by investigating the interactions among its respective community. The findings also suggested that ontologies are usually considered as incomplete ongoing projects; therefore, knowing and trusting their developer and maintenance team or organization before selecting them for reuse is essential. To confirm and expand these results, a second phase was designed and implemented using a questionnaire; the findings of this second phase are presented in the next chapter.

Chapter 5: SURVEY ANALYSIS

5.1 Introduction

In the second phase of this research, a survey was designed and sent to a large group of knowledge engineers to ask about the process of ontology evaluation and selection for reuse. This chapter describes the findings of the survey study. It starts by explaining the data preparation process and demographics of the participants and moves on to present the analysis and answers to different questions of this research, especially those asked in objective 3 and 4.

5.2 Data Preparation

The survey was conducted over a five-month period from September 2017 to early January 2018. The first step after closing the survey was to prepare the data for analysis. Data preparation is the process of converting raw data into a form that is useful for analysis; it includes clearing the database, coding data and assigning numeric values, and dealing with missing data (Creswell and Plano Clark, 2007, p. 130). In this study, data preparation started by removing some of the fields that had automatically been collected by Qualtrics, the application used to conduct the survey, but were not used in this research. Those fields included: “End data”, “Duration”, “Status”, and “Finished”.

The next step was to deal with the missing data. Missing data and survey nonresponse rate are among the most critical challenges in any survey research and are used as an indicator of the survey data availability and quality (Bates *et al.*, 2001). There are two main types of nonresponse identified in the literature: (1) unit nonresponse, that refers to the complete absence of the sampled unit (Yan and Curtin, 2010) and can be the result of failure to contact or persuade the sampled unit to participate in the survey (Brick, 2013), and (2) item nonresponse, that happens when some questions or items in a survey are left unanswered (Yan and Curtin, 2010).

In this research, a link to the survey was sent to 12 different mailing lists. The researcher was familiar with some of the mailing lists. Google search was also used to identify some of them. Calculating the size of the mailing lists is hardly possible as some of them do not provide access to their members’ list and info. Consequently, it is difficult to calculate the exact number of non-responders. However, the initial exploration of the survey data showed a large number of partial and incomplete responses. Some of the participants had only clicked on the link to the survey and had left it or did not come back to complete it, over a two-week period. The percentage of progress for those respondents was mostly around 13%. For the sake of completeness and quality, it was decided to only keep the complete responses with 100% progress (162 out of 314), even though it reduced the sample size.

Missing data can also happen when the screening questions are used, and the data is not required from some of the respondents as the result of their experiences and the skip logic rules used in the survey (Saunders, Lewis and Thornhill, 2012, p. 485). Screening questions were used in different sections of the survey. For example, before presenting respondents with the quality-related criteria, they were first asked if they had previously evaluated an ontology. According to the survey data, only one of the respondents had no prior experience of ontology evaluation and therefore, was not presented with the quality-related criteria. It was decided to keep this respondent and to address the missing data using multiple imputations in SPSS (IBM, 2016).

Moreover, a review of the data revealed that some of the respondents had provided invalid information for some of the essential and mandatory fields in the survey, such as demographic information. According to Lavrakas (2008), mandatory variables in a survey are those that are significantly correlated with the variables of interest. Five of the responses were eliminated, as they had entered invalid characters, e.g., "?" and "-", and letters, e.g., "b", in all the spaces provided for the questions about job title, organisation type, and the domain they had built ontologies for.

Some of the responders had also used abbreviations to answer the open-ended questions in the survey. For example, when asked about their job title, some had written "Prof" or "prof." instead of "Professor". To bring consistency to the data, all the abbreviations were changed to the full form of the words. Once the responses with less than 100% progress and those with invalid information were removed, the amount of missing data was minimal, and therefore, the missing data did not significantly impact the overall result. The final population of the survey was 157.

5.3 Demographics of Respondents

This study managed to access ontologists and knowledge engineers with many years of experience in building and reusing ontologies. More than 78% of the participants in the survey were actively involved in the ontology development process, and 95% of them would consider reusing existing ontologies before building a new one. The survey data was collected anonymously, meaning that no personally identifying information, such as name and geographic location, was collected. The 157 respondents of the study are categorised by the following demographics, all declared by the responders:

- **Job Title.** After conducting frequency analysis on the job titles provided by respondents, 78 unique job titles were identified, many of which were related to different roles and positions in academia, such as researcher, professor, and lecturer (see Appendix I).

- **Type of Organisation.** According to the frequency analysis conducted on the organisation types, more than 50% (80) of the respondents were working in universities; the remaining participants were working in 30 other types of organisations, ranging from research institutes to different companies and industries.

- **Years of Experience.** Interestingly, most of those who were surveyed were experts in their domain, and only around 10% of the respondents had less than two years of experience. Around 46% of the survey participants had more than ten years of experience. The second largest group of respondents were the knowledge engineers with five to ten years of experience (26.54%).

- **Domains Participants Had Built, or Reused Ontologies in.** Survey respondents had worked/were working in many different domains, and most of them had mentioned more than one domain, some of which were not related to each other.

5.4 Data Analysis

This section aims to present the results obtained from the survey analysis. It starts by discussing what reusability meant to participants in this study and then moves on to explore the process of ontology search and identification and different selection systems that are used in that process. The main part of this section discusses ontology evaluation and the metrics that it depends on in detail.

5.4.1 Ontology Reuse

Despite all the advantages, the popularity of ontology reuse amongst ontologists and knowledge engineers is not yet apparent (Fernández-López *et al.*, 2019). Participants in this phase of the study were asked a couple of questions about reuse, including how often they consider reusing ontologies and if they prefer to contribute to existing ontologies instead of building a new one. They were also asked about what they thought the main characteristics of a reusable ontology were and about the set of steps they tend to take or the information they provide while developing an ontology to make it reusable. The following section summarises the survey respondents' thoughts on ontology reuse.

5.4.1.1 Ontology Reuse Importance

One of the very first questions of the survey was "How often do you build an ontology?". This question aimed to identify the level of involvement of the survey respondents in the ontology development process. According to the responses, only 20% (32 out of 157) of the survey participants had never or rarely been involved in the ontology development process. Respondents were then asked how often they would consider reusing existing ontologies. Interestingly, all of the survey participants would consider reusing existing ontologies before developing a new one.

To clarify the role and importance of ontology reuse, respondents were also asked to express their level of agreement with the "I prefer to contribute to existing ontologies instead of developing a new one" statement. The responses to this question were interesting, as only 20 out of 157 (12.7%) of the respondents disagreed or strongly disagreed with it. Meaning that while all of the respondents would consider reusing ontologies, 12.7% of them were reluctant to do it, and 42% (66 out of 157) of them were neutral.

Survey participants were presented with one of the mentioned issues in the first phase, which was "In my domain, ontologies are not built to be shared and reused (because of the nature of the domain, intellectual property and/or financial concerns)" and were asked to express their level of agreement or disagreement with it. According to the results, only 23 out of 157 (14.6%) of the respondents agreed or strongly agreed with this statement and suggested that ontologies in their domain are not built to be shared and reused. The remaining respondents, however, stated that ontologies are built to be shared and reused.

This research was also interested in exploring the definition of the term "reusable" in the ontology domain. Doing this would not only help in clarifying one of the most important terms in this domain but would also help in clarifying some of the other terms and processes in this field, such as ontology quality and ontology evaluation. Therefore, survey respondents were asked how they make sure that the ontology they are building is reusable, what they thought the main features of a reusable ontology were, and what the best ontology they had ever reused was and why they thought it was the best.

5.4.1.2 How to Build a Reusable Ontology?

To find an answer to this question, respondents were presented with five different Likert items, each containing information that could be provided or the steps that could be taken to ensure an ontology is reusable. They were then asked how often they consider providing that information or taking those steps. They were also given a space at the end that would allow them to provide more answers and explanations, if they wanted to. The "Defining and clarifying the scope and the goals of the ontology" statement had the highest median of 5 (Mean:4.29), and 80 out of 152 (52.6%) respondents with previous experience of developing ontologies had stated that they would always define and clarify the scope and the goals of an ontology in order to make it reusable.

The other popular items were "providing proper documentation" and "reusing other ontologies", both with median of 4 and mean of 4.09 and 3.96 respectively. Considering the size of an ontology, 71.7% of the respondents had stated that they sometimes or very often make sure that the ontology they are developing is more reusable by making it smaller and more specialized (36.8% and 34.9% respectively). Using methods or other standard practices for ontology development had the lowest median and mean of 3; only 20.4% (31 out of 152) of the survey respondents would always use a method to make sure that the ontology they were building is reusable.

Screening questions were used to make sure that the respondents with no previous experience in ontology development are presented with a different set of questions, which would ask them "If you ever wanted to build an ontology, you would make sure that it is reusable by …"; the suggested set of actions and information, however, were the same. Those with no previous experience of ontology development had ranked the presented items slightly differently. For them, providing proper documentation was the item with the highest median of 5 (mean: 4.2, standard deviation (SD):1.3), followed by defining and clarifying scope and goals (median:4, mean:4, SD: 1.225), using a method or methodology (median: 4, mean: 3.6, SD: 1.14), reusing other ontologies (median: 3, mean: 3.6, SD: 0.894) and building smaller but more specialized ontologies (median: 3, mean: 3.6, SD: 1.342).

Survey respondents were provided with a space to enter the other steps they would take or the information they would provide to ensure that the ontology they are developing is reusable. The information provided in this part was analysed using NVivo (QSR International, 2015) and was grouped into different themes. What follows are some of the identified themes and aspects of ontologies that can affect their reusability.

- **Content.** Content was one of the most frequently mentioned aspects. Survey respondents had argued that for an ontology to be reusable, its content has to be explicit, accurate, precise, unambiguous, clearly defined and easy to understand. Many of them had also suggested that an ontology content should be generic, general, limited with minimum rules and *"broad enough to be applied to different contexts"*. Modularity was the other content related characteristic widely mentioned in the answers to this question; some of the respondents had argued that a reusable ontology is *"modular or easy to modularise"*. Structure and class hierarchy were the other internal aspects mentioned in this section.

- **Documentation.** Documentation was the second most mentioned characteristic of a reusable ontology. As it is seen in the literature and also the first phase of this study, documentation has always been linked to the reusability and quality of ontologies. The survey findings had suggested that ontologists and knowledge engineers would not only like to see different internal and natural language comments, such as "rdfs:label" and "rdfs:comment", but they would also like to see use cases, guidelines about how an ontology can be or should be used, and also published papers or technical reports that describe the ontology and its background assumptions.

- **Scope, Goal, Purpose, Context and Application.** These were amongst the other terms used to refer to different types of additional information that should be provided for a reusable ontology. One of the respondents, for example, had emphasised the importance of use case and context in ontology development and argued: *"we focus on 'usable' rather than 'reusable' and then build new ontologies from old"*. Another participant had also discussed some of the challenges faced in the reuse process by stating: *"the main thing that limits reusability has been the rules that apply in the original context, but not in a different or wider context"*. The importance of provenance information and metadata statements, such as ontology creator and license type were also discussed.

- **Being Based on or Having Reused Other Ontologies.** This criterion was the third most mentioned (35 times) characteristic of a reusable ontology. Some of the respondents had emphasised that an ontology needs to (re)use or be based on, connected and linked to other well-known ontologies in order to be reusable. One of the participants had suggested that instead of using many "owl:imports" statements, which might cause conflicts when merging, ontology developer should *"reuse/bind to upper ontologies by 'citing' the particular concepts only by their original IRIs (with "rdfs:isDefinedBy" links back to their ontology)"*.

- **Accessibility.** Respondents had used terms like open, public, online, available, accessible and findable to refer to accessibility and license aspects of ontologies. As one of the respondents had put it: *"no one can reuse an ontology that is not findable"*. The other respondent had also stated: *"it (ontology) needs to be publicly available and browsable, (have) proper licensing, (and be) modifiable. I need to be able to update terms, request clarification, add new information"*. Survey participants claimed that to be reusable, not only the ontology itself should be online and accessible via working URLs, but the information about its ownership and license type should also be published and available online.

- **Being Standardised or Based on a Design Patterns or a Set of Principles.** Standardisation was the other characteristic of a reusable ontology mentioned by some of the survey respondents.

- **Size.** Ontology size was mentioned 12 times in this section; some of the respondents had argued that small, lightweight and compact ontologies are more reusable. One of them, for example, had put it this way: *"complete ontologies are rarely reusable, except for very high-level ones, but the smaller pattern can be"*. The other respondent had also defined reusable ontologies as the ones *"that are small enough that it's not a major overhaul to import them"*.

- **Language.** The language that an ontology had been built in was also mentioned by some of the survey participants. One of the respondents, for example, had argued that a reusable ontology should be *"described with recommended language"*; the other one had stated that *"OWL/RDFS is a plus"*. Language has been previously mentioned both in the literature (Lozano-Tello and Gómez-Pérez, 2004) and the first phase of this study as a criterion for ontology evaluation. A Likert item was also used in the survey and asked how important respondents thought the language that an ontology is built in is in the evaluation process. This criterion was considered important and very important by 32.5% and 33.8% of the respondents respectively, with a median of 4 and mode of 5.

  Moreover, and at the end of the survey, respondents were asked about the primary language they built ontologies in. Being aware of the importance of language in ontology evaluation, this question aimed to identify the most popular language for ontology development. According to the answers to this question, 80% (126 out of 157) of the respondents had selected OWL as one of the languages they had built ontologies in. RDFS was the second chosen language; 42.7% of the respondents had some experience of developing ontologies using this language. DAML+OIL and SHOE were both mentioned once; other languages such as SKOS and FOL had also been suggested.

- **Community.** Some of the respondents had discussed how community and trust could impact the reusability of an ontology. For example, a reusable ontology was described as the one that has been *"vetted"* by other knowledgeable users or recognised authorities as being useful and readily reusable or the one that is well-used and accepted in the community. It was also mentioned that a reusable ontology is the one that *"comes from a trustable source"*. Some had also made a link between community, update and maintenance, and reusability by suggesting *"active support and development"* and *"promise of continuous maintenance"* as the characteristics of a reusable ontology.

- **General.** Some of the participants had provided some general definitions for a reusable ontology that worth mentioning. According to one of the interesting definitions, a reusable ontology *"covers what is not currently covered in an accessible format"*. The other respondent had defined a reusable ontology as an ontology that has *"clearly been* *developed for and is used in a number of different applications. In most cases, it is rather simple, covering only the core conceptualisation of a certain domain, and doesn't involve complex logical constraints - the general trade-off in practice seems to be: the less ontological commitment, the easier the reuse".*

5.4.1.3 Best Ontologies Respondents Had Previously Reused

This research was interested in identifying what ontologists and knowledge engineers consider as the "best ontology" and exploring a set of characteristics those ontologies had in common. Participants had provided 146 valid responses when asked about the best ontology they had ever reused. Table 5-1 presents some of the most mentioned ontologies.

Table 5-1 The Best Ontologies According to the Survey Respondents

#	Ontology Name	N	Domain	Size (Number of classes)	Language
1	Basic Formal Ontology (BFO)	14	Upper level	35	OWL
2	PROV Ontology (PROV-O)	13	Upper level	31	OWL2
3	Dublin Core (DC)	12	Upper level	11	NA
4	Simple Knowledge Organization System (SKOS)	12	Upper level	4	RDF/OWL
5	FOAF	11	Upper level	13	RDF
6	Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE)	9	Upper level	NA	OWL
7	Semantic Sensor Network Ontology (SSN)	9	Sensors & their observations	23	OWL

The reasons respondents had provided for calling the ontologies mentioned above "the best" are as follows:

- **Basic Formal Ontology (BFO)[20].** BFO is a small upper-level ontology that aims to support "integration of data obtained through scientific research" (Arp and Smith, 2008). There were different reasons why BFO was called the best, namely, because it is well documented, simple and is used by so many other ontologies. One of the respondents had described BFO as *"an upper ontology faithful to reality"* and had mentioned the fact that BFO is *"used by other ontologies of the same quality level"* as the reason of calling it the best. The other interesting reasons included, *"because it must be used in healthcare systems"*, and because *"there is a YouTube channel"* for it.

[20] https://basic-formal-ontology.org/

- **PROV Ontology (PROV-O)[21].** PROV-O is used to present provenance information generated in different systems. PROV-O was described as a light, simple, generic, versatile, adaptable, well-documented, well-established, standardised and high-quality ontology. Besides all the positive responses and the useful characteristics of this ontology, some of the respondents did not find this ontology ideal but used it anyhow because there was no other alternative ontology for describing provenance information or because it covers several use cases.

- **Dublin Core (DC)[22].** For some of the respondents, DC was the best ontology they had ever reused. DC is not an ontology but a set of fifteen different elements that can be used for describing resources. Reasons stated for using DC were because it is lightweight, well-known and well defined. Some would also use DC because *"it is a good way to be linked with others"*, it *"does what it sets out to do"*, and it is not too big. One of the interesting points observed during the first two phases of this study was that some of the ontologies are used, even if they are not the best or even good enough. As one of the respondents had put it: *"It's [DC is] not very clear, quite messy and irritating, but it covers a domain that everybody needs"*.

- **Simple Knowledge Organization System (SKOS)[23].** Similar to DC, SKOS is not an ontology but a common data model and a W3C recommendation that facilitates the process of "sharing and linking knowledge organisation systems via the Web" (Miles and Bechhofer, 2009). Some of the reasons mentioned for using SKOS included its flexibility, *"its uptake and ease of use"*, and because it fits different use cases.

- **Friend of a Friend (FOAF)[24].** As one of the very popular vocabularies or ontologies, FOAF includes a set of keywords that can be used to describe humans and the links and relationships between them (Brickley and Miller, 2010). FOAF was mentioned 11 times (out of 147) as the best ontology used by the respondents. Some of the reasons for calling FOAF the best were very similar to those that were mentioned for DC and SKOS, which included being well-defined, well-known, having proper documentation, being very compact and having a small size.

[21] https://www.w3.org/TR/prov-o/
[22] http://www.dublincore.org/specifications/dublin-core/dcmi-terms/
[23] https://www.w3.org/2004/02/skos/
[24] http://www.foaf-project.org/

The other reason mentioned for using FOAF was because it is the standard way of representing concepts related to humans. The previous phase of this study indicated that failing to reuse the standard ontologies might affect how people judge the overall quality of an ontology (Talebpour, Sykora and Jackson, 2017).

- **Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE)[25].** As the first module in the WonderWeb Foundational Ontologies Library (WFOL), DOLCE aims to capture the "ontological categories underlying natural language and human common sense" (Masolo *et al.*, 2002). Being upper-level, rigorously developed, OntoClean-approved and well documented were some of the reasons mentioned by the survey respondents for calling DOLCE the best. One of the respondents had also added: *"[DOLCE is] based on an accurate ontological analysis; it provides a sound framework for domain ontologies; it exists both in reference strongly axiomatised versions and in lighter OWL versions; it is clearly described in the scientific literature"*.

- **Semantic Sensor Network Ontology (SSN)[26].** This ontology is proposed and used to describe sensors and their observation. Survey respondents had mentioned some of the characteristics of this ontology, namely, being developed and endorsed by W3C as the reason for supporting it. SSN was also described as *"an intelligent and well-modelled ontology"* that has *"a nice level of abstraction and is kept quite generic"*.

- **Time and Geo Related Ontologies.** One of the respondents had linked the stability of the domain that an ontology has been built in/for and its reusability and had argued: *"ontologies in uncontroversial and stable domains tend to be reusable, e.g., the W3C Time ontology[27]"*. Similar reasons were provided while mentioning why ontologies related to geolocation are the best. One of the respondents, for example, had mentioned that GeoNames[28] is the best ontology they had reused because it *"takes the hassle out of modelling geographical data"*. Being well designed and well documented were among the other reasons for calling these standard ontologies the best.

[25] http://www.loa.istc.cnr.it/dolce/overview.html
[26] https://www.w3.org/TR/vocab-ssn/
[27] https://www.w3.org/TR/owl-time/
[28] http://www.geonames.org/ontology/documentation.html

- **Schema.org[29], DBpedia[30], Suggested Upper Merged Ontology (SUMO)[31], The Lexicon Model for Ontologies (Lemon)[32], QUDT[33], and Unified Foundational Ontology (UFO)[34]**. The reasons for calling these ontologies the best included being well documented, simple, modular, based on, compliant with, or recommended by a standard like W3C and continuously maintained and developed. Ontologies following or based on OBO principles or included in the OBO library were also considered as the best by some of the survey respondents. Moreover, GO was called the best a couple of times, because it has lots of documentation, it is well funded, and it is the oldest and the most developed ontology in biology.

Overall, the most apparent finding of this section is that no ontology could be called the "best" and usage of ontologies depend on many different factors, such as, application requirements and use cases. As one of the respondents had put it: *"this depends on what you mean by an ontology, and 'best'"*. It was also seen that ontologists and knowledge engineers tend to use some of the ontologies, even if they do not have the highest quality. One of the respondents, for example, had stated: *"none are good enough to call one the best. Though some biomedical ontologies are quite good"*. Moreover, while explaining why an ontology was reused, one of the respondents had said: *"not anything specific, but it catered to our requirement"*.

The survey respondents had interesting thoughts on the role of standardization. One of them had stated the best ontologies are *"all the standard ones - they are standard for a reason"*. The other responded had also discussed how reusing a standard ontology might affect the quality of the ontology someone is building by stating: *"W3C ontologies are among the best because their popularity. By reusing them you reach a large audience and makes your own ontology appealing and easily understandable for others"*. Interestingly, some respondents had pointed out the name of an author of an ontology or the organisation responsible for its development and had mentioned the best ontologies are the ones that are developed by them.

There was also some discussion around how the size of an ontology might affect its reusability. Ontology size, or the number of classes it has, has been used as a metric to evaluate ontologies by some of the well-known ontology selection and recommendation systems, like BioPortal.

[29] https://schema.org/
[30] https://wiki.dbpedia.org/
[31] http://www.adampease.org/OP/
[32] https://lemon-model.net/
[33] http://qudt.org/
[34] http://dev.nemo.inf.ufes.br/seon/UFO.html

In this system, smaller ontologies are thought to be "more specialized" and will get a higher score in the ranking and recommendation process (Martínez-Romero *et al.*, 2017). Despite this, most of the previous studies have failed to provide a definition for "small" or "large" ontologies or agree on the minimum or maximum number of classes that an ontology should have to be considered as small or large. Participants in this study referred to ontologies, such as FOAF, with less than 50 classes as small. SNOMED CT, however, with more than 357,533 classes was thought to be a large ontology.

Findings of the survey conducted in this research suggested that smaller ontologies are more favourable. As one of the respondents had put it: *"limited size made it easy to use"*. However, some respondents had suggested the steps that can be taken to facilitate reusing larger ontologies. This is what one of the respondents had to say about reusing CIDOC-CRM[35]: *"though it is huge, it has very good documentation"*. The other respondent had also explained how the modular design of an ontology might help in the reuse process: *"(reuse) worked best when the ontologies were very modular, and different types of content were specified using different vocabulary. In those cases, the more general modules, and even the mid-level modules were reused often and successfully"*.

The other interesting observation was that the most used ontologies are domain independent, general, and small; reusing some ontologies like FOAF, Time, and Geo is inevitable. Here is what one of the respondents had to say about the link between an ontology type and its reuse: *"I have never succeeded in reusing a truly domain-specific ontology other than perhaps to extend upon an existing one"*. Another respondent had also argued that top or upper-level ontologies are the best, while the other one had stated: "*the best ones are the middle-layer ones, such as Information Artifact Ontology (IAO)*[36]*, that provide common building blocks I can then extend"*.

Taken together, the best ontologies suggested in this section had the following characteristics:

- They are simple upper-level ontologies
- They are small (have less than 50 classes)
- They are built using OWL
- They have followed and/or been endorsed by well-known standards, such as, W3C and OBO Foundry

[35] http://www.cidoc-crm.org/
[36] http://www.obofoundry.org/ontology/iao.html

- They are well documented and maintained

5.4.2 Searching for a Reusable Ontology

Different search and selection systems have been proposed for ontologies since early 2004. However, the findings of the first round of this study suggested that some of the selection systems for ontologies are not that popular, and knowledge engineers might prefer to refer to the literature or their community to identify ontologies for reuse (Talebpour, Sykora and Jackson, 2017). To explore this matter further, the survey respondents were presented with a set of Likert questions, each mentioning a type of search and selection systems and were asked how often they use those systems. Moreover, participants in the survey were asked to share the main challenges they have faced while searching for a reusable ontology. The responses to these questions are presented in the following sections.

5.4.2.1 Ontology Search and Selection Systems

When asked about the selection systems, only 17.2% (27 out of 157) of the respondents stated that they would never or rarely search the literature to find an ontology for reuse; literature and scientific papers also had the highest median and the lowest SD, which indicated the highest level of agreement amongst the survey respondents. Google was ranked second with a median of 4, but a larger SD compared to literature and scientific papers. Interestingly, repositories and search engines for ontologies were ranked third and fifth, respectively. The *"I do not search; I know the ontologies I want to reuse"* statement was ranked fourth and 42 (out of 157) survey participants had stated that they very often or always know the ontology they want to reuse.

Some very interesting responses were also collected from the space provided for any "Other" search or selection systems used for finding ontologies. From the 67 responses collected in this section, 36 (53.7%) of them had referred to LOV, which is a "high-quality catalogue of reusable vocabularies for the description of data on the Web" (Vandenbussche *et al.*, 2017). OBO Foundry and Ontobee[37] (Ong *et al.*, 2017) were amongst the other selection systems mentioned by the respondents in this part of the survey, seven and six times, respectively. The early selection systems for ontologies were not that popular amongst the respondents. One of the respondents, for example, had called Swoogle and Watson *"outdated and useless"*.

[37] http://www.ontobee.org/

5.4.2.2 Challenges Faced while Searching for a Reusable Ontology

When asked about the main challenges faced while looking for a reusable ontology, survey participants discussed different issues. According to some of them, especially those who were new to the field, the first challenge was "having to know where to look" and *"knowing the best sources to start with"*. Participants in this phase of the study blamed ontology search and selection systems for not being reliable, not properly indexing ontologies, being context unaware or even *"rubbish"*. Google, which was named as one of the most popular tools for finding ontologies, was also blamed for not doing more regarding ontology search and selection.

How different selection systems work and identifying the type of input they accept was mentioned as one of the other challenges in the selection process. As one of the respondents had put it: *"finding keywords to identify relevant ontologies, especially when you are not familiar with the domain in the first place"*. Availability of different ontologies and the fact that many of them cover the same purpose and overlap was the next challenge some of the survey participants had faced while searching for a reusable ontology. As one of them had described: *"identifying the right ontology to rely on when multiple ontologies capture part of the knowledge we like to have [is a challenge]"*.

Finding an ontology that will best fit the requirements of an application or project was the other mentioned challenge. Some of the respondents had argued that it is hard to identify ontologies that precisely correspond to their needs because *"almost every ontology is specific to a task; new tasks and applications typically have specific requirements that an existing ontology doesn't fit"*. One of the respondents had added: *"even if you find an ontology that seems to be a good fit for reuse, it may turn out to have built-in dependencies on others that are not!"*.

Interestingly, a large number of the answers provided for the question about the challenges of ontology search were more about the quality of ontologies and their evaluation. Some respondents, for example, had argued that *"many of the ontologies out there are not very good and often abandoned"* or *"most of the ontologies are not complete and underspecified"*. The other ones had expressed their concern about the quality by mentioning that they *"cannot anticipate the quality of an ontology"* or saying that *"in general, it [ontology] is not a complete work"* and *"is often different from what it is said to be"*.

"Lack of documentation", *"poor quality documentation"* and *"bad documentation"* were also mentioned many times while talking about the challenges faced in ontology search. The results of this study had suggested that documentation is a critical factor in the selection process, as it can help to find an answer to one of the fundamental questions that ontology developers have while looking for an ontology, that *is "why one [ontology] would be more useful than another or more appropriate"*. Ontologists and knowledge engineers would also like to know *"whether it [ontology] was designed with reuse in mind or if it is, in fact, an application-specific"* and *"what are the background and modelling assumptions"*. Moreover, they would like to understand *"the authors conception of the domain"* and to have information about *"the use cases it [ontology] was designed for and whether it can be used for [their] use case"*.

Availability, accessibility, and maintenance were amongst the other mentioned issues. While referring to the literature was the most popular method of identifying ontologies, some of the respondents argued that *"some ontologies published only in research papers and not published in computerised form"*. Not being accessible, unclear development status, and the fact that *"authors cannot be reached as they changed institute"* and will not be able to maintain ontologies, or respond to requests, were the other discussed challenges.

Respondents also argued that they would want to find an ontology that is reputable in their domain and *"is supported and has an active community"*. They would also like to have access to expert opinions and reviews on an ontology and to know *"who is actually already reusing it"*; as one of the respondents had argued: *"usually you want to be sure you are reusing the ontology that others have also chosen to reuse"*. Some of the respondents had mentioned trust and had argued that it is difficult to *"trust an ontology and [its] background research"* or *"to judge whether a 'popular' ontology is good or not"*.

Ontology development and modelling process and its effects on quality had also been discussed. The respondents had argued that *"existing ontologies are built in different styles",* and it has led to a different range of quality. As one of the respondents had put it: *"there is no universal training of ontologists, so there are no commonly held best practices, and they come from all walks of life with different backgrounds and focus. This leads to huge variations in the quality of ontologies, large enough to render them useless"*. Some other respondents had argued that *"Nearly all 'ontologies' are just taxonomies where most of the definitions are implicit in the comments and term names"* and *"there are not that many [ontologies] encoding common sense knowledge"*.

Overall, the findings of this section confirmed one of the early assumptions of this research - the critical role and the importance of ontology quality and evaluation in the selection process for reuse.

5.4.3 Ontology Evaluation

One of the very important objectives of this survey study was to identify a set of metrics that can be used in the evaluation process; this survey also needed to determine how important each of those metrics were. To do that, respondents were presented with different sets of qualitative and quantitative questions. The following section presents the results in detail.

5.4.3.1 Quality Metrics for Ontology Evaluation-Qualitative Data

In the evaluation part of the survey, the respondents were first asked an open-ended question about how they evaluate the quality of an ontology before selecting it for reuse. This question aimed to gather respondents' opinions on different evaluation metrics and approaches, before presenting them with the previously identified ones. The responses to this question were coded according to different categories of quality metrics, namely, (1) internal, (2) metadata, (3) social, community, and popularity.

According to the analysis, quality metrics thought to be the most important were content and coverage (mentioned 51 times) and documentation (mentioned 41 times). If an ontology has been reused previously and the popularity of the ontology on the web, or amongst the community were the other frequently mentioned metric by the respondents (38 times). The community judgment on the quality of an ontology, activeness and responsiveness of the ontology developer team, and the reputation of them or the organisation responsible for an ontology were also mentioned by many of the participants (25 times).

Interestingly, 19 respondents had mentioned following or complying with different design guidelines and principles or being a part of standards, like W3C and OBO Foundry, as one of the features they would consider before selecting an ontology for reuse. Some had also stated that they check if an ontology is built by using any methodology like NEON (Suárez-Figueroa, Gómez-Pérez and Fernández-López, 2012). Respondents had ranked “The use of a method /methodology (e.g. NEON, METHONTOLOGY, or any other standard and development practice)” statement 29th (out of 31), with a mean of 2.8 and a median of 3 from the Likert Scale responses.

Ontology size was also mentioned and linked to the other quality related factors, such as coverage, level of specialisation and modularity. According to one of the respondents, while evaluating an ontology they make sure that *"coverage of the ontology is adequate for the particular notions we need to model, but not too big that it defines too many notions not relevant for our needs"*. The other respondent had mentioned that they would ask if an ontology "*is big enough that it worth linking to for general interoperability or specialised enough that it replaces a significant amount of the work in designing my own material"*.

Despite all the research on ontology evaluation and its affecting factors, one of the respondents had blamed the literature for lack of adequate criteria for ontology evaluation and had stated: *"we only reuse ontological patterns that have been tested over time in multiple projects. The evaluation criteria proposed by the current literature are, in my opinion, not adequate"*. Some other respondents had also argued that they know the ontology they wanted to reuse, and they do not need to evaluate it.

Besides discussing the characteristics of a reusable ontology, one of the participants had mentioned the characteristics of an ontology that they would not reuse:

"If it is a taxonomy only, I will not use it. If it has no structure, I will not use it. If it doesn't use class inheritance (class specialisation), I will not use it. If it embeds its logic in 'owl:equivalentClass' relations that aren't classes, I will not use it. If it embeds as its definition other languages, such as CLIF, I will not use it. If it doesn't use namespace prefixes, I will convert it".

Most of the criteria mentioned here had already been identified by this research and had been included in the quantitative part of the survey. However, two of the metrics, namely, "fit" and "format" were not covered by the Likert items. The format of an ontology was only mentioned two times, but fit was mentioned 37 times. Despite its significance, how well an ontology fits different selection requirements cannot be an indication of its quality. Using OOPS and loading the ontology into a software application were amongst the other evaluation approaches discussed in this section.

5.4.3.2 Quality Metrics for Ontology Evaluation-Quantitative Data

Survey respondents were presented with four different sets of Likert items, each, including metrics that can be used to evaluate the importance of different aspects of ontologies, namely, internal, metadata, social and community and popularity. They were then asked to rate how important they thought each of those metrics was, choosing from a five-point scale ranging from "not important" to "very important". This section aimed to identify the most important or the most used metrics for ontology evaluation. It also helped in confirming the new metrics that had been identified in the interview study.

5.4.3.3 The Most Important/Used Metrics for Ontology Evaluation

Descriptive statistics were used to analyse the ratings assigned to different quality metrics presented in this section and also to address one of the critical questions of this research and to identify the most important metrics used in the evaluation process. Considering the question type used in this section and the type of data that was collected, nonparametric tests were mostly applied for analysis (Corder and Foreman, 2014).

Table 5-2 shows the descriptive statistics of all the 31 quality metrics, sorted by mean. The metrics are ranked from 1 to 31, with 1 being the most important and 31 being the least important metric considered when evaluating the quality of an ontology for reuse. Median and mean are used to show the midpoint and centre of the data, respectively. Standard deviation is used to express the level of agreement on the importance of each metric in the ontology evaluation process; the lower value of standard deviation represents the higher level of agreement among the survey respondents.

As it is seen in Table 5-2, ontology content is the first feature ontologists, and knowledge engineers tend to look at when evaluating the quality of an ontology for reuse. Other internal aspects of ontologies like their structure, scope, syntactic correctness, and consistency are also amongst the top ten quality metrics used in the evaluation process.

According to Table 5-2, participants in this survey have given a very high rating to some of the metadata related metrics. Availability and accessibility, for example, is the second most important criterion used to assess the quality of ontologies. Survey respondents have also given a very high rating, four and eight respectively, to other metadata related metrics, such as documentation and availability of metadata and provenance information about an ontology. However, other criteria in the metadata group, e.g., availability of funds for ontology update and maintenance and use of a method/methodology were among the bottom ten least important metrics.

Table 5-2 Descriptive statistics of all the quality metrics in the survey

Rank	Metric	Mean	SD	Median
1	The Content (classes, properties, relationships, individuals, axioms)	4.59	0.57	5
2	The ontology is online, accessible, and open to reuse (e.g., License type)	4.52	0.85	5
3	The Scope (domain coverage)	4.42	0.84	5
4	The availability of documentation (both internal, e.g., adding comments and external)	4.38	0.79	5
5	The Structure (class hierarchy or taxonomy)	4.29	0.82	4
6	The Syntactic Correctness	4.15	0.92	4
7	The Consistency (e.g., Naming and spelling consistency all over the ontology)	4.03	1	4
8	Availability of metadata and provenance information about the ontology	3.92	1.01	4
9	The Semantic Richness and Correctness (e.g., level of details)	3.92	1.06	4
10	Having information about the purpose that ontology is used/has been used for (e.g., annotation, sharing data, etc.)	3.77	1.03	4
11	The Language that ontology is built in (e.g., OWL)	3.7	1.3	4
12	Having an active responsive (developer) community	3.62	1.09	4
13	Availability of published (scientific) work about the ontology	3.56	1.19	4
14	The popularity of the ontology in the community and among colleagues	3.51	1.17	4
15	Availability of wikis, forums, mailing lists and support team for the ontology	3.45	1.03	4
16	Knowing and trusting the ontology developers	3.42	1.11	4
17	The flexibility of the Ontology (being easy to change) and the ontology developer team	3.41	1.14	4
18	The number of times the ontology has been reused or cited (e.g., owl:imports, rdfs:seeAlso, daml:sameClassAs)	3.4	1.13	3
19	Knowing and trusting the organisation or institute that is responsible for ontology development	3.38	1.11	3
20	Having information about the other projects that the ontology is used/has been used in	3.34	1.1	3
21	The reputation of the ontology developer team, and/or institute in the domain	3.31	1.12	3
22	The frequency of updates, maintenance, and submissions to the ontology	3.22	1.16	3
23	The number of updates, maintenance, and submissions to the ontology	3.13	1.19	3
24	Having information about the other individuals or organisations who are using/have used the ontology	3.12	1.1	3
25	The number of times the ontology has been reused or cited (e.g., owl:imports, rdfs:seeAlso, daml:sameClassAs)	3.08	1.19	3
26	The popularity of the ontology on the web (number of times it has been viewed in different websites/applications across the web)	3.05	1.24	3
27	The reviews of the ontology (e.g., ratings)	3.03	1.25	3
28	The size of the ontology	3.02	1.19	3
29	The use of a method /methodology (e.g., NEON, METHONTOLOGY, or any other standard and development practice)	2.8	1.26	3
30	The availability of funds for ontology update and maintenance	2.77	1.23	3
31	The popularity of the ontology in social media (e.g., in GitHub, Twitter, or LinkedIn)	2.28	1.16	2

Community related metrics had some very interesting ratings too. The results indicated that ontologists and knowledge engineers would like to know about the purpose that an ontology is used/has been used for while evaluating it. Having an active, responsive developer community and knowing and trusting the ontology developers were among the other top-ranked community related aspects of ontologies. “Availability of wikis, forums, mailing lists and support team for the ontology” was also ranked 15th, with SD of 1.03, which shows the importance of this metric for the respondents of this survey.

As it is seen in Table 5-2, the popularity of an ontology in the community and amongst colleagues has the highest median and mean compared to the other popularity related metrics. The findings also suggested that ontologists tend to consider the reputation of the ontology developer team and/or institute in the domain while evaluating it for reuse. Popularity metrics previously identified in the literature, namely, the popularity of the ontology in social media (Martínez-Romero *et al.*, 2014), and the popularity of the ontology on the web (Martínez-Romero *et al.*, 2017) were amongst the metrics with the least median and mean.

5.4.4 Factor Analysis

After identifying the most important metrics used in the evaluation process, the next step was to conduct factor analysis. Factor analysis is one of the most used multivariate statistical procedures in research (Brown, 2014) and helps to "move from a large set of variables (the items) to a smaller set (the factors) that does a reasonable job of capturing the original information" (DeVellis, 2012, p. 127). Factor analysis has also been referred to as a category of procedures that helps in "reorganisation of a substantial amount of specific information into a more manageable set of more general but meaningful categories" (DeVellis, 2012, p. 116).

There are two main types of factor analysis: (1) exploratory factor analysis (EFA) and (2) confirmatory factor analysis (CFA). Both EFA and CFA aim to "reproduce the observed relationships among a group of indicators with a smaller set of latent variables" (Brown, 2014). However, EFA is more focused on exploring data and has no prior assumption about the number of factors or their loadings, whereas, in CFA, a set of predefined hypothesis and assumptions about the number of factors and their loadings are specified in advance (Suhr, 2006; Brown, 2014). The set of steps usually followed while conducting factor analysis includes: (a) selecting the variables, (b) computing correlations matrix amongst those variables, (c) extracting the unrotated factors, (d) rotating the factors, and (e) interpreting the rotated factor matrix (Comrey and Lee, 1992).

There are many different reasons for undertaking factor analysis, e.g., "determining how many latent variables underline a set of items" (DeVellis, 2012), identifying a set of constructs that "might be used to explain the intercorrelations among these variables" (Comrey and Lee, 1992), and summarizing the collected data to identify and understand potential patterns and relationships (Yong and Pearce, 2013). Factor analysis was used in this study for two main reasons, namely, to explore and identify the potential correlations among the quality metrics suggested by this research and also, to summarize the findings and to determine a number of latent variables that can be used to evaluate ontologies.

Sample size has been considered as one of the most critical issues to address before starting factor analysis. Some have suggested larges samples as the condition for a reliable factor analysis research; however, Maccallum *et al.* (1999) argued that linking the reliability of a study to its sample size is incorrect, and stated: "under some conditions, relatively small samples may be entirely adequate, whereas under other conditions, very large samples may be inadequate".

According to a survey conducted by Costello and Osborne (2005) on the best practices in EFA, subject to item ratio of most of the studies they reviewed (62.9%) was 10:1 or less. This study identified 31 metrics and manage to collect 157 complete responses, which means having subject to item ratio of 5:1, which is within an acceptable range. It worth mentioning that finding and contacting experts in the ontology domain is very challenging and much of the research in this field has been based on smaller samples; the recent study by Matentzoglu *et al.* (2018), for example, was based on 110 responses. The rest of this section presents the factor analysis conducted in this study

5.4.4.1 Identification of the Factors Used in Ontology Evaluation

The first step in conducting EFA is to examine the correlation between different variables (Comrey and Lee, 1992). For a study to be analysed using factor analysis, the correlation between different variables should exceed 0.3; moreover, and for the sake of validity, communalities value should be higher than 0.4 (Costello and Osborne, 2005). In running the EFA, the following methods and approaches were applied using SPSS (IBM, 2016):

- Correlation Matrix: Coefficient, significance levels, Kaiser-Meyer-Olkin (KMO) and Bartlett's test of sphericity.
- Extract factors using Principle Component Analysis and based on eigenvalues greater than 1 (Kaiser, 1960).
- Rotation Method: Varimax
- Method for Factor Scores: Regression

The Kaiser-Meyer-Olkin Measure of Sampling Adequacy in KMO and Bartlett's test had a high value of 0.842 with a significance of 0.000 (χ^2 (465, N = 31) = 2248.047, $p < .05$), which meant that the data was suited for factor analysis and the instrument was reliable and internally consistent. The very high score of Cronbach's Alpha (.909) also confirmed the internal consistency of the survey items. Moreover, the correlation matrix generated for the 31 metrics prove some powerful correlations between them and the communalities value for all of the variables were also higher than 0.5.

After computing correlations between different variables, the second step was to determine the number of factors that should be extracted. Eigenvalue rule and scree plot are two of the most widely nonstatistical guidelines used in the literature to help to choose the right number of factors that should be extracted (DeVellis, 2012). Eigenvalue indicates the amount of variance explained (Suhr, 2006) and information captured (DeVellis, 2012) for each factor. This study followed Kaiser's eigenvalue rule; therefore, factors with the eigenvalue of less than 1 were eliminated (Kaiser, 1960). A scree plot was also used to provide a graphical view of the right number of factors that should be extracted (Thompson, 2004, p. 33). As it is seen in Figure 5-1, around 69% of the total variance could be explained by nine factors with the eigenvalue of more than one.

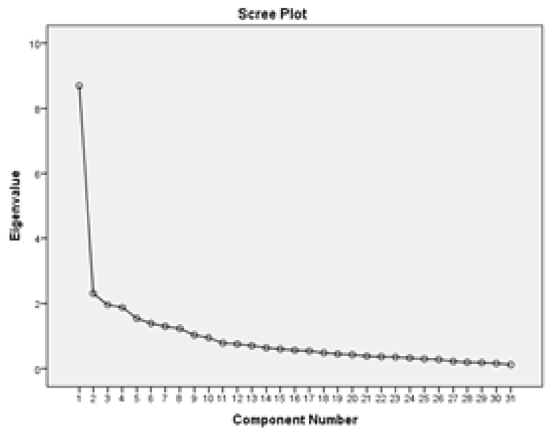

Figure 5-1 Scree Plot

The nine components identified in the previous step were then rotated using Varimax rotation as the method. Rotation helps in clarifying the factors that each variable belongs to and makes factor naming easier (Seiler, 2004, p. 177). In this study, rotation helped to determine the relationship between each variable, that was a quality metric for ontology evaluation, with each of the resulting factors.

Table 5-3 presents what is known as "loadings", that refers to the extent that different variables are related to the hypothetical factor(s) (Comrey and Lee, 1992, p.5); higher loadings are usually preferred in the literature. Comrey and Lee (1992, p. 243) has proposed the following guidelines for loading interpretation: 0.32=poor, 0.45=fair, 0.55=good, 0.63=very good, and 0.71=excellent. Some variables might also have negative loadings on some factors, which means that they are negatively correlated to the factor construct (ibid.). In this study, loadings less than 0.4 were eliminated from the rotated component matrix.

Table 5-3 Loadings-9 Factors

Factor	Item	Loading
Factor 1	QM4_1_Number_Of_Times_Ontology_Been_Reused	.713
	QM4_2_Popularity_On_Web_Website_Views	.771
	QM4_3_Popularity_In_Community_Among_Colleagues	.665
	QM4_4_Popularity_Ontology_Social_Media	.700
	QM4_6_Reviews_Rating_Of_Ontology	.550
	QM2_4_Reuse_Import	.440
Factor 2	QM3_2_Knowing_Trusting_Ontology_Developers	.832
	QM3_3_knowing_trusting_ontology_development_organization	.709
	QM3_4_Flexibility_Ontology_&_Developer_Team	.559
	QM4_5_Reputation_Developer_Team_Institute	.595
Factor 3	QM2_7_Number_of_update_maintenance	.832
	QM2_8_Frequency_Update_Maintenance	.820
	QM2_9_Funds_Availability_Update_Maintenance	.635
	QM3_1_Active_Responsive_Community	.509
Factor 4	QM3_5_Extra_Info_Usage_Individuals_Organisations	.690
	QM3_6_Extra_Info_Usage_Projects	.760
	QM3_7_Extra_Info_Usage_Purpose	.793
Factor 5	QM1_3_Structure	.626
	QM1_4_Semantic_Richness_&_Correctness	.768
	QM1_5_Syntactic_Correctness	.620
	QM1_6_Consistency	.520
Factor 6	QM2_10_Accessibility	.840
	QM3_8_Availability_Wikis_Forums_MailingLists_SupportTeam	.533
Factor 7	QM2_2_Documentation	.645
	QM2_3_Availability_of_metadata	.725
	QM2_5_Language	.432
Factor 8	QM1_1_Scope	-.530
	QM2_1_Methodology	.511
	QM2_6_Size	-.440
	QM2_11_Availability_Publication	.571
Factor 9	QM1_2_Content	.853
	QM1_3_Structure	.447

As it is seen in Table 5-3, the first component identified in factor analysis included all the popularity related metrics used in the survey. Metrics related to the trust and reputation of the developer team or organisation were highly loaded on the second reliable component. The third component included metrics that referred to the responsiveness of the developer team and the maintenance process. The rest of the components had grouped different metrics related to the internal aspects of ontologies, their accessibility, and additional information about them.

After identifying and extracting factors, the next step was to interpret them; researchers should consider different reliability issues before starting the interpretation process. They should also ask questions like “what is the potential value of this factor?” or “do the variables that define the factor reveal all its major aspects?” (Comrey and Lee, 1992). Reliability of different factors can also be tested by the number of variables that are loaded on them and the absolute value of the loadings. Some have recommended that each desirable factor should at least include three variables (Yong and Pearce, 2013).

Stevens (2009, p.333) suggested that “components with four or more loadings above 0.6 in absolute value are reliable, regardless of sample size”. He also argued that a factor can be considered reliable if “the average of the four largest loadings is > 0.60 or the average of the three largest loadings is $> .80$”. Reliability determination gets more complicated when the absolute value of loadings is lower.

As it is seen in Table 5-3, the loadings of different variables identified in this study are generally high and can mostly be rated as good or very good, sometimes even excellent (Comrey and Lee, 1992). However, the researcher had to deal with the problem of not having enough variables loaded on some of the factors. For example, two variables, namely, “accessibility” and “availability of online mailing list and support team” were highly loaded on component 6; however, having only two variables loaded on this component made it very difficult, if not impossible, to interpret it; also, it did not meet the minimum reliability requirements.

According to the literature, researchers should be cautious when interpreting factors that are based on a few low loadings variables (Comrey and Lee, 1992). To address the harmful effects of under extracting or over extracting factors, Costello and Osborne (2005) has suggested using scree plot, manually setting the number of items to retain and conducting multiple factor analysis, until identifying “best fit”; best fit has been defined as a model with no factor with less than three variables, with the minimum loading value of .3 for different items, and no or few items that cross-load on different factors (ibid.).

In this study, following Kaiser’s eigenvalue rule (Kaiser, 1960) led to having nine factors, some of which had reliability issues, e.g., did not include enough variables loaded on them. Crossloading, the situation in which a variable is loaded on two or more factors (Yong and

Pearce, 2013), was the other issue; size, for example, was fairly loaded on component 1 and 8, with the value of 0.488 and -0.440 respectably. However, it was difficult to link size to the popularity factor (component 1), as this factor is more about how ontologies are used in the community or how many times they are used. Thus, it was decided to add it to factor 8.

To address some of the discussed issues, it was decided to keep all the variables, even though some of them were not forming any reliable factor, and to re-conduct factor analysis (Costello and Osborne, 2005).

5.4.4.2 Rerunning EFA

EFA was reconducted by reducing and increasing the number of factors. Increasing the number of extracted factors to 10 resulted in one of the most unfeasible situations, with 2-3 components that only had two variables loaded on them. EFA was then rerun using five, six, seven, and eight variables. Extracting five factors led to having a large number of problematic items, some of which were loaded on 2 or even three factors or were not loaded on any factor at all; same results were obtained when six factors were extracted.

While repeating EFA with 8 factors, two of the variables, namely, "consistency" and "reuse_import" were dropped from the initial list as the result of crossloading. Reconducting EFA with 29 variables and eight factors did not solve the crossloading problem and also, led to having a factor with only two variables loaded on it. Thus, a model with eight factors was rejected.

All the 31 items were used again in an EFA, this time with seven factors to be extracted. According to the rotated component matrix table, two of the variables, namely, "consistency" and "availability of wikis, forums, mailing lists and support team for the ontology" were not loaded on any factor; thus, they were removed from further analysis. Maximum iterations for convergence value was also increased to 50, as 25 would cause a "Rotation failed to converge in 25 iterations" error. The final rotating component matrix is presented in Table 5-4. As it is observed, reducing the number of factors has increased the reliability of factors obtained, as well as the number of variables loaded on each factor.

EFA based on seven factors improved some of the issues faced earlier; however, some of the extracted factors were still unreliable. To increase the reliability of factors and subject to item ratio, subset EFA were conducted. The results of these analyses are presented in the following section.

Table 5-4 EFA with 7 Factors

Factor	Item	Loading
Factor 1	QM2_6_Size	.504
	QM2_11_Availability_Publication	.416
	QM4_1_Number_Of_Times_Ontology_Been_Reused	.684
	QM4_2_Popularity_On_Web_Website_Views	.717
	QM4_3_Popularity_In_Community_Among_Colleagues	.656
	QM4_4_Popularity_Ontology_Social_Media	.746
	QM4_6_Reviews_Rating_Of_Ontology	.652
Factor 2	QM2_7_Number_of_update_maintenance	.850
	QM2_8_Frequency_Update_Maintenance	.825
	QM2_9_Funds_Availability_Update_Maintenance	.581
	QM3_1_Active_Responsive_Community	.532
Factor 3	QM3_2_Knowing_Trusting_Ontology_Developers	.808
	QM3_3_Knowing_Trusting_Ontology_Development_Organization	.713
	QM3_4_Flexibility_Ontology_&_Developer_Team	.584
	QM4_5_Reputation_Developer_Team_Institute	.592
Factor 4	QM2_2_Documentation	.460
	QM3_5_Extra_Info_Usage_Individuals_Organisations	.642
	QM3_6_Extra_Info_Usage_Projects	.718
	QM3_7_Extra_Info_Usage_Purpose	.751
Factor 5	QM1_2_Content	.649
	QM1_3_Structure	.765
	QM1_4_Semantic_Richness_&_Correctness	.669
Factor 6	QM1_5_Syntactic_Correctness	.692
	QM2_5_Language	.618
	QM2_10_Accessibility	.722
Factor 7	QM1_1_Scope	-.493
	QM2_1_Methodology	.564
	QM2_3_Availability_of_metadata	.481
	QM2_4_Reuse_Import	.526

5.4.4.3 Subset Analysis

The 31 identified metrics were divided into two subsets for further analyses. As it is seen in Table 5-5 and Table 5-6, subset analyses led to the same number of factors, which was seven. However, all of the identified factors satisfied the minimum reliability requirements proposed by Stevens (2009, p. 333); the subject to item ratio was also enhanced as the result of reducing the number of factors. The following part of this chapter describes EFA results in greater detail.

Subset 1. This subset contained two main groups of metrics that could be used in the evaluation process: (1) metrics related to internal aspects of ontologies, and (2) metrics related to different metadata that can be used to provide additional information about ontologies. Before grouping internal and metadata related items in one subset, two different EFAs were conducted on each of these groups; both led to unreliable factors. Therefore, it was decided to group and analyse them together. Initially, EFA was conducted with 17 items; two of the items "availability of publication" and "syntactic correctness" were removed as they had a fair loading value on more than one factor. As it is seen in Table 5-5, the second run of EFA resulted in a model with four reliable factors.

Table 5-5 EFA Subset Analysis (1)

Factor	Item	Loading	Factor Name
Factor 1	QM3_5_Extra_Info_Usage_Individuals_Organisations	.803	Usage Information
	QM3_6_Extra_Info_Usage_Projects	.863	
	QM3_7_Extra_Info_Usage_Purpose	.817	
Factor 2	QM1_2_Content	.676	Internal
	QM1_3_Structure	.792	
	QM1_4_Semantic_Richness_&_Correctness	.651	
	QM1_6_Consistency	.430	
Factor 3	QM2_1_Methodology	.531	Metadata
	QM2_2_Documentation	.483	
	QM2_3_Availability_of_metadata	.775	
	QM2_4_Reuse_Import	.651	
Factor 4	QM1_1_Scope	.588	Other Metadata
	QM2_5_Language	.629	
	QM2_6_Size	.608	
	QM2_10_Accessibility	.637	

Subset 2. This subset contained all the criteria that can be used to measure the social and community aspects of ontologies. It was also decided to include the "ontology is based on (or has reused) other ontologies (e.g. Import)" item, as it refers to one of the definitions of popularity in the literature. However, and according to the initial analysis, this item had a very low extraction value (0.216) and did not load on any of the identified factors; thus, it was removed. All the remaining 14 items were then included in a single EFA. The items and their loadings on each factor are presented in Table 5-6. Similar to the previous rounds of factor analysis, loadings with the value of less than 0.4 were removed, and the same reliability rules were applied.

Subset EFA analyses led to the same number of factors (seven), all of which were reliable. Therefore, it was decided to keep the results of the subset analysis and use them in the next rounds of data analysis, which included testing different hypotheses.

Table 5-6 EFA Subset Analysis (2)

Factor	Item	Loading	Factor Name
1	QM4_1_Number_Of_Times_Ontology_Been_Reused	.794	Popularity
	QM4_2_Popularity_On_Web_Website_Views	.826	
	QM4_3_Popularity_In_Community_Among_Colleagues	.744	
	QM4_4_Popularity_Ontology_Social_Media	.689	
	QM4_6_Reviews_Rating_Of_Ontology	.695	
2	QM2_7_Number_of_update_maintenance	.858	Maintenance & Responsiveness
	QM2_8_Frequency_Update_Maintenance	.861	
	QM2_9_Funds_Availability_Update_Maintenance	.663	
	QM3_1_Active_Responsive_Community	.660	
	QM3_8_Availability_Wikis_Forums_MailingLists_SupportTeam	.400	
3	QM3_2_Knowing_Trusting_Ontology_Developers	.853	Community, Reputation & Trust
	QM3_3_Knowing_Trusting_Ontology_Development_Organization	.755	
	QM3_4_Flexibility_Ontology_&_Developer_Team	.427	
	QM4_5_Reputation_Developer_Team_Institute	.691	

5.4.5 Hypotheses Testing for Quality Metrics

This research aimed to clarify if the years of experience, type of organisation and domain of ontologists and knowledge engineers affect how they evaluate ontologies, their choice of metrics and the importance they assign to each metric. To address these questions, different hypotheses were tested. The Shapiro-Wilk test was conducted to check the normality of all of the seven identified factors. The results showed that none of the factors was normally distributed. Therefore, the Kruskal Wallis test was the appropriate test for this data. Kruskal Wallis H is a non-parametric test with no assumptions about the normality of data (Hecke, 2012) and is usually used to make a comparison among three or more independent groups (Vargha and Delaney, 1998).

5.4.5.1 Hypothesis 1: Ontologists and knowledge engineers with different years of experience rank the importance of quality metrics differently

As it is seen in Table 5-7, participants in the second phase of this study had varying years of experience in building and reusing ontologies. Kruskal-Wallis and Median tests for K

Table 5-7 Participants' Years of Experience

Years of Experience	Frequency	Percent
Less than 3 years	14	8.9
3 to 5 years	28	17.8
5 to 10 years	42	26.8
More than 10 years	73	46.5

Independent Samples were conducted to check whether ontologists and knowledge engineers with different years of experience evaluate ontologies differently.

- **H_0:** there is **no** statistically significant difference between the years of experience ontology developers have and how they rank the importance of quality metrics for ontology evaluation.
- **H_1:** there is **a** statistically significant difference between the years of experience ontology developers have and how they rank the importance of quality metrics for ontology evaluation.

Both Kruskal-Wallis and Median tests results indicated that people with different years of experience tend to rate the importance of two (out of seven) quality factors, "internal" and "popularity", differently. According to the results, the distribution of "popularity" rating is statistically significantly different among people with different years of experience ($\chi^2(3) = 14.655, p = 0.002$). The two groups who would rate this factor differently were the participants with less than three years of experience (Median=4) and the ones with more than ten years of experience (Median=2.6), U= 240, p = 0.002, *r*=-.336.

Kruskal-Wallis test also indicated that those with three to five years of experience would rate the internal related metrics differently compared to ones with five to ten and more than ten years of experience ($\chi^2(3) = 10.253, p = 0.017$). According to Mann-Whitney U test, respondents with three to five years of experience tend to rate internal related factor higher than ones with five to ten (*U*=353, *p*=0.004, *r*=-0.341) or more than ten years of experience (*U*=643.5, *p*=.004, *r*=-0.228). Medians of the importance of Internal factor for the above-mentioned groups were 4.625, 4.25, and 4, respectively.

5.4.5.2 Hypothesis 2: Ontologists and knowledge engineers working in different domains rank the importance of quality metrics differently

To test this hypothesis, the summated mean of the seven factors identified in subset EFA analysis were compared across different domains, namely, biomedical, non-biomedical and some biomedical.

- **H_0:** there is **no** statistically significant difference between the domain people build ontologies in and how they rank the importance of quality metrics for ontology evaluation.

- **H_1:** there is **a** statistically significant difference between the domain people build ontologies in and how they rank the importance of quality metrics for ontology evaluation.

The median test results showed that there was no statistically significant difference between the median of different factors across different domain except for the "Internal" and "Other metadata" factors, where the null hypothesis was rejected and there existed a statistically significantly difference between the median of these two factors across different domains.

According to a Mann-Whitney U test (U= 72.5, p = 0.014), respondents working on biomedical ontologies would rank internal related quality metrics higher than those working on some biomedical ontologies. Participants involved in building non-biomedical ontologies would also give a significantly statistically higher rank to "Other Metadata" factor compared to those who were building ontologies in different domains (general), U= 282.5, p = 0.013.

Kruskal-Wallis test results showed no statistically significant difference between the distribution of most of the quality-related factors across the domains. Therefore, the null hypothesis was retained for all the identified factors except the two of them. According to the results, the distribution of "Maintenance and Responsiveness" rating was statistically significantly different across different domains ($\chi^2(3) = 9.372, p = 0.025$). A Mann-Whitney test indicated that respondents who were building ontologies in the biomedical domain would rate "Maintenance and Responsiveness" factor statistically significantly higher than those who build ontologies in different domains, *U*=70.5, *p*=0.011.

The difference between the distribution of "Internal" factor was also statistically significant across different domains, $\chi^2(3) = 11.991, p = 0.007$. Mann-Whitney test showed that participants who were building non-biomedical ontologies would rate "Internal" factor statistically significantly higher than those who were building some biomedical ontologies, *U*=214.5, *p*=0.002.

5.4.5.3 Hypothesis 3: Ontologists and knowledge engineers working in different organisations rank the importance of quality metrics differently

Survey respondents were working in different types of organisations, namely, universities, companies, industry, and NGOs. Some of them were also working in more than one organisation, e.g. both university and company. Three main types of organisations, namely, academia, non-academia and some academia were identified (Table 5-8). Academia included ontologists and knowledge engineers who were only working in universities and research institutes. "Some Academia" referred to the group of ontologists and knowledge engineers who were working both in academia, and other types of organisation (e.g., companies). "Non Academia" group included the ontology developers who had no academic experience and were working in different companies and industries.

Table 5-8 Participants' Organisation Type

Organisation Type	Frequency	Percent
Academia	108	68.8
Non Academia	41	26.1
Some Academia	8	5.1

- **H_0:** there is **no** statistically significant difference between the type of organisation ontology developers work for and how they rate the importance of quality metrics for ontology evaluation.
- **H_1:** there is **a** statistically significant difference between the type of organisation ontology developers work for and how they rate the importance of quality metrics for ontology evaluation.

To test this hypothesis, Kruskal-Wallis and Median test for K Independent Samples were conducted. The results of these tests indicated that respondents working in different types of organisations tend to give the same importance to different quality metrics for ontology evaluation, meaning that the type of organisation does not affect how ontologies are evaluated. Therefore, the null hypothesis was retained for all of the seven factors identified in the subset EFA analysis.

5.4.6 The Role of Community in Ontology Evaluation, Selection and Reuse

To explore what survey respondents thought about the role of community in the ontology domain, they were presented with "Communities play a critical role in ontology engineering field" statement and were asked to state their level of agreement with it. According to the results, 129 out of 157 (82.1%) of the respondents agreed or strongly agreed with this statement; only 5% of the survey respondents strongly disagreed or disagreed with it. The next statement was about the ontology development process and asked how much in agreement or disagreement survey respondents were with the "Building ontology is difficult in isolation, and I prefer to work and collaborate with other people" statement. According to the results, 114 out of 157 (72.6%) of respondents agreed or strongly agreed with this statement.

Survey participants were also asked if they find it useful to know the people in their community and to follow their work. Similar to the previous community-related statements, only a small number of respondents (1.9%) expressed their disagreement that it was useful. Moreover, only eight out of 157 (5.1%) of the survey respondents disagreed or strongly disagreed with the "I think that it is useful to find and know people with similar experience to mine" statement.

The next two statements presented in the survey aimed to examine if the connection and relationships among ontologists and knowledge engineers can affect their judgment about the quality of an ontology. Only 14.6% (23 out of 157) of those who responded to the survey expressed their disagreement with the "I trust other people's judgment about an ontology and I believe that ontologies can be evaluated via community feedback and reviews" statement. Other participants expressed some level of agreement with it. Similar results were obtained for the "There are some ontologies that are not the best, but I use them anyhow because other people are using them and I want to standardise with them" statement, where only 19.2% of the respondents (27 out of 157) reported some level of disagreement.

Overall, the very high level of agreement with community-related statements and the high ratings of community-related Likert items for ontology evaluation proves the importance of one of the understudied aspects of ontologies, that is the role of community and social interactions.

5.4.6.1 Hypotheses Testing for the Role of Community

This research was interested in identifying if different demographic characteristics of survey respondents influence their views towards the role of the community in ontology evaluation and reuse; therefore, Kruskal-Wallis and Median tests were conducted. The following part of this section presents the results.

As it is seen in Table 5-9, “I trust other people's judgment about an ontology, and I believe that ontologies can be evaluated via community feedback and reviews” is the only statement which distribution is not the same across different categories of experience years, $\chi^2(3) = 22.248, p = 0.000$. According to the Mann-Whitney U test, participants with more than ten years of experience showed a lower level of agreement (median=3) to this statement compared to people with three to five years of experience, *U*=615, *p*=.001, *r*=-0.326. Another Mann-Whitney test (*U*=199, *p*=.000, *r*=-0.411) indicated that ontologists and knowledge engineers with less than three years of experience had given a higher rating to this statement compared to people with more than ten years of experience, with median of 4 and 3 respectively.

Kruskal-Wallis test was also conducted to measure if the domain ontologists and knowledge engineers develop ontologies in can significantly affect how they think about the role of community in ontology evaluation and reuse. The median value of four (out of eight) statements were statistically significantly different across domains. Multiple Mann-Whitney tests were conducted to identify the categories that had statistically significant differences. The median of ratings for the first statement, that was “I prefer to contribute to existing ontologies instead of developing a new one”, was significantly different between respondents working in the biomedical domain compared to those who were building not only biomedical ontologies, but also other types of ontologies, 4 and 3 respectively with *U*=59.5, *p*=.004, *r*=-0.469.

Respondents with “Some Biomedical” experience had also assigned a lower level of agreement to the “Building ontology is difficult in isolation, and I prefer to work and collaborate with other people” statement, compared to those who were building biomedical or general ontologies, with a median of 3.5, 5, and 5 respectively. Moreover, a Mann-Whitney test showed that participants working on biomedical ontologies believe that community plays a more important role in the ontology domain compared to those working on some biomedical ontologies, a median of 5 and 4 respectively, and *U*=75, *p*=.018, *r*=-0.411.

Similar results were obtained when comparing the median of ratings survey respondents working in biomedical domain had assigned to the second statement compared to the median of ratings those in non-biomedical domain had assigned to that statement, 5 and 4 respectively;

Mann-Whitney U test confirmed that the difference between the median of these two groups is statistically significant, U=1054, p=.002, r=-0.27.

Table 5-9 Years of Experience, Domain and Organisation Type, and Role of Community

Statements	Years of Experience	Ontology Domain	Organisation Type
I prefer to contribute to existing ontologies instead of developing a new one	Null Hypothesis Retained	$\chi^2(3) = 11.654, p = 0.009$	Null Hypotheses Retained
Building ontology is difficult in isolation, and I prefer to work and collaborate with other people	Null Hypothesis Retained	$\chi^2(3) = 14.287, p = 0.003$	
Communities play a critical role in ontology engineering field.	Null Hypothesis Retained	$\chi^2(3) = 11.644, p = 0.009$	
It is important to know the people in my community and to follow their work	Null Hypothesis Retained	Null Hypothesis Retained	
I trust other people's judgment about an ontology, and I believe that ontologies can be evaluated via community feedback and reviews.	$\chi^2(3) = 22.248, p = 0.000$	$\chi^2(3) = 9.287, p = 0.026$	
There are some ontologies that are not the best, but I use them anyhow because other people are using them and I want to standardise with them	Null Hypothesis Retained	Null Hypothesis Retained	
I think that it is useful to find and know people with similar experience to mine	Null Hypothesis Retained	Null Hypothesis Retained	
In my domain, ontologies are not built to be shared and reused (because of the nature of the domain, intellectual property and/or financial concerns)	Null Hypothesis Retained	Null Hypothesis Retained	

The next statement aimed to measure the role of trust in ontology evaluation and asked survey respondents how agree or disagree they were with “I trust other people's judgment about an ontology, and I believe that ontologies can be evaluated via community feedback and reviews" statement. The results indicated that there existed a statistically significantly difference between the level of agreement of the ontologists working on general ontologies compared to those working on non-biomedical and biomedical ontologies, the median of 3, 4 and 4 respectively. It means that ontologists working in biomedical and non-biomedical domain tend to trust the judgment of their peers about the quality of an ontology more.

5.5 Summary

The main aim of this phase of the study was to clarify, confirm and generalize the findings of the first phase. To do that, a large group of ontologists and knowledge engineers were asked about some of their everyday activities, namely, ontology evaluation and selection for reuse. Their views regarding social interactions and the role of community in the ontology domain were also explored.

Ontology reuse was discussed in the first part. Different selection systems for ontologies and their usefulness as well as the challenges ontologists and knowledge engineers face while looking for a reusable ontology were discussed in the second part. The metrics used in the evaluation process and their importance were discussed in the third part of this chapter. Factor analysis was also conducted to identify a smaller set of metrics that can be used in the evaluation process. Finally, different hypotheses were tested to find answers to the last question of this research, which was whether the demographic features of participants affect how they evaluate ontologies.

A summary of the findings is as follows:

- Ontology reuse was found to be very popular amongst the participants of this study
- Small, upper level ontologies such as BFO and DC were chosen as the best ontologies participants in this study had reused
- Despite the availability of different search and selection systems for ontologies, literature and Google were chosen as the main sources of finding ontologies for reuse
- Availability of documentation was found to be very important, both while building reusable ontologies and evaluating them for reuse
- The findings identified three main dimensions namely internal, metadata, and social aspects of ontologies, that can be evaluated while selecting them for reuse

- Each of the identified dimensions included different factors, such as, (1) internal, (2) usage related information, (3) documentation and standardization, (4) other metadata, (5) popularity, (6) maintenance and responsiveness and (7) reputation and trust
- Hypothesis tested in this phase indicated that the process of evaluating and selecting ontologies is very similar across different domains and organisations.

The findings of this phase helped in answering different research questions and clarified the main quality metrics used in the evaluation process and their importance, especially compared to the ones used by the search and selection systems for ontologies. Taken together, these results provided a new understanding for some of the most important processes in the ontology domain, namely, ontology search, evaluation and selection for reuse. These findings will be validated in the next chapter.

Chapter 6: VALIDATION

6.1 Introduction

This chapter aims to review, clarify and validate the findings of the previous two phases. It starts by explaining and clarifying two of the dimensions identified in this research, namely, metadata and social dimension. The metrics that can be used for evaluating each one of those dimensions and also the measures that can be used to quantify and evaluate some of the identified metrics are discussed. The internal dimension and its related metrics have been the main focus of the research within prior literature; therefore, they will not be the subject of any further exploration.

The framework proposed by this research is also presented and discussed in this chapter. The final part of this chapter includes the evaluation of the usefulness of the framework and metrics identified in the first two phases of this study. It starts by reviewing different validation strategies used in the literature and then moves on to explain the two different experiments that were conducted to validate the findings of this study.

6.2 Dimensions and Metrics

An overview of different dimensions, metrics, measures and the value that can be assigned to each of them is presented in Table 6-1. One of the unique and interesting characteristics of the metrics identified by this research is that most of them are query independent. Meaning that to measure the quality of an ontology using these metrics, no predefined query is required.

6.2.1 Metadata Dimension

The metadata dimension includes different sets of additional information that participants in the first two phases of this study found to be helpful and sometimes essential in the evaluation and selection process. A more detailed account of different metrics of the metadata dimension is explained in the following section.

- **Usage Related Information (Factor 1).** In broad terms, usage information refers to the set of information that helps in exploring how an ontology has previously been (re)used or is currently used. It includes information about the individuals and organisations that have or are currently employing an ontology, the projects that an ontology has been/is being used in, and the purpose that an ontology has been/is being used for. Usage related information is provided by some of the selection systems in the literature (e.g., BioPortal). However, they do not state the purpose (e.g., annotation) that an ontology has been (re)used for; this information is not considered in the evaluation, ranking and recommendation process (Martínez-Romero *et al.*, 2017). According to the findings of this study, however, ontologists and knowledge engineers assess this information while deciding which ontology to select for reuse. Considering this information in the evaluation, ranking and recommendation process is therefore recommended.

Table 6-1 Dimensions, Metrics and Measures for Ontology Evaluation and Selection

Dimension	Metric(s)	Measure	Value
Internal	Internal aspects of ontologies e.g. content, structure, consistency, and correctness.	Out of the scope for this research	NA
Metadata	Usage Information	Is there any information available about the other people/organisation that have reused the ontology? or the other projects that have reused the ontology? or the purpose that an ontology has been reused for?	yes ¦ no
	Documentation, Metadata, Publication and Standardization	Availability of internal comments and labels	yes ¦ no
		Availability of external documentation	yes ¦ no
		Availability of Metadata	yes ¦ no
		Developed using a methodology for ontology development	yes ¦ no
		Is there any publication available?	yes ¦ no
		Has reused other ontologies?	yes ¦ no
	Language	OWL, RDFS	
	Size	Number of Classes	
Social	Popularity	Popularity of an ontology in the community and among colleagues The reviews and ratings of an ontology	
	Maintenance and Responsiveness	Frequency of updates	Number
		Availability of contact information	yes ¦ no
	Reputation and Trust		

- **Documentation, Metadata and Standardization (Factor 2).** Documentation and publication, metadata (e.g., version), standardization, and reusing existing ontologies were among some of the features that get assessed while selecting an ontology for reuse. According to the findings of the first two phases of this research, ontologists and knowledge engineers would like to have access to as much additional information as possible while selecting an ontology for reuse. Documentation is one of the main sources of additional information and can be used to learn about different aspects of an ontology, from how it was built to how it should be used. Some ontology selection tools like OLS and OBO Foundry provide different metadata and additional information about the ontologies in their repository. However, they do not use it as a metric in the ranking process (Côté *et al.*, 2006).

Availability of publications was found to be very important in the ontology search and evaluation process. However, none of the current selection tools for ontologies provide any information regarding the availability of publications about respective ontologies or consider it as a selection criterion. The findings of this study also indicated that ontologists and knowledge engineers would like to know if an ontology is based on any of the recent ontology development standards like OBO Foundry and W3C, before selecting them for reuse.

Reusing and importing existing ontologies can also be helpful in the selection process. As the findings of this study suggested, using some of the ontologies is inevitable, and not using them can show the lack of quality. Currently, it is not very easy to evaluate the quality of an ontology based on the ontologies that it is reusing. However, it can be argued that reusing ontologies with a higher quality might affect the overall quality of the developed ontology.

- **Other Metadata (Factor 3).** Scope, accessibility, size, and language are among the other aspects of ontologies that can be assessed while evaluating and selecting them for reuse. Identifying the scope of an ontology or how well it covers a domain is a complicated task; however, ontology developers can make their ontologies more reusable by providing documentation that includes information about the goal and scope of an ontology and what it aims to cover, or even what it is not covering. Accessibility was the other discussed feature; to be reusable, ontologies have to be online and accessible.

  The results of this study found the size of an ontology to be an effective factor in the ontology evaluation and selection process. Size is a complicated metric, and as it is seen in the literature, there is no consensus if smaller ontologies are better or the larger ones. Moreover, OWL was found to be the most popular languages for ontology development. Therefore, it can be suggested that ontologies developed in OWL will have a higher chance of being reused.

6.2.2 Social and Community Related Aspects

The findings of this study identified a set of social-related metrics that can be used in the evaluation, selection and recommendation process.

- **Popularity (Factor 4).** According to the findings of this study, the popularity of an ontology does not only depend on what has previously been known and used in the literature, but it also depends on the popularity of an ontology in the community and amongst colleagues and the reviews and ratings of an ontology. As was discussed in section 2.5.3, the role and essentiality of ratings and reviews for ontologies has been one of the discussed topics in the field of ontology engineering, especially between 2005-2008. However, most of the current selection systems for ontologies do not provide their users with any facility of adding ratings and reviews for ontologies. Therefore, the results of this research have important implications for developing search and selection systems for ontologies.

- **Maintenance and Responsiveness (Factor 5).** The findings of this research indicated that ontology development is usually an ongoing process, and ontologies need to be maintained and updated over time. It was also argued that the team or organisation responsible for ontology development should be responsive and flexible to the potential changes. Therefore, maintenance and responsiveness are used by the framework proposed in this research.

- **Reputation and Trust (Factor 6).** The results of this study indicate that knowing and trusting the team and/or organisation responsible for ontology development and their reputation in the particular domain, can affect how the quality of an ontology is evaluated. None of the current selection systems for ontologies provide information that can be used to measure reputation and trustworthiness (reliability) of ontology developers in a community. However, one possible way of measuring this feature might be via assessing the popularity of the organisation, institute or research group responsible for ontology development. Availability of reviews and ratings about ontologies could also help in measuring the reputation of an ontology.

6.3 Framework

As it is seen in Figure 6-1, the framework proposed by this study accepts three different inputs, namely, a set of ontologies and the scores assigned to them, a set of selection requirements and also weights and importance that should be assigned to each selection criteria. As the output, this framework recommends a set of ontologies that best cover the selection requirements.

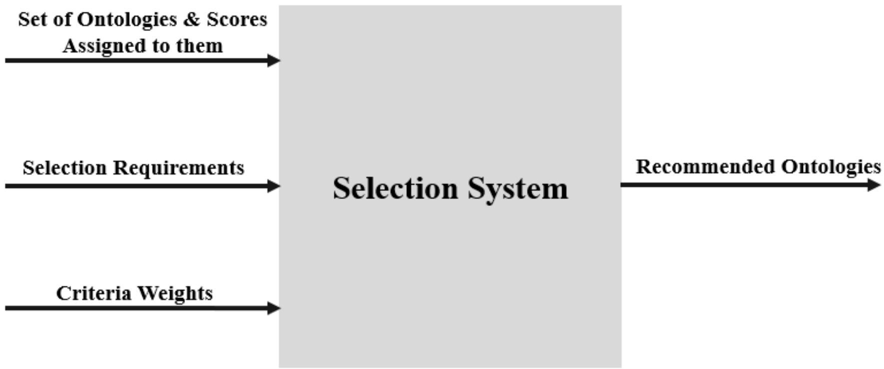

Figure 6-1 Framework Overview

As it is seen in the detailed view of the framework, presented in Figure 6-2 , the ontology repository used in this framework is formed by collecting ontologies from either external repositories, web crawling, or allowing users to submit ontology URLs. Most of the metrics used in this framework are query independent. Therefore, ontologies can be pre-processed and evaluated using the evaluation factors identified in the previous phases, before the process of ontology selection begins. A set of ontologies, as well as the scores assigned to their different features, will then be kept in the repository of this framework.

The second part of this framework is responsible for ontology identification and recommendation. It consists of the following components:

1. **Query Pre-processing.** This component is responsible for performing different pre-processing techniques on the selection requirements submitted by users, namely, expanding the query (Cantador, Fernandez and Castells, 2007) and removing the stop words (Butt, Haller and Xie, 2016).

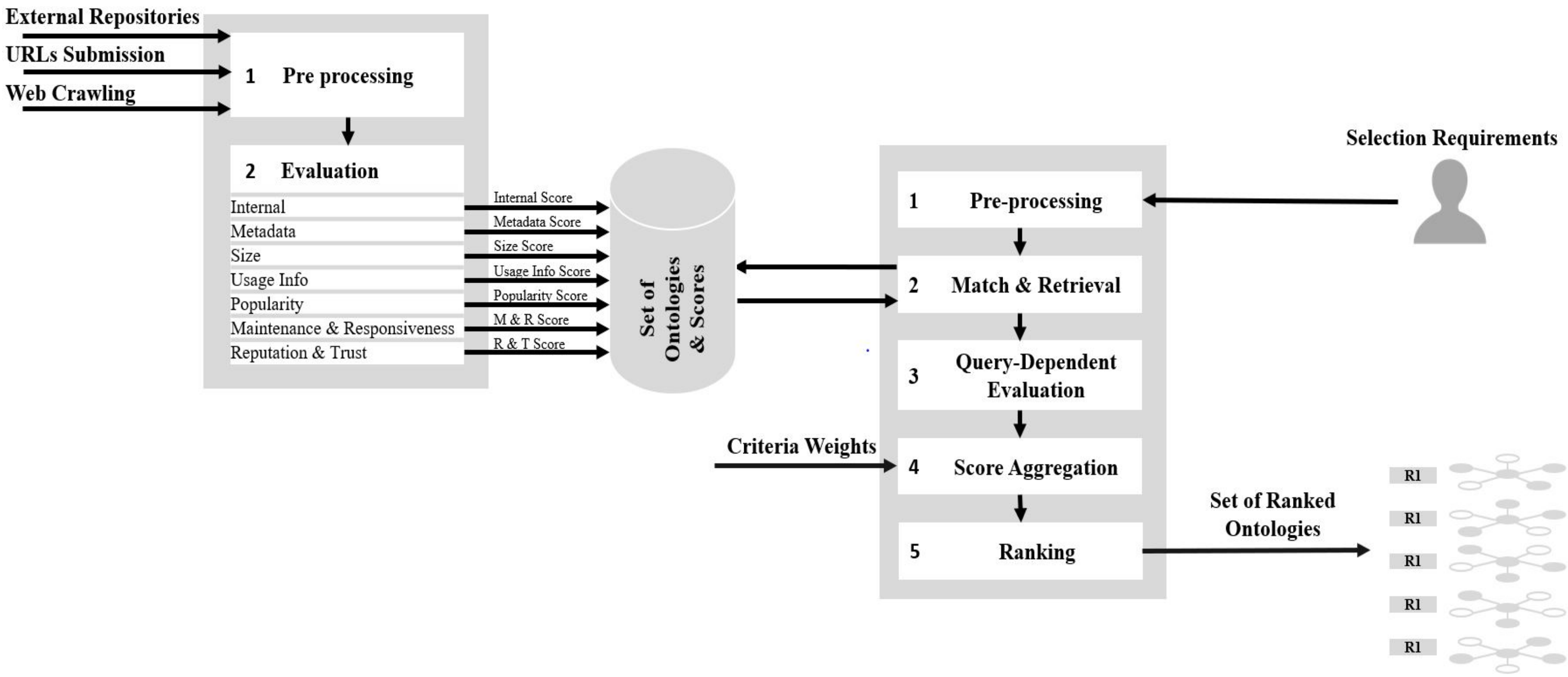

Figure 6-2 Detailed View of the Framework

2. **Match and Retrieval.** Having the ontology repository and the selection requirements, this component aims to identify a set of ontologies that best matches the users' requirements. Six different scores have already been assigned to each of the ontologies that are retrieved in this step.
3. **Query-Dependent Evaluation.** The main focus of this research was on query-independent criteria; however, the findings of this study suggested that some of the query-dependent evaluation criteria, such as, how well an ontology covers the selection requirements is also important when deciding which ontology to reuse. Therefore, the third component of the second part of the framework evaluates ontologies using different query-dependent criteria and assigns a score to them.
4. **Score Aggregation.** This component aims to aggregate the score assigned to each of the ontologies retrieved in the previous steps. The importance of different evaluation factors might vary based on the project requirements. Therefore, the fourth component of this framework allow users to assign different weights to the evaluation criteria.
5. **Ranking.** This component is responsible for finalizing the ranking and ordering of ontologies, based on the scores and weights assigned to their different features.

6.4 Approaches, Tools and Techniques for Validating the Framework

Before validating the findings, the validation processes and techniques used in similar studies are discussed to help identify and justify the most appropriate method to be used in this phase of the research. Maiga & Williams (2009) have proposed two primary approaches for validating the performance of different ontology evaluation and selection tools/techniques: human-based validation and tool validation. However, as it is seen in the literature, selection algorithms and frameworks can also be evaluated by being applied to a collection of ontologies.

In the first approach, usually a group of ontologists are asked to evaluate a set of ontologies; their evaluation, ranking and ratings are then compared with the ratings and rankings of the tool or technique under study. To validate AKTiveRank, for example, a set of ontologies were first retrieved using Swoogle and were then evaluated and re-ranked using the measures proposed by this technique. A user-based experiment was then conducted and compared the rankings ontologists assigned to different ontologies to the rankings that were obtained by AKTiveRank (Alani, Brewster and Shadbolt, 2006).

A similar approach was used by Ungrangsi, Anutariya and Wuwongse (2007), where a group of ontology users were presented with a collection of 63 ontologies and some queries and were then asked to identify and rank the ontologies that best covered the queries. The users' rankings were then compared to the rankings of the combiSQORE algorithm using Spearman's Rank Correlation Coefficient (Ungrangsi, Anutariya and Wuwongse, 2008). Alani *et al.* (2007) also presented five experts in the biomedical domain with a repository of 55 ontologies and four different queries and asked them to identify the ontologies they thought were relevant to the queries. The assessments and rankings of users and the algorithm were then compared with each other.

User-based validation can also include asking ontologists about the usefulness of a set of metrics or the performance of a system. Lozano-Tello (2002), for example, validated the metrics proposed in ONTOMETRIC by sending questionnaires to 10 experts and asking them to express how important they thought each of the 160 aspects of ontologies they had identified were; the options included: it is not important, it is not fundamental, it is important, it is very important, and it is fundamental. At the end, participants were also provided with space which would allow them to add comments on those criteria (Lozano-Tello and Gómez-Pérez, 2004).

Moreover, to validate BiOSS, which is a more recent example, Martínez-Romero *et al.* (2014) contacted several experts via a questionnaire and asked them to evaluate the system's performance considering five different test cases. The main aim of that questionnaire was to find out how helpful and useful BiOSS was. As a result, evaluators provided the system developers with some of the main strength and weaknesses of BiOSS.

The second validation method proposed by Maiga and Williams (2009) included using different tools to evaluate the quality of a set of ontologies and then comparing the assessment results with each other, to identify the tool or technique that is more efficient or successful in finding a set of ontologies that best meets the predefined requirements (Ungrangsi, Anutariya and Wuwongse, 2008). The performance of algorithms proposed by Wang, Guo and Fang (2008) and Butt, Haller, & Xie (2016), for example, were compared with other similar techniques, e.g., AKTiveRank (Alani, Brewster and Shadbolt, 2006). Some argue that this approach is only applicable when the ontology/ontologies under study are mature and well-known enough, and some previous evaluation data is available for them (Maiga and Williams, 2009).

As an example of the third validation approach, Tartir *et al.* (2005) applied their metrics to three ontologies, namely, SWETO[38], TAP, and GlycO[39] and compared how good each one of those ontologies was; the main aim of this comparison was to demonstrate the applicability of those metrics. Burton-Jones *et al.* (2005) also validated the feasibility of their proposed metrics by applying them to different ontologies in DAML library[40].

6.5 Validating the Findings of this Research

Two different experiments were conducted to validate the findings of this research and test the usefulness of the identified metrics in the evaluation process. The first experiment, which is presented in 6.5.1, included applying the identified metrics to some of the most and least used ontologies in BioPortal. The second experiment included an interview study and is presented in section 6.5.2. There are two reasons why BioPortal was used in both of the experiments, namely, because it is a very popular and well-known repository in the ontology domain and also because some of the additional information needed in these experiments was provided by this system; for example, the acceptance score provided by BioPortal can partially be linked to one of the factors identified and used in this study called popularity (Martínez-Romero *et al.*, 2017).

Pilot studies had been conducted in the previous phases of this research. The experiments conducted in the third phase, however, were very similar to some of the previous validation experiments in the ontology evaluation domain, such as the ones conducted by Tartir *et al.* (2005) and Burton-Jones *et al.* (2005). Therefore, the researcher was sufficiently informed about how best to design and implement the experiments and pilot study was not deemed necessary in this phase.

6.5.1 Experiment One

This experiment was designed to assess the usefulness of the quality metrics identified by this research in the process of evaluating and selecting ontologies for reuse. It also helped in

[38] http://knoesis.org/sweto
[39] http://research.bmh.manchester.ac.uk/cancerarchive/glycooncologyangiogenesis/glycooncology/
[40] http://www.daml.org/ontologies/

determining the effectiveness of those factors in predicting the reusability of ontologies. The following steps were taken in this experiment: first, a group of 20 ontologies were selected from BioPortal, based on the number of times they had been reused. They were then divided into two groups: (1) those that had never been reused, and (2) those which had previously been reused (11 times or more). Finally, they were evaluated using the metrics identified in this study, with the findings presented below.

- **Usage Information (Factor 1).** In BioPortal, there is no information on why an ontology has previously been reused. Therefore, it was very difficult to quantify this metric and to determine if there is a relationship between the usage related information provided for each ontology and the number of times it has been reused.

- **Documentation, Metadata and Standardization (Factor 2).** It was very interesting to see that a link to external documentation was provided for nine (out of 10) ontologies that had previously been reused. However, there was no external documentation or additional information provided for almost all of the ontologies that had not been reused. There was also a very clear link between being a part of a standard, e.g., OBO Foundry, with the number of times an ontology had been reused. Most of the ontologies (8 out of 10) in the first group were not a part of any recognised standard. In contrast, ontologies that were reused more often had mostly followed the principles proposed by OBO Foundry. Moreover, most of the ontologies that had been reused more often had web pages, which had provided different additional information and browsing facilities.

- **Other Metadata (Factor 3).** The comparison between the 10 most reused ontologies on BioPortal with the 10 least reused ones in this repository did not confirm the link between reusability of an ontology and its development language; only 4 (out of 10) ontologies that were reused more often were developed in OWL, whereas, most of the ontologies in the first group were built using OWL. Usage information in the BioPortal website did not indicate any meaningful relationship between the size of an ontology and the number of times it has been reused. For instance, SNOMED CT[41] and Ontology for Biomedical Investigations (OBI)[42] as two of the most reused ontology in this repository had 347,358 and 3,380 classes, respectively. In contrast, Surgical Secondary

[41] https://digital.nhs.uk/services/terminology-and-classifications/snomed-ct
[42] http://obi-ontology.org/

Events[43], as an example of a small ontology, had not been previously reused by BioPortal users.

- **Popularity (Factor 4).** The results of comparing the acceptance score of the selected sample suggest that 60% of ontologies that were frequently reused had an acceptance score of 80 or above (out of 100). In contrast, the maximum acceptance score for the ontologies in the first group was 26.1. Due to the limitations of the current selection systems, this experiment was unable to measure other aspects of popularity identified by this research, namely, how popular the developer team or organisation responsible for an ontology were.

- **Maintenance and Responsiveness (Factor 5).** In this experiment, ontologies that had been reused more often were the ones that were frequently updated (e.g. monthly). Moreover, some of the ontologies in this group had mailing lists and/or GitHub pages that users could refer to, in case they had any queries about the ontology or needed any alteration to it. In contrast, it was observed that ontologies in the first group were updated less frequently; some had not been updated since 2013. Contact information, however, was provided for all of the 20 ontologies in this experiment. Overall, it can be concluded that responsiveness of the developer team or organisation responsible for ontologies is likely to affect how they are being reused.

- **Community, Reputation and Trust (Factor 6).** Calculating and quantifying the level of reputation and trust amongst the community of ontologists and knowledge engineers is hardly possible, as none of the current selection systems for ontologies collect and provide relevant information for these factors. However, by combining popularity related metrics with different information about the organisation that some of the ontologies had been built at, it is likely that a connection exists between the reputation of the team and/or organisation responsible for ontology development and the number of times an ontology gets reused. SNOMED CT, as one of the very well-known and reused ontologies, is developed by NHS and College of American Pathologists (CAP).

Overall, it can be concluded that most of the factors identified in this study can be used to judge the reusability of an ontology. They can also be used as a guideline for developing reusable ontologies. In other words, ontologists and knowledge engineers can make the ontologies they

[43] https://bioportal.bioontology.org/ontologies/SSE

are building more reusable by following and providing different sets of additional information, as suggested by this research. These include documentation, metadata, and also maintenance of ontologies.

Due to the limitations of the current selection systems and lack of data concerning social aspects of ontologies, e.g., community ratings and reviews, measuring some of the proposed metrics, such as popularity and reputation and trust, was not easy. Notwithstanding these limitations, these findings have important implications for the ontology selection systems and the kinds of facilities they should provide for their users.

6.5.2 Experiment Two

The second experiment in this phase included interviews with eight experts in the ontology domain. In the information science domain, users can get involved while building or evaluating scientific knowledge (Mandran and Dupuy-Chessa, 2018). The user centred experiment in this phase aimed to determine how well the findings of this study can predict the ontology that knowledge engineers would likely select for reuse. This experiment also investigated whether ontologists and knowledge engineers find having information about the factors proposed in this study helpful in the selection process.

The NCBO BioPortal Recommender was employed as the baseline for this experiment. The inclusion criteria for this phase, therefore, was being expert and actively involved in the process of ontology evaluation and selection in the biomedical domain. The researcher identified some of the ontologies that were recently updated in the NCBO BioPortal and contacted the people responsible for developing and maintaining them. Some participants were also chosen from the experts who the researcher had come in contact with while doing this research but had not had a chance to interview them in the previous phases. The eight participants in this phase were asked to conduct an ontology search using NCBO BioPortal Recommender and to send their search results to the researcher.

The top 10 recommended ontologies for each user's query were then evaluated and (re)ranked using the query-independent factors identified in this study. An explanation, e.g. "ontology A is very similar to ontology B, but it is ranked higher because it is smaller" or "because it has been developed in OWL", was also provided to clarify the ranking assigned to each ontology. Finally, interviewees were presented with the new ranking and were asked *"How useful do you* *think the new ranking is compared to the recommended ontologies by BioPortal"* and *"Do you find the explanations and the extra information helpful in the selection process?"*. They were also asked to select an ontology that best covers their query; this question aimed to determine how successful the identified factors were in identifying an ontology that human experts find the most suitable, and to compare the ranking of this study with the one recommended by BioPortal (see Appendix K). A detailed account of the findings of the experiment is presented in the following section.

6.5.2.1 Ranking comparison

This study was very successful in predicting the ontology that knowledge experts would select for reuse as it ranked those ontologies higher, compared with how NCBO BioPortal Recommender had rank them. One of the participants in this phase used NCBO BioPortal Recommender to find ontologies that best covered a text about quitting smoking. National Cancer Institute Thesaurus (NCIT)[44], MESH[45], LOINC[46] and SNOMEDCT were the first four ontologies recommended by BioPortal; however, when ranked by the factors identified in this study, SNOMED CT had the highest score and therefore, was the first recommended ontology, followed by NCIT, MESH and RXNORM. When asked about the new ranking, respondent E stated:

"I guess, I agree that SNOMED CT is probably the most useful terminology for the kinds of concepts that are discussed in the text that I used because it is the broadest of all the clinical vocabularies out there really in terms of coverage".

In respondent C's case, the ontology they would select, and reuse was ranked 10th by NCBO BioPortal Recommender and second by the ranking based on the findings of this study. When asked about the new ranking and how it compared to what they would select, respondent C argued that they are not sure why the ontology they would reuse for this query is ranked very low in BioPortal. Moreover, and when asked about the most suitable ontology for their query, participant F argued that Gene Ontology (GO) would be the second appropriate ontology for their query. This ontology had the highest score and therefore, was ranked first in the ranking proposed by this research. In the recommendation by BioPortal, however, GO was ranked 10th.

[44] https://bioportal.bioontology.org/ontologies/NCIT
[45] https://meshb.nlm.nih.gov/search
[46] https://loinc.org/document-ontology/

For each of the ontologies in the ranking output, an explanation was provided stating different information about them and why a specific rank is assigned to them. Participants in this experiment found this information very useful. Respondent B, for example, stated: *"What is very useful is to have the comments provided, that can, in a glimpse, provide more confidence for a user to make an informed decision while picking an ontology of his choice"*.

Interviewee A emphasised the importance of transparency and argued that *"information is king"*. They stated that users of the selection systems should not only see the ranking of the ontologies, but they should also know why an ontology has got a specific rank. Respondent F also found it *"super useful"* to have all the additional information for each of the ontologies. Interviewee H added: *"I think it is interesting to try and take this extra information about [ontologies and] try to assess some sort of level of quality of these ontologies to change the ranking"*.

One of the most striking observations to emerge from the ranking comparison was that NCBO BioPortal had recommended NCIT as the first ontology for 7 (out of 8) queries used in this experiment. However, some of the interviewees found this ontology hardly relevant to their query. Respondent F, for example, argued that NCIT is *"like a dictionary of almost everything"* and while it might match many terms in each query, it is more like a random ontology to be recommended for their query. Interviewee D also stated that their query was more about natural language processing and argued that NCIT is not the best ontology for their query. The overall quality of NCIT was also criticized by interviewee G.

The very high rank of NCIT can be explained by the number of classes it has and the very high weight of the coverage metric (0.55 out of 1) in the BioPortal Recommender algorithm (Martínez-Romero *et al.*, 2017). NCIT ontology has 156,172 classes and there is a very good chance that it covers many terms in the biomedical domain, while not being a good or the best match for some of the queries in this domain. This finding has important implication for developing the next generation of the selection systems for ontologies. It suggests that the algorithms that are mainly based on coverage and other internal aspects of ontologies might not always be helpful in the selection process.

Overall, many of the participants found the rankings based on the factors of this study more useful and closer to how they would rank the ontologies. Respondent E, for example, stated that the new ranking is similar to *"the way that I would evaluate the suitability of these vocabularies for classifying this test"*.

6.5.2.2 Overall Usefulness of Factors

Besides re-ranking the top 10 ontologies for each query, participants of this experiment were asked about the usefulness of the metrics employed in the ranking process and the information that was provided for each ontology. Table 6-2 provides an overview of what participants had to say about different metrics; the rest of this section discusses their views in more detail.

- **Usage information.** When asked about the usage related information, many of the respondents mentioned that they would like to have access to that information. Respondent F put it this way: *"That [knowing why an ontology was reused before] is very useful; if it is mostly for indexing, or annotation or text mining. That matters because some ontologies are better for something, some for something else".* When asked about the importance of usage related information, Interviewee H had the following to say:

  "Yes, we care about that; typically for us, we are looking to use the ontologies that describe our data here. If we want to use someone else's ontology, then our assumption is that ontology is being used somewhere else to describe similar data, so that means that we can integrate with our data".

  Moreover, and similar to the previous phases of data collection, participants in this experiment stated that they would care more about why an ontology has been reused and who have reused it, rather than the number of times an ontology has been reused.

- **Documentation.** Most of the participants in this experiment emphasized the importance of documentation in the selection and reuse process. Respondent G, for instance, used OBO Foundry as an example and mentioned that having documentation is one of the requirements of getting accepted in that repository. Interviewee E also added: *"I think that [documentation] is extremely important; if a published ontology didn't have fair documentation, it would be pretty useless to me, because I would not feel like I was able to use it in the intended way".*

Despite its importance, some participant, namely, Interviewee F argued that finding documentation about an ontology might be challenging, as documentation for ontologies are published in different places, e.g., GitHub or various separate websites.

Table 6-1 Summary of Respondents' Comments

Evaluation Dimension	Evaluation Metric	Positive Comments	Negative Comments
Internal	Internal aspects of ontologies e.g. content, structure, consistency, and correctness.	Out of the scope of this phase of the research.	
Metadata	Usage Information	Usage information is very important	That might give me some idea about in what domain it is useful to use it but again it doesn't say anything about the quality.
	Documentation	If a published ontology didn't have fair documentation it would be pretty useless to me	Some would have documentation on GitHub, some would have it somewhere else and it [finding them] is a bit hard.
	Following Standard Approaches and Principles	This is important for me.	Nowadays many of the ontologies do not care if they are OBO Foundry or not; so, maybe it is not so important any more I would say.
	Frequency of Updates	The frequency of update is very important	The amount of time that is updated, is rather a lack of quality than a criterion of a quality
	Language	Yes, that is very important	For me it is not important, but I think for many research projects it will be important.
	Size	The size matters	[Ontology size] doesn't matter
Social	Popularity	It is important but not the most important	If the ontology exactly matches of what I need then maybe I do not care how popular it is.
	Maintenance and Responsiveness	I think it is very important	NA
	Reputation and Trust	I suppose I will be more likely to trust something that was developed by a group that I heard of compared to one that I have never heard of.	That is irrelevant I would say.

- **Standardization.** The importance of being based on or endorsed by ontology standards was also discussed by some of the participants. As respondent E put it: *"just the domain of an ontology is not enough to know whether it is usable for the project I am working on; I also need to know the approach they have taken with it"*. Interviewee H found being a part of a standard, like OBO Foundry, very helpful and stated that they *"prefer* *that ontologies are part of OBO Foundry, it makes things easier"*. However, they argued that *"there are also some very good ontologies being built in the life sciences that are not a part of OBO Foundry"*. Principles proposed by OBO Foundry were also blamed by participant G for ignoring the quality of an ontology.

- **Metadata.** When asked about the importance of having additional information about different aspects of ontologies, such as their version and license, participants had interesting thoughts. Interviewee H, for example, stated that SNOMED CT is a very good ontology but blamed BioPortal for not providing information about its license restrictions. They started:

  "I would not use SNOMED CT ever, because we are in academic institute working with public open data and SNOMED CT has a license. So, the license restriction on SNOMED CT is like if you have public open-source data, then you cannot use SNOMED CT, and that is not reflected here [in BioPortal]. So, while SNOMED CT is a very good clinical terminology, it comes with a license and a cost".

- **Size.** Size of the ontology or the number of classes it has was one of the other metrics discussed in the interviews. Interviewee A argued that very large or very small ontologies are not useful. When asked if smaller ontologies are better or the bigger ones, participant D argued that the acceptable size for an ontology depends on the purpose of reusing it; they explained that smaller ontologies might be more useful for indexing. Interviewee H also argued that there is a link between the size of an ontology and the area it is covering and stated: *"Obviously if I need an ontology of disease and there are only 50 classes, then I will probably be a bit suspicious that is going to give me the coverage that I need".*

- **Language.** According to the survey conducted in the second phase of this research, OWL was chosen as the most used language for ontology development. Interviewee G, however, blamed OWL for not being expressive enough for modelling some of the topics in their domain, e.g., time indexing. Participant H agreed with the expressivity limitation of OWL but argued that they are "*happy to give up a bit of expressivity in order to get access to a tool that works with OWL"*. They justified using OWL by stating:

"I think you have also got to kind of weight expressivity with sort of scalability and how much you can actually compute with these languages. So yes, I could take some other first-order language, but I might not be able to find anything that can make the scale for the ontology sizes that we work with".

They also added that they use OWL, not because of the expressivity, but because it is a well-known standard and there are tools (e.g., Protégé) and reasoners that can be used with it.

- **Popularity.** Many of the survey respondents found how popularity is currently measured unhelpful. Interviewee E, for example, stated that *"If an ontology is designed for a very specific purpose then it might not be broadly used but that does not necessary mean that it is not a good piece of work and that it wouldn't be useful for something that I am doing".* Respondent G also argued that the number of times an ontology has been reused is not a good metric for two reasons, namely, because many of the ontologies have their own websites and some may prefer to get the ontology directly from the ontology website, instead of using the repositories.

  Respondent A also mentioned that just because some people use an ontology in a specific repository like BioPortal may not mean that it is useful. Interviewee H also agreed and found the acceptance metric used in BioPortal *"dangerous"* and *"sort of misleading"*. Interviewee D supported the popularity factor proposed by this study and argued that popularity of an ontology depends on who is using the ontology; they added that they tend to use the ontologies that are used by their own community and are popular in their country.

- **Maintenance.** Most of the participants in this experiment would consider the availability of contact information and ontology maintenance while selecting an ontology for reuse. Interviewee E put it this way:

  "Both of those are super important. In terms of contact information, I basically never reused a model directly from BioPortal without, you know, trying to see if it exists, if I can get it directly from the source. Like if there is some organisation that maintain the ontology on their website, I'd rather get it there because I know that I am getting the most recent version and I know that it is going to have the contact information that's

going to lead me to the right people. So, I would say these are things that are important".

- **Frequency of Updates.** While most of the interviewees emphasized the importance of maintenance, there was no general agreement regarding the number of times an ontology should be updated. Some participants, like interviewee A, argued that ontologies should be updated more than once a year and stated that actively developed ontologies are updated every month. Respondent H agreed and stated that they expect to see updates every month in a healthy project and would consider it a good sign, as it shows that someone is actively working on the ontology. Respondent E, however, stated that the number of times an ontology has been updated is not important, as long as it has been updated once. They also argued that if an ontology has never been updated before, it will be a *"big red flag"* for them as *"nobody gets everything perfect on the first try"*.

  For participant G, frequent updates would indicate *"rather a lack of quality than a criterion of a quality"*; they argued that a perfect ontology does not need to be updated. They also stated that an ontology *"should describe what is generic in reality"* and those generic types usually never change. Respondent H, however, disagreed and stated: *"our understanding of science changes every month, so our knowledge does change quite regularly. I think it might be other domains and areas where things can be fairly stable and understood, but if you look at something like the Gene Ontology, our knowledge about what proteins do changes every month or not necessarily changes, but we are adding to it"*.

  Interviewee H argued that the context matters and that frequent changes could *"potentially [be] a bad sign because it might mean that this ontology is changing and therefore it is unstable and therefore, I should not be using"*. They added that users *"have to understand the nature of what is being changed. Are they just altering labels all the time or are they actually adding new classes? are they deleting classes? are they moving things around the hierarchy?"*. They also suggested that good changes in an ontology would include addition, rather than deletion and hierarchy rearrangements.

- **Community, Reputation and Trust.** Interviewees were asked if they thought that the interactions in the community and the reputation of the ontology developer team and organisation can affect the selection process. Interviewee A stated that they prefer to use an ontology that has been built by a professional team or organisation rather than someone whom they do not know. Respondent H supported the idea by stating: *"there is certainly a reputation around who has built a particular ontology and that might influence. If I had to decide between two, and I knew some of the developers of one, that can mean two things: one that I trust them and I think that they build good ontologies, and the other is that I know them and I know that they will be willing to collaborate and we could work on extending it together; so it is good"*.

6.5.2.3 General Comments

Participants in this phase had interesting thoughts on the usefulness of the available selection systems for ontologies. Some, for example, found certain ontologies recommended by BioPortal inappropriate and stated: "*I wouldn't consider using those ontologies for classifying a text like this"*. For interviewee G, moreover, the ontology they would select for their query was not in the list of ontologies recommended by the NCBO BioPortal Recommender.

Some of the interviewees also compared different selection systems for ontologies. Respondent F, for instance, mentioned that they use both OLS and BioPortal and stated: *"I think they complement each other in some way and none of them is great, none of them is perfect, but each one of them has its own advantages"*. Interviewee H also compared these two repositories and argued that the main difference between them is that OLS does not allow everyone to upload their ontologies there and ontologies are checked and chosen to be added to that repository; therefore, it *"gives you a certain level of quality assurance that you do not necessarily get in BioPortal"*.

Moreover, participants in the interviews were asked about the advantages and disadvantages of open selection systems and closed selection systems. Interviewee H stated that both open and closed systems are needed; however, selection systems should assure people about the quality of the ontologies in their repository. They said that they do not want what they called *"toy ontologies"* or the ontologies that *"someone has created for fun"* in their archive and argued that other people might end up using and annotating their data with an obsolete ontology that will never get updated again.

In answer to the question about the weights of different criteria in the evaluation process, all of the survey respondents firmly stated that different weights should be assigned to different metrics. They also argued that users should be able to personalize the assigned weights to each metric, based on their personal preferences or the purpose of the ontology selection. Some of the respondents, however, talked about the challenges of weight assignment and argued that they do not know what the best way of assigning and figuring out the weights for each metric is. Interviewee C, for example, argued that identifying the weights can be a separate research project in itself. Participant E also used BioPortal as an example and argued that the selection systems might have to come up with some "universal decisions about what is the most important".

6.6 Summary

This chapter aimed to validate the findings of the first two phases of this research. Two experiments, including a query-independent and a query-dependent one, were conducted. The first experiment aimed to determine how useful the results of this study were in predicting the reusability of ontologies. The second experiment presented the findings of the previous phases to a group of ontology experts and asked them how useful they thought the factors proposed by this study were in the selection process.

Overall the results of this chapter confirmed the findings of the previous phases of this study. Ontologists and knowledge engineers consider the factors proposed by this study important in the selection process and would like to have access to them while evaluating and selecting ontologies for reuse. Moreover, the rankings based on the findings of this study were very successful in predicting the ontology that the knowledge experts found relevant to their query and would reuse. These findings have important implication for developing ontology selection systems.

Chapter 7: Discussion

The notion of ontology quality, and the process of evaluating it, have been considered as two of the most significant and also complicated challenges in the ontology domain since 1995. Despite the importance of this matter and the extensive research on this topic, there are still many unanswered questions and challenges when it comes to evaluate and select ontologies for reuse. The main aim of this study was to address these issues by exploring the perspective of those who are working in the ontology domain. To do that, different phases of data collection, including a survey and two sets of interviews were conducted.

The findings of this study provided a new understanding of the notions of quality and reusability in the ontology domain. These results not only confirmed the importance of internal aspects of ontologies in the evaluation and selection process, but they also identified and clarified two other dimensions that can be assessed while selecting ontologies for reuse, these being different metadata and social aspects of ontologies. Moreover, these findings clarified the role of community in the ontology domain and suggested that the interactions in the community can affect how the quality of an ontology is evaluated. The following parts of this chapter discuss the results from the different phases of this study in greater detail.

7.1 Ontology Development and Quality

Since 1995, different approaches have been proposed to facilitate the process of ontology development and to support ontology reuse (Uschold and King, 1995; Suárez-Figueroa, Gómez-Pérez and Fernández-López, 2012). However, very little was known about how popular those methods were and how often they were used in the development process. The overall findings of this study suggested that the early versions of methodologies for ontology development, such as METHONTOLOGY (Fernabdez, Gomez-Perez and Juristo, 1997), are not very popular amongst ontology developers and knowledge engineers.

However, following and being endorsed by some of the recent development practices and principles, like the ones proposed by OBO Foundry and W3C, was found to be very important in the ontology domain, both for those who develop ontologies and those who evaluate and select them for reuse. Despite their importance, these principles, especially those of the OBO Foundry, were blamed for covering very specific domains, e.g., biomedical, and being limited to a small set of ontologies. The evidence provided in this research suggested that the notion of quality and some of the metrics it depends on are very similar across different domains. One implication of this, therefore, is the possibility of applying principles used in the biomedical domain to the other domains.

7.2 Ontology Search

Since the early 2000s, different search engines and selection systems, e.g., Swoogle, have been proposed to facilitate the process of ontology search and selection. Despite a large number of available systems, there has been very little discussion and investigation about the set of characteristics those systems should have, the metrics they should use in the evaluation and ranking process and the functionalities they should provide to be useful (Maiga and Williams, 2009).

The findings of this research indicated that ontology selection is still a manual task because none of the available systems provides users with all the facilities and functionalities they need. Contrary to expectations, comparing different selection systems suggested that ontologists and knowledge engineers mostly tend to refer to the literature and scientific papers to find ontologies; this finding was interesting as the availability of scientific papers has not previously been linked to ontology development or search and selection. Google was also widely mentioned as a popular source of finding ontologies; however, some of the participants blamed it for not "doing more" regarding ontology search and selection.

The findings of this study also showed that repositories for ontologies, like BioPortal, OLS, and LOV are more popular than the early versions of ontology search engines like Swoogle and Watson (d'Aquin and Motta, 2011). A possible explanation for this might be that libraries and repositories provide different sets of additional information and metadata for each of the ontologies in their collection. As the findings of this research suggested, this information is what knowledge engineers look for when selecting an ontology. However, most of these systems are limited to a specific domain and therefore, cannot be used by the knowledge engineers working in other domains.

Findings of this research have important implications for developing ontology selection systems. To be useful, these systems should not only provide sets of additional information about ontologies, but they should also use some of that information in the evaluation and ranking process. They should also facilitate interactions amongst the community of ontologists and knowledge engineers, e.g., by allowing them to add ratings and reviews. Finally, those ratings and reviews should be considered in the evaluation and ranking process.

7.3 Ontology Evaluation

The focus of this research was on the criteria-based evaluation approaches, also known as metric-based, multiple-criteria (Brank, Grobelnik and Mladenic, 2005) or feature-based (Arpinar, Giriloganathan and Aleman-Meza, 2006) approaches. According to this method, the suitability of an ontology for a particular task or requirement is evaluated by being compared against a set of predefined criteria (Maiga, 2008). Finding a set of metrics for ontology evaluation and reuse has always been a key research topic in the ontology domain. Therefore, different sets of quality metrics for ontology evaluation and selection have been proposed in the literature (Martínez-Romero *et al.*, 2014, 2017; Butt, Haller and Xie, 2016).

However, most of the previous works are based on a limited set of similar metrics, and much uncertainty still exists about the importance and usefulness of those metrics in the evaluation process. Therefore, this study set out with the aim of asking ontologists and knowledge engineers about the notion of quality, and also assessing the importance of the previously identified quality metrics. It also investigated if and how the interactions amongst the ontology developers and users can affect the evaluation and selection process. The rest of this section will discuss the metrics identified in this research.

7.3.1 Internal Aspects of Ontologies

As it is seen in the literature, different internal characteristics of ontologies have been used in their evaluation process. Coverage (Buitelaar and Eigner, 2008), content and structure (Buitelaar, Eigner and Declerck, 2004), consistency (Burton-Jones *et al.*, 2005; Raad and Cruz, 2015), completeness (Yu, Thom and Tam, 2009) and correctness (Jonquet, Musen and Shah, 2010) are amongst some of the popular metrics used in the literature. When asked about the quality of an ontology, participants in this study mentioned different internal characteristics of ontologies as the criteria they would consider while evaluating an ontology. These results were consistent with those of the previous studies.

Despite the importance of the internal characteristics, some issues need to be addressed. Firstly, measuring some of the internal characteristics is complicated, as there is no consensus regarding the definition of those metrics or the measurement strategies that should be used for them. For example, the findings of this research suggested that the content of a reusable ontology needs to be explicit, accurate, precise, unambiguous, generic, clearly defined and easy to understand. However, it is very complicated, if not impossible, to measure and quantify the accuracy, preciseness, clarity, or genericness of the content. Moreover, while ontologies are built to describe, capture and conceptualize different elements of different domains (Chandrasekaran, Josephson and Benjamins, 1999), ontology developers tend to describe the world differently; therefore, what is clear and accurate content to some people, may seem unclear and inaccurate to some other people.

7.3.2 Metadata's Role in the Ontology Evaluation

Participants in this study had some interesting views to share about the type of additional information they would like to have access to while evaluating and selecting ontologies for reuse; this section will discuss some of those views.

7.3.2.1 Usage Information

The reason an ontology had been reused before (e.g., annotation), or who and what organisation has reused it, and in what project, were amongst some of the additional information participants of this research would like to have access to while selecting and evaluating ontologies. The importance of metadata has been discussed by the researchers of some of the previous studies in the ontology domain (Hartmann, Palma and Sure, 2005; d'Aquin and Noy, 2012); however, they have not dealt with the usage related information suggested by this research and have not linked them to the ontology evaluation and selection process.

Some selection systems, like BioPortal, provide a list of projects that have reused an ontology, but not the reason an ontology has been reused. These systems, therefore, could facilitate and enhance the selection process by providing the reason an ontology has been reused, and, also, information about the people who have reused it. Recommendation algorithms can also be applied to these additional sets of information to identify similar users or similar ontologies.

7.3.2.2 Documentation, Metadata, and Standardization

The importance of documentation in the ontology development and evaluation process has been the subject of some of the previous studies; however, far too little attention has been paid to what documentation should include. In 1995, Gómez-Pérez proposed a long list of items, such as ontological commitments and summary of ontology definitions that need to be included in the documentation (Gómez-Pérez, 1995). ONTOMETRIC also suggested evaluating three different types of documentation, namely, "Documentation Using Access Interfaces", "Documentation Programming Access Interfaces", and "Tool Supplies Documentation About Built Products" while assessing an ontology (Lozano-Tello and Gómez-Pérez, 2004).

Apart from these two studies, there is a general lack of research on the role of documentation and additional information, especially when ontology selection is concerned. Moreover, no previous study has investigated what ontology users and developers would like to know about ontologies and see in their documentation. Therefore, this is the first study reporting additional evidence with respect to documentation, from the users' point of view. The findings of this study suggest that ontologists and knowledge engineers expect the documentation to include background and modelling assumptions and to explain why an ontology was built (e.g., specific purpose, reuse). Moreover, they would like to have access to different use cases and guidelines that clarify how an ontology should be used.

Participants of this study also argued that finding documentation about an ontology might be challenging for two reasons, namely, because they might not be available at all, or because they are published in different places, e.g., GitHub, ontology website, or Google Scholar[47]. Some of the selection systems, like OLS and OBO Foundry, have tried to address this issue by providing a link to the ontology's homepage or a list of publications about an ontology. It can be improved by aggregating all the additional information about ontologies in one place.

Availability of documentation was found to be the fourth most important metric used in the evaluation process. However, none of the current selection systems for ontologies consider the availability of documentation as a metric in the evaluation and ranking process. Put another way, the availability of documentation for ontologies will not affect how they are ranked by the current selection systems. It can, thus, be suggested that the overall performance of selection systems can be enhanced by using the availability of documentation as a metric in the evaluation and ranking process.

Many of the well-known selection systems for ontologies provide sets of metadata about each ontology in their repository. However, similar to documentation, none of this information is used in the ranking process. The findings of this study confirmed the importance of the

[47] https://scholar.google.co.uk/

availability and accessibility of this type of information in the selection process. Providing published papers or technical reports, and clarifying scope, goal, purpose, context, and application were amongst the other type of additional information that, if provided, would facilitate the process of selecting an ontology for reuse. Provenance information, such as ontology creator, license type, and the version of the ontology were also found to be important by the participants of this study.

7.3.2.3 Size and Language

As it is seen in the literature, there is no consensus about how the size of an ontology might affect its quality. Some have claimed that larger ontologies are more complete (Burton-Jones *et al.*, 2005), while others have argued that the smaller ones are more specialized (Martínez-Romero *et al.*, 2017). Most of the participants in this research stated that reusing smaller ontologies is easier. Moreover, some of the very small ontologies, such as BFO, PROV-O, DC, SKOS, and FOAF were mentioned as examples of the best ontology participants had reused.

In this study, ontology size was also linked to the domain it is built and used in. SNOMED CT and Gene Ontology were amongst some examples of ontologies that are not small but are very popular and have been reused many times. To explain this, some participants argued that "huge ontologies" are reusable if there is enough documentation about them or if they are modular. Overall, it can be assumed that ontology size is an important factor in the evaluation and selection process. However, it might only affect the decision-making process if two ontologies are equally good, and the only difference between them is their size. In this kind of scenario, the findings of the study suggested that the smaller ontologies have a higher chance of being selected and reused.

Ontology development language was also discussed as a factor in the selection process. As it is seen in the literature, some of the evaluation frameworks, e.g., ONTOMETRIC (Lozano-Tello and Gómez-Pérez, 2004) have suggested using the development language as a criterion in the evaluation process. Selection systems like BioPortal also provide their users with information about the language that each of the ontologies in their repositories has been built in. In the second phase of this research, 80% of the participants chose OWL as the language they use for building ontologies. It is, therefore, likely that building an ontology in OWL will increase its chance of being reused.

7.3.3 Social Aspects of Ontology Evaluation

Participants in this study had some very interesting views when asked about the role of community in ontology evaluation and selection for reuse. The rest of this section discusses the social features of ontologies in more detail.

7.3.3.1 Popularity

Popularity is the most defined and used term in the literature to refer to the role of community in the quality assessment process, and some of the prior studies have noted the importance of this metric. Different terms and phrases, like history (Burton-Jones *et al.*, 2005), connectedness (Buitelaar and Eigner, 2008), authority (Burton-Jones *et al.*, 2005), and direct popularity (Fernández *et al.*, 2009) have been used in the literature to refer to the acceptance (Martínez-Romero *et al.*, 2017) of an ontology in a domain. Moreover, popularity is used in the ranking process by some of the selection systems for ontologies, e.g., BioPortal.

As seen in the literature, popularity usually refers to the number of visits or page views of an ontology in a repository during a recent specific period (Martínez-Romero *et al.*, 2017). Popularity can also be measured by applying the PageRank algorithm (Page et al.,1999) to the ontology domain and counting the number of ontologies that import a particular one (Fernández et al., 2009; Wang et al., 2008; Supekar et al., 2004).

When asked about the importance of popularity in the evaluation and selection process, participants in different phases of this research had some interesting thoughts. Some of them, for example, doubted the link between the quality and reusability of an ontology and the number of times it has been visited in any particular repository. Some also stated that they would care more about the projects that an ontology has been or is being used in, compared to the number of times it has been used. It was also argued that the number of times an ontology has been reused depends on different factors, such as its size, the level of specialisation and the domain that it is built for; therefore, it cannot be used as a metric to measure quality and reusability.

Participants in the second phase of the study were asked to rate the importance of six different popularity related metrics, four of which were previously mentioned in the literature. The survey results indicated that ontologists and knowledge engineers tend to care more about the popularity metrics identified in this research, such as popularity of an ontology in the community and among colleagues (ranked 14 out of 31) and the reputation of the ontology developer team, and/or institute in the domain (ranked 21 out of 31) than the popularity related metrics that have been widely identified in the literature and used by the selection systems for ontologies.

Metrics used in the literature, including the number of times an ontology has been reused or cited (Supekar, Patel and Lee, 2004; Wang, Guo and Fang, 2008), the popularity of an ontology on the web (Burton-Jones *et al.*, 2005; Martínez-Romero *et al.*, 2017), the reviews of an ontology (Lewen and Aquin, 2010) and the popularity of an ontology on social media (Martínez-Romero et al., 2014), were found to be less important in the selection process and were ranked 25, 26, 27 and 31 (out of 31) respectively. The third phase of data collection for this study also confirmed the findings of the previous phases. Overall, the findings of this study do not support how popularity is defined and measured in the literature; many of the participants of this research found the current definition not useful, or even misleading and dangerous.

7.3.3.2 Maintenance and Responsiveness

In this study, maintenance and frequency of updates were highlighted as some of the very significant factors in the process of ontology evaluation and selection for reuse. Participants of this research argued that there is no such a thing as a "complete" or "finished" ontology, and the quality of ontologies is generally limited. Thus, ontology users often need to count on the responsiveness of the ontology developer team and organization, as well as their attitude and flexibility toward the requests for changes.

Despite what the findings of this study suggested, far too little attention has been given to ontology maintenance, the responsiveness of ontology developer team, and how it affects the quality and reusability of an ontology in the literature. Recently, an experiment was conducted by Geller, Keloth and Musen (2018) to determine the reasons some of the ontologies in BioPortal are not maintained. When they tried to contact the ontology developers using the email addresses they had provided in that system, 42.2% of the emails did not get any response. This figure is not very encouraging, especially if someone needs to request for changes in an ontology.

Some of the participants in the first round of this research linked ontology maintenance to community related metrics, like the activeness and reputation of the ontology developer team and organization in a domain. Therefore, as a part of the survey in the second phase, participants were asked how important they thought "Having an active responsive (developer) community" was in the evaluation process. It was very interesting to see that this metric was ranked 12th (out of 31), which was much higher than some of the widely used metrics in the evaluation and selection process, namely, popularity or the number of times the ontology has been reused or cited (Burton-Jones *et al.*, 2005; Butt, Haller and Xie, 2016).

Participants in this study were also asked how important they thought having "wikis, forums, mailing lists and support team for the ontology" was in the evaluation process. Mailing lists are used by some of the very well-known ontologies, like Gene Ontology, to respond to general questions and comments (The Gene Ontology Consortium, 2004). This metric was ranked 15th, which can be an indication of the important role it plays in the ontology domain. Hence, it could conceivably be hypothesised that using and providing mailing lists might be an appropriate solution to address the maintenance and responsiveness issues and to support ontology reuse.

The frequency of updates was the other discussed evaluation factor in this study. Some of the selection systems for ontologies, such as BioPortal, show a list of updates for each of the ontologies in their repositories. However, none of them considers it as a metric in the evaluation and ranking process. Moreover, no previous study has investigated if there exists a link between the quality of an ontology and the number of times it has been updated.

The findings of this research suggest that the number of times ontologies are updated is important when it comes to ontology evaluation and selection for reuse. However, there was no general consensus amongst the participants about the number of times an ontology should be updated. Some argued that a good and healthy project should be updated every month, while others argued that once every couple of years should be enough. Some participants explained this disagreement by arguing that ontologies built and used in some domains need to be updated more frequently, than others.

Hypothesis testing conducted in section 5.4.5 revealed that survey respondents who were building ontologies in the biomedical domain would rate the "Maintenance and Responsiveness" factor statistically significantly higher than some of the other participants, especially those who were building ontologies in various different domains. However, this analysis indicated that participants with different years of experience or those who work for different types of organisations (e.g., academic and non-academic) do not rank the importance of “Maintenance and Responsiveness” factors statistically significantly different.

In this research, the frequency of updates and maintenance was also linked to the availability of funds for ontology development. A recent experiment conducted by Geller, Keloth and Musen (2018) confirmed this finding and suggested the lack of funding or interruption in funding as the main reason why ontologies were not being maintained and updated in BioPortal. However, the availability of funds was ranked 30th (out of 31), when its importance in the evaluation process was tested. One possible explanation for this might be what one of the participants stated about this metric: “it [an ontology] does not necessarily have to be funded if there is a reasonable group of people with some motivation to carry on working on it”.

The combination of findings discussed here confirmed the importance of updates and maintenance as a metric for evaluating the quality and reusability of ontologies. However, this research is unable to suggest the number of times that ontologies should be updated in order to be reusable or selectable. In general, therefore, it seems that selection systems for ontologies should share the update history of ontologies with their users. However, they should not use it as a metric in the evaluation process and should leave it to the users to decide if an ontology has been updated enough or not.

7.3.3.3 Community, Reputation and Social Interactions

One of the very fundamental questions of this study was whether the social interactions amongst the community of ontologists and knowledge engineers affect how ontologies are evaluated and selected for reuse. Communities play a crucial role in the fields that are similar to ontology engineering, like software engineering. GitHub is one of the most successful examples of social-based software development environments and has facilitated collaboration among software engineers (Dabbish et al., 2012). In the ontology domain, however, there are fewer examples of community and social collaboration, especially when it comes to evaluating the quality of ontologies.

Some studies have tried to investigate the role of community in the evaluation and selection process. Hlomani and Stacey (2014), for example, defined user-based ontology evaluation as the process of evaluating an ontology through users’ experiences and by capturing different subjective information about ontologies. Lewen and D’aquin (2010) also argued that relying on the experiences of other users and community for evaluating ontologies is beneficial, as it lessens the efforts needed to assess an ontology and reduces the problems that users face while selecting an ontology.

When asked about the role of community, participants in this study had some very interesting thoughts to share. In the first phase, for example, science was called a social enterprise, and it was argued that knowing the ontology developer team and organisation and also trusting them can affect how ontologies are evaluated and selected for reuse. Some of the interviewees argued that they would like to have some information about who else is using an ontology and why they are using it, before selecting it for reuse. More than 82% of the participants in the second phase also agreed with the statement about the critical role of communities in the ontology domain.

Trust was the other emerging factor in the analysis. More than 85% of the participants in the second phase of this research agreed to the statement about trusting other people's judgment about an ontology. In the literature, however, far too little attention has been paid to the importance of trust amongst the community of ontologists and knowledge engineers. The study by Lewen *et al.* (2006) is one of the very few examples of using the notion of trust in the ontology evaluation context. According to them, "trust in Open Rating Systems corresponds to the feeling that the information delivered by a certain reviewer will be correct and useful". Information in this statement refers to the ratings and reviews provided by other users.

The findings of this study also suggested that having access to community feedback and reviews can be very helpful in the selection process. More than 85% of the survey respondents agreed that ontologies could be evaluated via community feedback and reviews. A reusable ontology was also defined as the one that has been vetted by other knowledgeable users or recognised authorities as being useful and readily reusable or the one that is well-used and accepted in the community. Looking at the literature and the available selection systems, it can be seen that apart from popularity, there is a general lack of research about the other types of social metrics, such as ratings, feedback, and reviews for ontologies.

Community and social interactions amongst them have been discussed when comparing the usefulness of open and closed rating systems. When discussing the advantages and disadvantages of these systems, some of the participants in this study supported the gatekeeping policies applied by closed rating systems and argued that it would guarantee a certain level of quality. However, using rigorous inclusion conditions in closed rating systems has led to having repositories with a limited number of ontologies in specific domains, like biomedical. The very subjective nature of evaluation in closed rating systems and being based on what a limited group of experts think might also be problematic.

Regarding open rating systems, the combination of findings in different phases of this study provided some support for the idea of open rating systems and evaluating ontologies by community feedback and reviews. However, apart from the examples mentioned above and to the best of the researcher's knowledge, there is no live ontology selection system that allows its users to provide reviews, ratings or feedback on ontologies or uses ratings and reviews in the evaluating and ranking process.

Overall, the findings of this study highly recommend applying community related metrics to the process of ontology evaluation and selection for reuse. The results of the survey conducted by this research indicated that ontologists and knowledge engineers would like to have a selection system that not only applies different inclusion criteria and ensures that the ontologies in its repository have a minimum quality, but also allows them to comment on different aspects of ontologies and to interact with the people in their community.

If an open rating system is implemented, future work will be required to deal with difficulties and complexities of it, such as a large number of qualitative reviews, and to find answers to questions like "who will rate the raters" (Noy, Guha and Musen, 2005). However, dealing with all these issues and trying to find answers to these questions is hardly possible, unless there exists a selection system that allows the community of ontologists and knowledge engineers to express their views on the quality of ontologies.

7.4 Metrics Comparison and Usefulness

This research contributed to the existing knowledge by asking the largest group of ontologists and knowledge engineers (so far) about the quality of ontologies and the metrics it depends on. The findings of this study suggested 31 metrics, 14 of which were not previously discussed in the literature. This study also compared the importance of the identified metrics with each other and the ones previously used in the literature and ranked those metrics accordingly (Table 5-2). It was very interesting to see that some of the metrics identified in this research, especially the community related ones, were ranked higher than some of the metrics widely used in the literature. Participants in the third phase of this study also found it very helpful to have access to the information that this research is suggesting. These findings have important implications, especially for developing selection systems for ontologies.

7.5 Weight Assignment

Ontology search and selection systems not only need a set of metrics that can be used to assess the suitability of an ontology, but they should also assign weights to the importance of each of those metrics. Burton-Jones *et al.* (2005) argued that ontology quality could be considered as a formative construct, meaning that it is formed by different measures, all of which can equally be important but do not need to be, and may have a different level of importance assigned to them.

Assigning weights and values to the quality metrics can be considered as one of the most complicated tasks in the ontology domain. Some of the selection systems and frameworks like AKTiveRank (Alani, Brewster and Shadbolt, 2006) and RecoOn (Butt, Haller and Xie, 2016) assigned fixed weights to each of the metrics used in their evaluating and ranking process. NCBO BioPortal Recommender has predefined weights for each of the metrics used in its recommendation process; however, users of this system are provided with the facility of changing the weights of each metric.

Participants in different phases of this research, especially phase three, found the process of assigning weights to different evaluation and selection criteria very complicated and were not sure what the best way for figuring those weights was. Most of them believed that an initial weight should be assigned to the evaluation metrics by the selection systems. However, they were against using fixed weights for all the metrics and argued that the importance of each metric in the selection process depends on the purpose of the ontology selection, and therefore, users should be able to assign different weights and personalise the importance of each metric.

While being preliminary, the findings of this research, especially the ones related to metric comparison, can be used as a guideline for the initial weight assignment. However, the findings of this research firmly suggested that the metrics' weight should not be fixed, and users should be provided with facilities of changing the weight of each metric, based on their requirements.

7.6 Demographic Information

Despite all the development in ontology domain and the availability of different search and selection systems, it was not clear if the concept of quality and the factors it depends on has any association with the domain an ontology has been built in, or the years of experience the ontology evaluator and user have and the type of organisation they work for. Therefore, one of the questions of this research was to test different hypotheses and find out if demographic features of ontology users can affect their perspective on quality and how they evaluate it.

7.6.1 Domain Comparison

In this research, the link between the domain and ontology quality was explored using both qualitative and quantitative data. Interviews in the first round of the study showed some differences between how ontologies are used and evaluated in different domains. To investigate this matter further, different hypotheses were developed and tested in the second phase. Despite some minor differences, the overall results suggested that the set of metrics used in the evaluation and selection process are very similar across different domains. A very important implication of these findings is the possibility of reusing the available technologies in domains like biomedical by applying them to the search and selection systems in other domains.

This research also investigated the role and importance of community in the ontology domain. Overall, the findings indicated that the significance of the role of the community varies across different domains. People working in the biomedical domain, for example, suggested that communities play a more important role than those who were working in non-biomedical domains. They were also more in favour of collaborative ontology development and showed more tendency to contribute to existing ontologies in their domain rather than developing a new one.

Explaining this result is difficult, especially considering the fact that only 5% of the participants disagreed with the statement about the critical role of communities in the ontology domain. These differences can be related to the fact that some of the biomedical ontologies, like GO, are too big to be built and maintained only by a small group of people and, therefore, community collaboration and support are essential. Or, “maybe it is a part of a historical accident”, as one of the participants stated.

7.6.2 Years of Experience

This study investigated, for the first time, if there is any correlation between the years of experience ontologists and knowledge engineers have and how they rank the importance of quality factors in the evaluation process. The results suggested that people with different years of experience tend to rate the importance of most of the quality metrics, except the internal and popularity factors, similarly. Differences in the years of experience did not affect how people judge the importance of the community dramatically, either. In general, therefore, it seems that the number of years people have worked on ontologies does not affect their perception of quality and reusability.

7.6.3 Type of Organisation

Participants in different phases of this study were working across different types of organisations, including different universities, industries and companies. This study was interested in exploring if there is any link between the type of organisations ontologists and knowledge engineers work for and the importance they assign to each quality metric. Overall, the findings suggested that the type of organisation does not affect the perspective of ontologists and knowledge engineers about the quality and the factors that can be used to assess it.

7.7 Summary and Recommendations

The presented study was designed to clarify and address some of the most important topics in the ontology domain, which are the concept of quality, what it depends on, and how it is evaluated. This research also set out with the aim of assessing the importance of metrics used in the evaluation and selection process. The role and importance of community and social interactions, especially regarding the process of selecting and evaluating ontologies for reuse was examined. Finally, this research investigated whether there are differences in ontology domains, or other important idiosyncrasies deserving further attention.

The findings of this study suggested several courses of action for those who are involved in building and developing ontologies. Firstly, ontologies should be built by following the well-known principles and standards, such as the ones proposed by OBO Foundry. Secondly, developers should try to provide as much information as possible about their ontology, e.g., by providing proper documentation, GitHub page, and province information; ontology users are particularly interested in knowing who has built an ontology, why it has been built, how it has been built, what does it cover, and how it can be (re)used.

Moreover, ontology developers can facilitate the process of reuse by being responsive and flexible. The availability of wikis, forums, and mailing lists was suggested to be very important and helpful, especially if someone has questions or need alterations to be made to an ontology. Keeping an ontology up to date will also be useful, especially in domains like biomedical. The findings of this study also suggested that smaller ontologies and those built in OWL have a higher chance of being reused; the bigger ontologies, however, can be as reusable, if they are modular.

Regarding selection systems, several significant changes need to be made. Firstly, to address the needs of ontologists and knowledge engineers, selection systems should not only apply gatekeeping policies and have inclusion requirements but should also facilitate social interactions and allow ontologists to interact and review different aspects of ontologies in the system. Comparing open rating systems with closed rating systems showed that none of them is useful on their own, and a combination of them is needed in the ontology domain.

Participants in this research did not find some of the metrics used by the evaluation and selection systems useful. There is, therefore, a definite need for having selection frameworks, similar to the one proposed in this study, that not only consider different internal characteristics of ontologies but would also assess the metadata and social metrics while evaluating ontologies. Some of the proposed metrics by this study, namely, availability of metadata, availability of documentation, and being based on a standard can directly be used to evaluate and rank ontologies in the selection systems.

The findings of this research also recommended that information about some of the factors like frequency of updates and ontology size should be provided but cannot directly be used in the evaluation and ranking process, as there was no consensus amongst the respondents about how they should be measured. In terms of community related factors, e.g., responsiveness of ontology developer team, their reputation and popularity, there should first be some selection systems that allow interactions in the community and collection of these sets of information. When available, this information can then be analysed using recommendation algorithms and be used in the ontology recommendation process.

Finally, the weight assigned to different selection criteria should be flexible. Selection systems should allow their users to assign different importance to the metrics used in the selection process. The results indicated that ontologists and knowledge engineers assess different features of ontologies before selecting them for reuse. Therefore, it can be useful to allow them to choose the set of metrics they want to be used in the evaluation and ranking process.

Chapter 8: Conclusion

8.1 Introduction

This chapter aims to summarise the findings of this research in line with the aim and objectives previously discussed in Chapter 1:. It also discusses some of the limitations of this study and provides recommendations for future work.

8.2 Research Overview

This research set out to clarify the notions of quality and reuse in the ontology domain and to identify the set of metrics that ontologists and knowledge engineers tend to consider when assessing the suitability of an ontology for reuse. It also determined the process of ontology evaluation and selection and the set of steps that are usually taken in that process. Moreover, this research investigated the potential role of community and social interactions in the process of ontology evaluation and selection for reuse. The following parts of this chapter provide a summary of the main findings, together with the contributions and limitations of this study and some suggestions for future research.

8.3 Findings Summary

To achieve the research aim, five different objectives were set and have been met by the three phases of this research.

Objective 1. The first objective was to conduct an extensive critical survey of ontology evaluation techniques and systems. Ontology evaluation and selection for reuse is a very complicated task and depends on many different factors. Therefore, the researcher reviewed the literature in different domains, such as, ontology selection, libraries, search engines, and evaluation, ranking and recommendation approaches. More than 30 different selection algorithms, frameworks and systems were also reviewed to determine the general process of ontology selection for reuse and to identify the areas of potential improvement. These reviews are presented both in Chapter 2 and Appendix A-F.

Objective 2. The second objective was to study the notion of quality in the ontology domain, determine how ontologies are evaluated and selected, and identify and classify the set of metrics that are used in that process. To do that, first, an exploratory interview study was conducted (Chapter 4) and asked a group of ontologists and knowledge engineers about the general process of ontology evaluation, selection, and reuse. In the second phase (Chapter 5), quantitative data was collected through a survey questionnaire to confirm and generalize the findings of the first phase, which included a set of metrics that can be used in the evaluation and selection process. This objective was also addressed in the third phase (Chapter 6), by interviewing a group of experts in the ontology domain.

Objective 3. The third objective was to determine whether social and community interactions can affect the reusability of ontologies. The importance of community and social interactions amongst ontologists and knowledge engineers was first discussed by the participants in the first phase of this study (Chapter 4). In the second phase (Chapter 5), participants were presented with different metrics and statements about the role and importance of community in the process of evaluating and selecting ontologies for reuse. Overall, the findings of all the three phases of this study suggested that the community plays a significant role in the ontology domain and social interactions do affect the process of ontology selection for reuse.

Objective 4. The fourth objective was to determine if the choice of metrics used in the evaluation and selection process can be linked to the years of experience, domain, and organisation type of the ontology users. To address this objective, the following hypotheses were developed and tested in the second phase of this study (Chapter 5):

1. Ontologists and knowledge engineers with different years of experience rank the importance of quality metrics differently
2. Ontologists and knowledge engineers working in different domains rank the importance of quality metrics differently
3. Ontologists and knowledge engineers working in different organisations rank the importance of quality metrics differently

Overall, the findings suggested that people working in different domains and different types of organisations with different years of experience tend to evaluate ontologies very similarly.

Objective 5. The fifth objective was to construct and test a framework to facilitate the process of ontology selection for reuse. The findings of the first two phases and also reviewing different systems in the literature helped in constructing this framework. As it was seen in Chapter 6, this is the first time that an ontology selection framework is based on two different evaluation components, one responsible for evaluating ontologies using different query independent criteria, and the other one to compare the ontologies with users' queries and to find the one that best fits the their requirements.

Objective 6. The sixth objective was to provide recommendations to the ontology engineering community on how to build, evaluate and select ontologies. A summary of this research findings and different sets of recommendations for ontology developers and evaluators are presented in section 7.7. This study also identified a set of metrics that ontologists and knowledge engineers use in the evaluation process. This finding has important implication for developing the next generation of selection systems for ontologies.

8.4 Contributions

The findings of this study make several contributions to the current literature. Firstly, this study is the largest research project so far to document the general process of ontology evaluation and selection for reuse. This is also the first study to propose evaluation and selection metrics based on the findings of exploratory interview studies and a confirmatory survey study that asked more than 180 ontologists and knowledge engineers what they thought the most important metrics for evaluating ontologies were. This differs from most of the other studies in this domain, which first proposed a set of metrics the researchers find important in the evaluation process and then validate those metrics by asking a small number of respondents.

Secondly, unlike many of the other studies in the ontology selection domain, the findings reported in this thesis highlight the importance of metadata and social interactions among the community of ontologists and knowledge engineers in the process of ontology evaluation and selection for reuse. The results of this research suggested that the availability of different sets of metadata and additional information about ontologies is much more important than some of the metrics used by the current selection systems for ontology evaluation and recommendation.

Thirdly, the findings of this research provide a new understanding of one of the most used, and maybe the only used, social metric for ontology evaluation, the popularity of an ontology. Participants in different phases of this research found the popularity metrics identified by this study, such as, (1) popularity of an ontology in the community and amongst colleagues, and (2) the reputation of the ontology developer team, and/or institute in the domain more important and useful than the popularity metrics that have been widely used in the literature and by different selection systems, e.g., the presence of an ontology in different repositories and the number of page views (Martínez-Romero *et al.*, 2017).

Finally, the findings of this research were used to develop an ontology evaluation and recommendation framework, which differs from the available ones in the literature in a number of important ways. This framework is based on two evaluation components; one to evaluate the query independent criteria, like popularity, social aspects, and metadata, and the second, to identify an ontology that best matches the users' requirements in terms of coverage and internal aspects of ontologies.

8.5 Limitations

Despite the novelty of this research in clarifying the notion of quality in the ontology domain and its empirical and theoretical contributions, there remain limitations. Due to practical constraints, this research was unable to measure some of the metrics that were identified in different phases of this research. None of the current systems for ontologies provide their users with the facility of adding ratings and reviews for ontologies. Therefore, this study was limited by the lack of information on the social aspects, and it could not quantify the value of two of the identified metrics, namely, the reputation of the ontology developer team and their trustworthiness in the community.

The third phase of this study validates the findings of the previous phases by applying them to a set of ontologies and also interviewing a group of experts. While being useful, the scope of this experiment was limited to the biomedical domain. The reason for this is that the information required in the validation phase was mostly available for ontologies in the biomedical domain, and not some of the other domains.

The overall findings of this study suggested that ontology quality depends on three different dimensions. The focus of this research, however, was on two of the identified dimensions, metadata and social aspects of ontologies. Therefore, this thesis did not engage with the internal aspects of ontologies and the methods that can be used to measure them. Moreover, due to practical and time constraints, the framework proposed by this research could not be implemented.

8.6 Recommendations for Future Work

As a part of this study, participants were asked how they would make an ontology reusable and also about the characteristics of reusable ontologies or the best ontology they have ever reused. Features identified and suggested as the answer to these questions can be used for proposing a formal set of principles that ontologists and knowledge engineers should follow while building an ontology and to ensure that their ontology is reusable. The third experiment presented in Chapter 6 can also be extended by applying the identified features to a larger set of ontologies in different domains. This will help verify the usefulness and applicability the proposed features in different domains.

The findings of this study provided a new understanding of the notion of quality and reusability in the ontology domain and the factors they depend on. Many of the metrics identified in this study are not used by the current selection systems for ontologies. Therefore, the findings of this research have important implications for developing the next generation of selection systems for ontologies. The systems that should not only evaluate ontologies according to their internal characteristics, but also judge the quality of ontologies by considering sets of additional information about them and the social interactions amongst the community, such as ratings and reviews about different aspects of ontologies.

Despite the availability of different search and selection systems for ontologies, participants in the second phase of this study chose literature (Google Scholar) and Google as the first two means of searching for ontologies for reuse. While being very popular amongst most of the participants, Google was blamed for not doing enough regarding ontology search and selection. Some similarities exist between the ontology search and document/text search, as it is done in Google. Both processes are mostly keyword-based and use page rank related algorithms to identify the most suitable match for different queries. However, more research needs to be undertaken to investigate how Google can be more helpful to the community of ontologists and knowledge engineers.

In the literature, one way to help users find what they look for is to analyse their interactions in a social environment by using different recommendation algorithms. Despite their importance and the potential benefits, recommendation algorithms have rarely been applied to the ontology domain. The findings of this research suggested that social interactions can affect the process of ontology evaluation and selection for reuse. Therefore, further research can introduce the notion of recommendation, as used in Amazon[48], to the ontology domain and determine the usefulness of different recommendation algorithms, e.g., collaborative filtering in this domain.

8.7 Benefits of this Research

This research extends our knowledge of ontology quality and the factors that it depends on. The findings of this study help the community of ontologists and knowledge engineers by facilitating the process of ontology evaluation and selection for reuse. Moreover, the framework proposed by this research can be used to develop the next generation of selection and recommendation systems for ontologies. Lastly, and by discussing the characteristics of reusable ontologies, this research helps ontology developers and knowledge engineers in developing ontologies that have higher level of quality and are easier to reuse.

[48] https://www.amazon.co.uk/

References

Acharya, A. S. *et al.* (2013) ‘Sampling: Why and How of it?’, *Indian Journal of Medical Specialities*, 4(2). doi: 10.7713/ijms.2013.0032.

Alani, H., Brewster, C. and Shadbolt, N. (2006) ‘Ranking Ontologies with AKTiveRank’, in *International Semantic Web Conference*. Athens, GA, USA: Springer, Berlin, Heidelberg, pp. 1–15. doi: 10.1007/11926078_1.

Alani, H., Noy, N., Shah, N., Shadbolt, N. and Musen, M. (2007) ‘Searching Ontologies Based on Content: Experiments in the Biomedical Domain’, in *The Fourth International Conference on Knowledge Capture (K-Cap)*. ACM, pp. 55–62.

Ameen, A., Khan, K. U. R. and Rani, B. P. (2012) ‘Creation of Ontology in Education Domain’, in *2012 IEEE Fourth International Conference on Technology for Education*. Hyderabad India: IEEE, pp. 237–238. doi: 10.1109/T4E.2012.50.

Amith, M. ‘Tuan’, Manion, F., Liang, C., Harris, M., Wang, D., He, Y. and Tao, C. (2018) ‘OntoKeeper: Semiotic-driven Ontology Evaluation Tool For Biomedical Ontologists’, in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Madrid, Spain, Spain: IEEE, pp. 1614–1617. doi: 10.1109/BIBM.2018.8621458.

Annamalai, M. and Sterling, L. (2003) ‘Guidelines for Constructing Reusable Domain Ontologies’, in *Proceedings of the Workshop on Ontologies in Agent Systems (OAS 2003) at the 2nd International Joint Conference on Autonomous Agents and Multi-Agent Systems*. Melbourne, Australia, pp. 71–74.

Arp, R. and Smith, B. (2008) ‘Function, role and disposition in Basic Formal Ontology’, in *Nature Proceedings*.

Arpinar, I. B., Giriloganathan, K. and Aleman-Meza, B. (2006) ‘Ontology quality by detection of conflicts in metadata’, in *4th International EON Workshop*. Edinburgh, United Kingdom.

Arpírez, J., Gómez-Pérez, A., Lozano Tello, A. and Pinto, H. S. A. N. P. (1998) ‘(ONTO)2Agent: An ontology-based WWW broker to select ontologies’, in *Proceedings of ECAI98’s Workshop on Application of Ontologies and Problem Solving Methods*. Brighton, pp. 16–24.

Bandeira, J., Bittencourt, I. I., Espinheira, P. and Isotani, S. (2016) *FOCA: A Methodology for Ontology Evaluation*.

Bard, J. B. L. and Rhee, S. Y. (2004) ‘Ontologies in biology: design, applications and future challenges’, *Nature Reviews Genetics*. Nature Publishing Group, 5(3), pp. 213–222. doi: 10.1038/nrg1295.

Basili, V. R. (1993) ‘Applying the Goal/Question/Metric paradigm in the experience factory’, *Software Quality Assurance and Measurement: A Worldwide Perspective*, 7(4), pp. 21–44.

Bates, N., Atrostic, B. K., Burt, G. and Silberstein, A. (2001) ‘Nonresponse in U.S. Government Household Surveys: Consistent Measures, Recent Trends, and New Insights’, *Journal of Official Statistics*, 17(2), pp. 209–226.

Bernard, H. R. (2017) *Research methods in anthropology : qualitative and quantitative approaches*. Fourth Edi. Rowman & Littlefield.

Bontas, E. P., Mochol, M. and Tolksdorf, R. (2005) 'Case Studies on Ontology Reuse', in *Proceedings of the IKNOW05 International Conference on Knowledge Management*, pp. 345–353. doi: 10.1016/j.sysarc.2006.02.002.

Brank, J., Grobelnik, M. and Mladenic, D. (2005) 'A survey of ontology evaluation techniques', in *Proceedings of the conference on data mining and data warehouses (SiKDD 2005)*. Ljubljana, Slovenia, pp. 166–170.

Brank, J., Mladenic, D. and Grobelnik, M. (2006) 'Gold standard based ontology evaluation using instance assignment', in *Workshop on Evaluation of Ontologies for the Web, EON*.

Brewster, C., Alani, H., Dasmahapatra, S. and Wilks, Y. (2004) 'Data Driven Ontology Evaluation', in *4th International Conference on Language Resources and Evaluation, LREC'04*. Lisbon, Portugal.

Brick, J. M. (2013) 'Unit Nonresponse and Weighting Adjustments: A Critical Review', *Journal of Official Statistics*, 29(3), pp. 329–353. doi: 10.2478/jos-2013-0026.

Brickley, D. and Miller, L. (2010) *FOAF Vocabulary Specification 0.98*. Available at: http://xmlns.com/foaf/spec/20100809.html.

Brown, T. A. (2014) *Confirmatory Factor Analysis for Applied Research*. Second Edi. The Guilford Press.

Bryman, A. (2016) *Social Research Methods*. Fifth Edit. Oxford : Oxford University Press.

Buitelaar, P. (2004) 'OntoSelect: Towards the Integration of an Ontology Library, Ontology Selection and Knowledge Markup', in *Proc. of the Workshop on Knowledge Markup and Semantic Annotation (Semannot 2004) at the International Semantic Web Conference*. Hiroshima, Japan, pp. 127–128.

Buitelaar, P. and Eigner, T. (2007) 'Evaluating Ontology Search', in *5th International EON Workshop (EON2007)*, pp. 11–20.

Buitelaar, P. and Eigner, T. (2008) 'Ontology Search with the OntoSelect Ontology Library', in *6th International Conference on Language Resources and Evaluation (LREC'08)*. Marrakech, Morocco.

Buitelaar, P., Eigner, T. and Declerck, T. (2004) 'OntoSelect: A Dynamic Ontology Library with Support for Ontology Selection', in *Proceedings of the Demo Session at the International Semantic Web Conference*. Hiroshima, Japan, pp. 3–6.

Bürger, T. and Simperl, E. (2008) 'Measuring the Benefits of Ontologies', in *OTM Confederated International Conferences 'On the Move to Meaningful Internet Systems'*. Monterrey, Mexico: Springer, Berlin, Heidelberg, pp. 584–594. doi: 10.1007/978-3-540-88875-8_82.

Burton-Jones, A., Storey, V. C., Sugumaran, V. and Ahluwalia, P. (2005) 'A semiotic metrics suite for assessing the quality of ontologies', *Data and Knowledge Engineering*. North-Holland, 55(1), pp. 84–102. doi: 10.1016/j.datak.2004.11.010.

Butt, A. S. (2017) *Ontology Ranking: Finding the Right Ontologies on the Web*. The Australian National University. Available at: https://openresearch-repository.anu.edu.au/bitstream/1885/125224/1/Butt Thesis 2017.pdf.

Butt, A. S., Haller, A. and Xie, L. (2016) 'RecOn: Ontology Recommendation for

Structureless Queries', *Applied Ontology*. IOS Press, 11(4), pp. 301–324. doi: 10.3233/AO-160173.

Caldarola, E. G., Picariello, A. and Rinaldi, A. M. (2015) 'An Approach to Ontology Integration for Ontology Reuse in Knowledge Based Digital Ecosystems', in *Proceedings of the 7th International Conference on Management of computational and collective intElligence in Digital EcoSystems*. ACM, pp. 1–8. doi: 10.1145/2857218.2857219.

Caldarola, E. G. and Rinaldi, A. (2016) 'An Approach to Ontology Integration for Ontology Reuse', in *2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)*. IEEE, pp. 384–393.

Cantador, I., Fernandez, M. and Castells, P. (2007) 'Improving Ontology Recommendation and Reuse in WebCORE by Collaborative Assessments', in *Proceedings of the 1st International Workshop on Social and Collaborative Construction of Structured Knowledge CKC 2007 at the at the 16th International World Wide Web Conference WWW 2007*.

Cantador, I., Fernández, M. and Castells, P. (2006) 'A Collaborative Recommendation Framework for Ontology Evaluation and Reuse', in *Actas de International Workshop on Recommender Systems, en la 17th European Conference on Artificial Intelligence (ECAI 2006)*. Garda, Italy, pp. 67–71.

Cao, H., Jiang, D., Pei, J., Chen, E. and Li, H. (2009) 'Towards context-aware search by learning a very large variable length hidden markov model from search logs', in *Proceedings of the 18th international conference on World wide web*. New York, New York, USA: ACM, pp. 191–200. doi: 10.1145/1526709.1526736.

Cardoso, J. (2007) 'The semantic web vision: Where are we?', *IEEE Intelligent Systems*, 22(5), pp. 84–88. doi: 10.1109/MIS.2007.4338499.

Chandrasekaran, B., Josephson, J. R. and Benjamins, V. R. (1999) 'What Are Ontologies, and Why Do We Need Them?', *IEEE Intelligent systems*, 14(1), pp. 20–26. doi: 10.1109/5254.747902.

Cheng, G., Ge, W. and Qu, Y. (2008) 'Falcons: searching and browsing entities on the semantic web', in *Proceedings of the 17th international conference on World Wide Web*. Beijing: ACM, pp. 1101–1102. doi: 10.1145/1367497.1367676.

Churchill, J. G. A. (1979) 'A paradigm for Developing Better Measures of Marketing Constructs', *Journal of Marketing Research*, 16(1), pp. 64–73. doi: 10.1017/CBO9781107415324.004.

Comrey, A. L. and Lee, H. B. (1992) *A first course in factor analysis*. Second Edi. Psychology Press. doi: 10.4324/9781315827506.

Corcho, Ó., Gómez-Pérez, A., González-Cabero, R. and Suárez-Figueroa, M. C. (2004) 'ODEval: A Tool for Evaluating RDF(S), DAML+OIL, and OWL Concept Taxonomies', in *IFIP International Conference on Artificial Intelligence Applications and Innovations*. Boston: Springer, pp. 369–382. doi: 10.1007/1-4020-8151-0_32.

Corder, G. W. and Foreman, D. I. (2014) *Nonparametric statistics : a step-by-step approach*. Second Edi. John Wiley & Sons.

Costello, A. B. and Osborne, J. W. (2005) 'Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis', *Practical assessment, research & evaluation*, 10(7).

Côté, R. G., Jones, P., Apweiler, R. and Hermjakob, H. (2006) ‘The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries’, *BMC Bioinformatics*. BMC, 7, p. 97. doi: 10.1186/1471-2105-7-97.

Côté, R. G., Jones, P., Martens, L., Apweiler, R. and Hermjakob, H. (2008) ‘The Ontology Lookup Service: more data and better tools for controlled vocabulary queries’, *Nucleic Acids Research*. Narnia, 36(Web Server issue), pp. W372–W376. doi: 10.1093/nar/gkn252.

Creswell, J. and Clark, V. (2017) *Designing and conducting mixed methods research*. Sage publications.

Creswell, J. W. (1994) *Research design : qualitative & quantitative approaches*. Sage Publications.

Creswell, J. W. (2009) *Research design : qualitative, quantitative, and mixed methods approaches*. Third Edit. SAGE Publications.

Creswell, J. W. (2014) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Forth Edit. SAGE Publications, Incorporated, 2013.

Creswell, J. W. and Creswell, J. D. (2018) *Research Design Qualitative, Quantitative & Mixed Methods Approaches*. Fifth Edit. SAGE Publications.

Creswell, J. W. and Plano Clark, V. L. (2007) *Designing and Conducting Mixed Methods Research*. SAGE Publications, Inc.

Crotty, M. J. (1998) *The Foundations of Social Research: Meaning and Perspective in the Research Process*. first. SAGE Publications.

Cui, Z., Damiani, E. and Leida, M. (2007) ‘Benefits of ontologies in real time data access’, in *2007 Inaugural IEEE International Conference on Digital Ecosystems and Technologies (IEEE DEST 2007)*, pp. 392–397.

d’Aquin, M., Baldassarre, C., Gridinoc, L., Angeletou, S., Sabou, M., Motta, E., Aquin, M., Baldassarre, C., Gridinoc, L., Angeletou, S., Sabou, M., Motta, E. and D’Aquin, M. (2007) ‘Watson: A Gateway for Next Generation Semantic Web Applications’, in *The 6th International Semantic Web Conference (ISWC 2007)*. Busan, Korea, p. 3.

d’Aquin, M. and Motta, E. (2011) ‘Watson, more than a Semantic Web search engine’, *Semantic Web*, 2(1), pp. 55–63.

d’Aquin, M. and Noy, N. F. (2012) ‘Where to publish and find ontologies? A survey of ontology libraries’, *Journal of Web Semantics*. Elsevier B.V., 11, pp. 96–111. doi: 10.1016/j.websem.2011.08.005.

d’Aquin, M., Sabou, M., Motta, E., Angeletou, S., Lopez, V. and Zablith, F. (2008) ‘What can be done with the Semantic Web ? An Overview of Watson-based Applications’, in *Proceedings of the 5th Workshop on Semantic Web Applications and Perspectives (SWAP2008)*.

Dabbish, L., Stuart, C., Tsay, J. and Herbsleb, J. (2012) ‘Social Coding in GitHub: Transparency and Collaboration in an Open Software Repository’, in *Proceedings of the ACM 2012 conference on computer supported cooperative work*. ACM, pp. 1277–1286.

De Nicola, A. and Missikoff, M. (2016) ‘A Lightweight Methodology for Rapid Ontology Engineering’, *Communications of the ACM*, 59(3), pp. 79–86. doi: 10.1145/2818359.

De Nicola, A., Missikoff, M. and Navigli, R. (2009) 'A software engineering approach to ontology building', *Information Systems*. Pergamon, 34(2), pp. 258–275. doi: 10.1016/J.IS.2008.07.002.

DeVellis, R. F. (2012) *Scale development : theory and applications*. Third Edit. SAGE Publications.

Ding, L., Finin, T., Joshi, A., Peng, Y., Cost, R. S., Sachs, J., Pan, R., Reddivari, P. and Doshi, V. (2004) 'Swoogle: A Semantic Web Search and Metadata Engine', in *Proc. 13th ACM Conf. on Information and Knowledge Management*.

Ding, L., Pan, R., Finin, T., Joshi, A., Peng, Y. and Kolari, P. (2005) 'Finding and ranking knowledge on the semantic web', in *International Semantic Web Conference*. Springer, Berlin, Heidelberg, pp. 156–170. doi: 10.1007/11574620_14.

Ding, Y. and Fensel, D. (2001) 'Ontology Library Systems: The key to successful Ontology Re-use', in *Proceedings of the First International Conference on Semantic Web Working Symposium (SWWS'01)*. California, USA, pp. 93–112.

Ding, Y. and Foo, S. (2002) 'Ontology research and development. Part 2 - a review of ontology mapping and evolving', *Journal of Information Science*, 28(5), pp. 375–388. doi: 10.1177/016555150202800503.

Djedidi, R. and Aufaure, M.-A. (2010) *ONTO-EVOAL an Ontology Evolution Approach Guided by Pattern Modeling and Quality Evaluation*, *Foundations of Information and Knowledge Systems*. doi: \url{10.1007/978-3-642-11829-6_19}.

Domingue, J. (1998) 'Tadzebao and WebOnto: discussing, browsing, and editing ontologies on the Web', in *Eleventh Workshop on Knowledge Acquisition,Modeling and Management*. Banff, Alberta.

Dramé, K., Diallo, G., Delva, F., Dartigues, J. F., Mouillet, E., Salamon, R. and Mougin, F. (2014) 'Reuse of termino-ontological resources and text corpora for building a multilingual domain ontology: An application to Alzheimer's disease', *Journal of Biomedical Informatics*, 48, pp. 171–182. doi: 10.1016/j.jbi.2013.12.013.

Duque-Ramos, A., Fernández-Breis, J. T., Iniesta, M., Dumontier, M., Egaña Aranguren, M., Schulz, S., Aussenac-Gilles, N. and Stevens, R. (2013) 'Evaluation of the OQuaRE framework for ontology quality', *Expert Systems with Applications*. Pergamon, 40(7), pp. 2696–2703. doi: 10.1016/J.ESWA.2012.11.004.

Ekstrand, M. D., Riedl, J. T. and Konstan, J. A. (2011) 'Collaborative Filtering Recommender Systems', *Foundations and Trends® in Human–Computer Interaction*. Now Publishers, Inc., 4(2), pp. 81–173. doi: 10.1561/1100000009.

Elo, S. and Kyngäs, H. (2008) 'The qualitative content analysis process', *Journal of Advanced Nursing*, 62(1), pp. 107–115. doi: 10.1111/j.1365-2648.2007.04569.x.

Etikan, I., Musa, S. A. and Alkassim, R. S. (2016) 'Comparison of Convenience Sampling and Purposive Sampling', *American Journal of Theoretical and Applied Statistics*. Science Publishing Group, 5(1), pp. 1–4. doi: 10.11648/j.ajtas.20160501.11.

Evans, J. R. and Mathur, A. (2005) 'The value of online surveys', *Internet Research*. Emerald Group Publishing Limited, 15(2), pp. 195–219. doi: 10.1108/10662240510590360.

Fahad, M. and Qadir, M. A. (2008) 'A framework for ontology evaluation', *CEUR Workshop*

Proceedings, 354, pp. 148–156. doi: 10.1007/978-3-642-17449-0_2.

Farquhar, A., Fikes, R. and Rice, J. (1997) 'The Ontolingua Server: A Tool for Collaborative Ontology Construction', *International journal of human-computer studies*, 46(6), pp. 707–727.

Feilmayr, C. and Wöß, W. (2016) 'An analysis of ontologies and their success factors for application to business', *Data & Knowledge Engineering*. North-Holland, 101, pp. 1–23. doi: 10.1016/J.DATAK.2015.11.003.

Fernabdez, M., Gomez-Perez, A. and Juristo, N. (1997) *METHONTOLOGY: From Ontological Art Towards Ontological Engineering*.

Fernández-López, M. and Gómez-Pérez, A. (2002) 'The integration of OntoClean in WebODE', in *Proceedings of the 1st Workshop on Evaluation of Ontology-based Tools (EON2002) , held at the 13th International Conference on Knowledge Engineering and Knowledge Management EKAW 2002 (2002)*, pp. 38–52.

Fernández, M., Cantador, I. and Castells, P. (2006) 'CORE: A Tool for Collaborative Ontology Reuse and Evaluation', in *Proceedings of the 4th International EON Workshop (EON 2006) at the 15th International World Wide Web Conference (WWW 2006)*. Edinburgh, UK.

Fernández, M., Overbeeke, C., Sabou, M. and Motta, E. (2009) 'What makes a good ontology? A case-study in fine-grained knowledge reuse', in *Proceedings of the 4th Asian Conference on The Semantic Web*. Shanghai, China: Springer, Berlin, Heidelberg, pp. 61–75. doi: 10.1007/978-3-642-10871-6_5.

Fernández-López, M., Poveda-Villalón, M., Suárez-Figueroa, M. C. and Gómez-Pérez, A. (2019) 'Why are ontologies not reused across the same domain?', *Journal of Web Semantics*. Elsevier, 57. doi: 10.1016/J.WEBSEM.2018.12.010.

Fetters, M. D., Curry, L. A. and Creswell, J. W. (2013) 'Achieving Integration in Mixed MethodsDesigns—Principles and Practices', *Health services research*, 48(6pt2), pp. 2134–2156.

Finin, T., Ding, L., Pan, R., Joshi, A., Kolari, P., Java, A. and Peng, Y. (2004) 'Swoogle : Searching for Knowledge on the Semantic Web ∗', *Data Engineering*, pp. 1682–1683.

Finin, T., Ding, L., Pan, R., Joshi, A., Kolari, P., Java, A. and Peng, Y. (2005) 'Swoogle: Searching for knowledge on the Semantic Web', in *Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI 05)*. Pittsburg, Pennsylvania, pp. 1682–1683.

Fox, M. S., Barbuceanu, M. and Gruninger, M. (1996) 'An organisation ontology for enterprise modeling: Preliminary concepts for linking structure and behaviour', *Computers in industry*, 29(1–2), pp. 123–134. doi: 10.1109/ENABL.1995.484550.

Franco, M., Vivo, J. M., Quesada-Martínez, M., Duque-Ramos, A. and Fernández-Breis, J. T. (2019) 'Evaluation of ontology structural metrics based on public repository data', *Briefings in Bioinformatics*. doi: 10.1093/bib/bbz009.

Frolov, V., Mengel, D., Bandara, W., Sun, Y. and Ma, L. (2009) 'Building an ontology and process architecture for engineering asset management', in *Proceeding of the 4th World Congress on Engineering Asset Management*. Athens: Springer, pp. 86–97. doi: 10.1007/978-0-85729-320-6_11.

Gangemi, A., Catenacci, C., Ciaramita, M. and Lehmann, J. J. (2006) ‘Modelling Ontology Evaluation and Validation’, in *Proceedings of the 3rd European Semantic Web Conference (ESWC2006)*. Springer, pp. 140–154. doi: 10.1007/11762256_13.

Geller, J., Keloth, V. K. and Musen, M. A. (2018) ‘How Sustainable are Biomedical Ontologies?’, in *AMIA Annual Symposium proceedings*. American Medical Informatics Association, pp. 470–479.

Gerbing, D. W. and Anderson, J. C. (1988) ‘An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment’, *Journal of Marketing Research*, 25(2), pp. 186–192.

Gill, T. (2008) ‘Metadata and the Web’, in Baca, M. (ed.) *Introduction to metadata*. Second Edi. Los Angeles: The Getty Research Institute, pp. 20–37.

Gómez-Pérez, A. (1994) *From Knowledge Based Systems to Knowledge Sharing Technology: Evaluation and Assessment*. Palo Alto, CA.

Gómez-Pérez, A. (1995) ‘Some ideas and examples to evaluate ontologies’, in *Proceedings of the 11th Conference on Artificial Intelligence for Applications*. IEEE, pp. 299–305. doi: 10.1109/CAIA.1995.378808.

Gómez-Pérez, A. (1999) ‘Evaluation of Taxonomic Knowledge in Ontologies and Knowledge Bases’, in *In Proceedings of the 12th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop*. Banff, Alberta, Canada.

Gray, D. E. (2013) *Doing research in the real world*. Third Edit. SAGE.

Gruber, T. R. (1993) ‘A Translation Approach to Portable Ontology Specifications’, *Knowledge Acquisition*, 5(2), pp. 199–220.

Guarino, N., Oberle, D. and Staab, S. (2009) ‘What Is an Ontology?’, in Staab, S. and Studer, R. (eds) *Handbook on Ontologies*. Springer, Berlin, Heidelberg, pp. 1–17.

Guarino, N. and Welty, C. (2002) ‘Evaluating ontological decisions with OntoClean’, *Communications of the ACM*, 45(2), pp. 61–65. doi: 10.1145/503124.503150.

Guba, E. G. and Lincoln, Y. Y. S. (1989) *Fourth generation evaluation*. Sage Publications.

Guest, G., Bunce, A. and Johnson, L. (2006) ‘How Many Interviews Are Enough?An Experiment with Data Saturation and Variability’, *Field Methods*, 18(1), pp. 59–82. doi: 10.1177/1525822X05279903.

Gunter, B., Nicholas, D., Huntington, P. and Williams, P. (2002) ‘Online versus offline research: implications for evaluating digital media’, *Aslib Proceedings*, 54(4), pp. 229–239. doi: 10.1108/00012530210443339.

Hartmann, J., Palma, R. and Sure, Y. (2005) ‘OMV-ontology metadata vocabulary’, in *Proceedings of the International Workshop on Ontology Patterns for the Semantic Web (OPSW) at ISWC2005*. Galway.

Hartmann, J., Spyns, P., Giboin, A. and Maynard, D. (2005) *Methods for Ontology Evaluation*, *Knowledge Web Deliverable D*. doi: KWEB/2004/D1.2.3/v1.3.

Hecke, T. Van (2012) ‘Power study of anova versus Kruskal-Wallis test’, *Journal of Statistics and Management Systems*. Taylor & Francis Group, 15(2–3), pp. 241–247. doi: 10.1080/09720510.2012.10701623.

Heflin, J. and Hendler, J. (2000) ‘Dynamic Ontologies on the Web’, in *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI-2000)*. Menlo Park, CA: AAAI/MIT Press, pp. 443–449.

Hinkin, T. R., Tracey, J. B. and Enz, C. A. (1997) ‘SCALE CONSTRUCTION: DEVELOPING RELIABLE AND VALID MEASUREMENT INSTRUMENTS’, *Journal of Hospitality & Tourism Research*, 21(1), pp. 100–120.

Hlomani, H. and Stacey, D. (2014) ‘Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey’, *Semantic Web Journal*, 1(5), pp. 1–11.

Hoehndorf, R., Slater, L., Schofield, P. N. and Gkoutos, G. V (2015) ‘Aber-OWL: a framework for ontology-based data access in biology’, *BMC Bioinformatics*. BioMed Central, 16, pp. 1–9. doi: 10.1186/s12859-015-0456-9.

Hong, T. P., Chang, W. C. and Lin, J. H. (2005) ‘A Two-Phased Ontology Selection Approach for Semantic Web’, in *Knowledge-Based Intelligent Information and Engineering Systems*. Melbourne, Australia: Springer, Berlin, Heidelberg, pp. 403–409. doi: 10.1007/11554028_56.

Hox, J. J. and Boeije, H. R. (2005) ‘Data Collection, Primary vs. Secondary’, in Kempf-Leonard, K. (ed.) *Encyclopedia of Social Measurement*. Elsevier, pp. 593–599. doi: 10.1016/B0-12-369398-5/00041-4.

IBM (2016) ‘SPSS’. Available at: https://www.ibm.com/analytics/spss-statistics-software.

Iqbal, R., Azrifah Azmi Murad, M., Mustapha, A. and Mohd Sharef, N. (2013) ‘An Analysis of Ontology Engineering Methodologies: A Literature Review’, *Research Journal of Applied Sciences, Engineering and Technology*, 6(16), pp. 2993–3000.

ISO (2005) *IEC 25000 Software and system engineering–Software product Quality Requirements and Evaluation (SQuaRE)–Guide to SQuaRE*. Geneva, Switzerland: International Organization for Standardization.

Jones, D., Bench-Capon, T. and Visser, P. (1998) ‘Methodologies For Ontology Development’, in *Proceedings of ITi & KNOWS Conference of the 15th IFIP World Computer Congress*, pp. 62–75.

Jonquet, C., Toulet, A., Arnaud, E., Aubin, S., Dzalé-Yeumo, E., Emonet, V., Graybeal, J., Musen, M. A., Pommier, C. and Larmande, P. (2016) ‘Reusing the NCBO BioPortal technology for agronomy to build AgroPortal’, in Jaiswal, P. and Hoehndorf, R. (eds) *7th International Conference on Biomedical Ontologies, ICBO’16, Demo Session*. Corvallis, Oregon, USA, pp. 4–6.

Jonquet, C. (2017) ‘Challenges for ontology repositories and applications to biomedicine & agronomy’, in *SIMBig: Symposium on Information Management and Big Data*. Lima, Peru, pp. 25–37.

Jonquet, C., Toulet, A., Arnaud, E., Aubin, S., Dzalé Yeumo, E., Emonet, V., Graybeal, J., Laporte, M.-A., Musen, M. A., Pesce, V. and Larmande, P. (2018) ‘AgroPortal: A vocabulary and ontology repository for agronomy’, *Computers and Electronics in Agriculture*. Elsevier, 144, pp. 126–143. doi: 10.1016/J.COMPAG.2017.10.012.

Jonquet, C., Musen, M. A. and Shah, N. H. (2010) ‘Building a Biomedical Ontology Recommender Web Service’, *Journal of biomedical semantics*. BioMed Central, 1(1), p. S1. doi: 10.1186/2041-1480-1-S1-S1.

Jonquet, C., Shah, N. and Musen, M. (2009) ‘Prototyping a biomedical ontology recommender service’, in *Bio-Ontologies: Knowledge in Biology, ISMB/ECCB SIG*. Stockholm.

Kaiser, H. F. (1960) ‘The Application of Electronic Computers to Factor Analysis’, *Educational and Psychological Measurement*. Sage PublicationsSage CA: Thousand Oaks, CA, 20(1), pp. 141–151. doi: 10.1177/001316446002000116.

Kalfoglou, Y. and Schorlemmer, M. (2003) ‘Ontology Mapping: The State of the Art’, *The knowledge engineering review*, 18(1), pp. 1–31.

Kamdar, M. R., Tudorache, T. and Musen, M. A. (2017) ‘A Systematic Analysis of Term Reuse and Term Overlap across Biomedical Ontologies’, *Semantic web*. NIH Public Access, 8(6), pp. 853–871.

Kehagias, D. D. D. *et al.* (2008) ‘A methodological approach for ontology evaluation and refinement’, in *ASK-IT Final Conference*, pp. 1–13.

Killam, L. (2013) *Research terminology simplified: Paradigms, axiology, ontology, epistemology and methodology*.

Kolbe, N., Kubler, S., Le Traon, Y., Robert, J. and Zaslavsky, A. (2019) ‘Linked Vocabulary Recommendation Tools for Internet of Things: A Survey’, *ACM Computing Surveys*, 51(6), p. 127.

Kvale, S. (2008) *Doing Interviews*. SAGE Publications.

Lavrakas, P. (2008) *Encyclopedia of survey research methods*. Edited by P. Lavrakas.

Lewen, H. and d’Aquin, M. (2010) ‘Extending Open Rating Systems for Ontology Ranking and Reuse’, in *International Conference on Knowledge Engineering and Knowledge Management*. Springer Berlin Heidelberg, pp. 441–450.

Lewen, H., Supekar, K., Noy, N. F. and Musen, M. A. (2006) ‘Topic-Specific Trust and Open Rating Systems : An Approach for Ontology Evaluation’, in *Proceedings of the 4th International Workshop on Evaluation of Ontologies for the Web (EON 2006) at the 15th International World Wide Web Conference (WWW 2006)*.

Lim, S. C. J., Liu, Y. and Chen, Y. (2015) ‘Ontology in design engineering: status and challenges’, in *International Conference on Engineering Design (ICED15)*. Milan, Italy, pp. 1–10.

Lozano-Tello, A. (2002) *Métrica de idoneidad de ontologías*. Universidad de Extremadura.

Lozano-Tello, A. and Gómez-Pérez, A. (2004) ‘ONTOMETRIC: A Method to Choose the Appropriate Ontology’, *Journal of Database Management*, 15(2), pp. 1–18. doi: 10.1109/MC.2002.1046975.

Maccallum, R. C., Widaman, K. F., Zhang, S. and Hong, S. (1999) ‘Sample Size in Factor Analysis’, *Psychological Methods*, 4(1), pp. 84–99.

Mace, J. C., Parkin, S. and van Moorsel, A. (2010) ‘A collaborative ontology development tool for information security managers’, in *Proceedings of the 4th Symposium on Computer Human Interaction for the Management of Information Technology - CHiMiT ’10*. New York, New York, USA: ACM, p. 5. doi: 10.1145/1873561.1873566.

Maedche, A. and Staab, S. (2002) ‘Measuring similarity between ontologies’, in *International*

Conference on Knowledge Engineering and Knowledge Management, pp. 251–263. doi: 10.1007/3-540-45810-7_24.

Maiga, G. (2008) 'A Flexible Approach for User Evaluation of Biomedical Ontologies', *International Journal*, 2(2), pp. 62–74.

Maiga, G. and Williams, D. (2009) 'A Flexible Biomedical Ontology Selection Tool', *International Journal of Computing and ICT Research, Special Issue*, 3(1), pp. 53–66.

Mandran, N. and Dupuy-Chessa, S. (2018) 'Supporting experimental methods in Information System research', in *12th IEEE International Conference on Research Challenges in Information Science RCIS'2018*. Nantes, France.

Martínez-Romero, M., Jonquet, C., O'Connor, M. J., Graybeal, J., Pazos, A. and Musen, M. A. (2017) 'NCBO Ontology Recommender 2.0: an Enhanced Approach for Biomedical Ontology Recommendation', *Journal of Biomedical Semantics. BioMed Central*. BioMed Central, 8(21). doi: 10.1186/s13326-017-0128-y.

Martínez-Romero, M., Vázquez-Naya, J. M., Pereira, J. and Pazos, A. (2014) 'BiOSS: A system for biomedical ontology selection', *Computer Methods and Programs in Biomedicine*. Elsevier Ireland Ltd, 114(1), pp. 125–40. doi: 10.1016/j.cmpb.2014.01.020.

Masolo, C., Borgo, S., Gangemi, A., Guarino, N. and Oltramari, A. (2002) *WonderWeb Deliverable D17*, *Computer Science Preprint Archive*.

Matentzoglu, N., Malone, J., Mungall, C. and Stevens, R. (2018) 'MIRO : guidelines for minimum information for the reporting of an ontology', *Journal of Biomedical Semantics*. Journal of Biomedical Semantics, 9(1), p. 6. doi: 10.1186/s13326-017-0172-7.

Mcdaniel, M., Storey, V. C. and Sugumaran, V. (2016) 'The Role of Community Acceptance in Assessing Ontology Quality', in *In International Conference on Applications of Natural Language to Information Systems*. Springer International Publishing Switzerland 2016, pp. 24–36. doi: 10.1007/978-3-319-41754-7_3.

McPeake, J., Bateson, M. and O'Neill, A. (2014) 'Electronic surveys: how to maximise success', *Nurse Researcher*, 21(3), pp. 24–26.

McQuilton, P., Gonzalez-Beltran, A., Rocca-Serra, P., Thurston, M., Lister, A., Maguire, E. and Sansone, S.-A. (2016) 'BioSharing: curated and crowd-sourced metadata standards, databases and data policies in the life sciences', *Database*, 2016, pp. 1–8. doi: 10.1093/database/baw075.

Miles, A. and Bechhofer, S. (2009) 'SKOS Simple Knowledge Organization System Reference'. Available at: https://www.w3.org/TR/skos-reference/.

Miles, M. B. and Huberman, A. M. (1994) *Qualitative Data Analysis: An Expanded Sourcebook*. Second Edi. SAGE Publications.

Miller, G. A. (1995) 'WordNet: a lexical database for English', *Communications of the ACM*. ACM, 38(11), pp. 39–41. doi: 10.1145/219717.219748.

Mizoguchi, R., Vanwelkenhuysen, J. and Ikeda, M. (1995) 'Task ontology for reuse of problem solving knowledge.', in *Towards Very Large Knowledge Bases*. Amsterdam: IOS Press Amsterdam, The Netherlands, pp. 46–59.

Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S. and Ferreira, M. E. C.

(2017) 'Scale development: ten main limitations and recommendations to improve future research practices', *Psicologia: Reflexão e Crítica*. Springer International Publishing, 30, p. 3. doi: 10.1186/s41155-016-0057-1.

Morris, M. W., Leung, K., Ames, D. and Lickel, B. (1999) 'Views from Inside and Outside: Integrating Emic and Etic Insights about Culture and Justice Judgment', *The Academy of Management Review*, 24(4), pp. 781–796. doi: 10.5465/amr.1999.2553253.

Mossakowski, T., Kutz, O. and Codescu, M. (2013) 'Ontohub-a repository engine for heterogeneous ontologies and alignments', in. Available at: http://ontolog.cim3.net/forum/ontoiop-forum/2013-05/pdfqyVdCx7UaU.pdf.

Mossakowski, T., Kutz, O. and Codescu, M. (2014) 'Ontohub A semantic repository for heterogeneous ontologies', in *Proceedings of the Theory Day in Computer Science (DACS-2014), satellite workshop of ICTAC-2014*. Bucharest.

Musen, M. A. and Protégé Team, the P. (2015) 'The Protégé Project: A Look Back and a Look Forward', *AI matters*. NIH Public Access, 1(4), pp. 4–12. doi: 10.1145/2757001.2757003.

Myers, M. D. (1997) 'Qualitative research in information systems', *Management Information Systems Quarterly*, 21(2), pp. 241–242. doi: 10.2307/249422.

Naskar, D. (2014) *Ontology and Ontology Libraries : A critical study*.

Netzer, Y., Gabay, D., Adler, M., Goldberg, Y. and Elhadad, M. (2009) 'Ontology Evaluation through Text Classification', in *Advances in Web and Network Technologies, and Information Management*. Suzhou, China: Springer, Berlin, Heidelberg, pp. 210–221. doi: 10.1007/978-3-642-03996-6_20.

Ning, H. and Shihan, D. (2006) 'Structure-Based Ontology Evaluation', in *2006 IEEE International Conference on eBusiness Engineering (ICEBE06)*. IEEE, pp. 132–137. doi: 10.1109/ICEBE.2006.97.

Noy, N. F., Shah, N. H., Whetzel, P. L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D. L., Storey, M.-A., Chute, C. G. and Musen, M. a (2009) 'BioPortal: ontologies and integrated data resources at the click of a mouse.', *Nucleic acids research*, 37(suppl_2), pp. W170–W173. doi: 10.1093/nar/gkp440.

Noy, N. F., Alexander, P. R., Harpaz, R., Whetzel, P. L., Fergerson, R. W. and Musen, M. A. (2013) 'Getting lucky in ontology search: A data-driven evaluation framework for ontology ranking', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8218 LNCS(PART 1), pp. 444–459. doi: 10.1007/978-3-642-41335-3_28.

Noy, N. F., Griffith, N. and Musen, M. (2008) 'Collecting community based mappings in an ontology repository', in *The International Semantic Web - ISWC 2008. Lecture Notes in Computer Science*.

Noy, N. F., Guha, R. and Musen, M. A. (2005) 'User Ratings of Ontologies : Who Will Rate the Raters ?', in *AAAI Symp. on Knowledge Collection from Volunteer Contributors*. Menlo Park: American Association for Artificial Intelligence, pp. 56--63.

Noy, N. F. and McGuinness, D. L. (2001) *Ontology Development 101: A Guide to Creating Your First Ontology*. Available at: http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html.

Noy, N. F., Musen, M. A. and Guha, R. (2005) ‘User Ratings of Ontologies : Who Will Rate the Raters ?’, *Proceedings of the AAAI Spring Symposium on Medical Informatics*, pp. 56--63.

Obrst, L., Ceusters, W., Mani, I., Ray, S. and Smith, B. (2007) *The Evaluation of Ontologies*, *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*. Edited by C. J. O. Baker and K.-H. Cheung. Springer. doi: 10.1007/978-0-387-48438-9_8.

Ochs, C., Perl, Y., Geller, J., Arabandi, S., Tudorache, T. and Musen, M. A. (2017) ‘An empirical analysis of ontology reuse in BioPortal’, *Journal of Biomedical Informatics*, 71, pp. 165–177. doi: 10.1016/j.jbi.2017.05.021.

Ong, E., Xiang, Z., Zhao, B., Liu, Y., Lin, Y., Zheng, J., Mungall, C., Courtot, M., Ruttenberg, A. and He, Y. (2016) ‘Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration.’, *Nucleic acids research*, 45(D1), pp. D347–D352. doi: 10.1093/nar/gkw918.

Ong, E., Xiang, Z., Zhao, B., Liu, Y., Lin, Y., Zheng, J., Mungall, C., Courtot, M., Ruttenberg, A. and He, Y. (2017) ‘Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration.’, *Nucleic acids research*, 45(D1), pp. D347–D352. doi: 10.1093/nar/gkw918.

Page, L., Brin, S., Motwani, R. and Winograd, T. (1998) *The PageRank Citation Ranking: Bringing Order to the Web*, *Technical report, Stanford Digital Library Technologies Project*. doi: 10.1.1.31.1768.

Pan, J. Z., Serafini, L. and Zhao, Y. (2006) ‘Semantic Import: An Approach for Partial Ontology Reuse’, in *Proceedings of the 1st International Workshop on Modular Ontologies (WoMO’06) co-located with the International Semantic Web Conference (ISWC’06)*, pp. 71–84.

Pan, J. Z., Thomas, E. and Sleeman, D. (2006) ‘OntoSearch2: Searching and Querying Web Ontologies’, in *Proceedings of the IADIS International Conference on WWW/Internet*. Murcia, Spain, pp. 211–219.

Park, H., Yoon, A. and Kwon, H.-C. (2012) ‘Task Model and Task Ontology for Intelligent Tourist Information Service’, *International Journal of u-and e-Service*, 5(2), pp. 43–58.

Patel, C., Supekar, K., Lee, Y. and Park, E. K. (2003) ‘OntoKhoj: A Semantic Web Portal for Ontology Searching, Ranking and Classification’, in *Proceedings of the 5th ACM international workshop on Web information and data management*. New Orleans, LA, USA: ACM, pp. 58–61.

Patton, M. Q. (1990) *Qualitative evaluation and research methods*. Second Edi. Thousand Oaks, CA, US: Sage Publications, Inc.

Patton, M. Q. (2015) *Qualitative research & evaluation methods:integrating theory and practice*. Fourth Edi. SAGE Publications.

Pinto, H. S. and Martins, J. P. (2001) ‘A Methodology for Ontology Integration’, in *Proceedings of the 1st international conference on Knowledge capture (K-CAP ’01)*. Victoria, British Columbia, Canada: ACM, pp. 131–138.

Porzel, R. and Malaka, R. (2004) ‘A Task-based Approach for Ontology Evaluation’, in *ECAI-2004 Workshop on Ontology Learning and Population*. Valencia, Spain, pp. 1–6.

Poveda-Villalón, M., Gómez-Pérez, A. and Suárez-Figueroa, M. C. (2014) 'Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation', *International Journal on Semantic Web and Information Systems (IJSWIS)*, 10(2), pp. 7–34.

Poveda-Villalón, M., Suárez-Figueroa, M. C. and Gómez-Pérez, A. (2012a) 'Did you validate your ontology? OOPS!', in *9th Extended Semantic Web Conference*. Heraklion, Crete: Springer, Berlin, Heidelberg, pp. 402–407.

Poveda-Villalón, M., Suárez-Figueroa, M. C. and Gómez-Pérez, A. (2012b) 'The Landscape of Ontology Reuse in Linked Data', in *1st Ontology Engineering in a Data-driven World (OEDW 2012) Workshop at the 18th International Conference on Knowledge Engineering and Knowledge Management (EKAW2012)*. Galway, Irlanda.

Pressman, R. S. (2005) *Software Engineering: A Practitioner's Approach*. 6th, illustr edn. McGraw-Hill, 2005.

QSR International (2015) 'NVivo'. Available at: https://www.qsrinternational.com/nvivo/home.

R Core Team (2013) 'R: A Language and Environment for Statistical Computing'. Vienna, Astria. Available at: http://www.r-project.org.

Raad, J. and Cruz, C. (2015) 'A Survey on Ontology Evaluation Methods', in *Proceedings of the International Conference on Knowledge Engineering and Ontology Development, part of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*. Lisbonne. doi: 10.5220/0005591001790186ï.

Ramesh, R. and Iyer, S. (2016) 'Integrating the Learning Objectives and Syllabus into a Domain Ontology for Data structures Course', in *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education*. Arequipa, Peru, pp. 266–271. doi: 10.1145/2899415.2899453.

Ricci, F., Rokach, L. and Shapira, B. (2011) *Introduction to recommender systems handbook, Recommender systems handbook*. Edited by F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor. Springer. doi: 10.1007/978-0-387-85820-3.

Ricci, F., Rokach, L., Shapira, B. and Kantor, P. (2010) *Recommender Systems Handbook*. Springer-Verlag Berlin, Heidelberg ©2010.

Rowley, J. and Slack, F. (2004) 'Conducting a literature review', *Management Research News*, 27(6), pp. 31–39. doi: 10.1108/01409170410784185.

Rudestam, K. E. and Newton, R. (2007) *Surviving Your Dissertation: A Comprehensive Guide to Content and Process*. Third Edit. SAGE Publications, Inc.

Rudestam, K. and Newton, R. (2014) *Surviving Your Dissertation: A Comprehensive Guide to Content and Process*. Fourth Edi. SAGE Publications, Inc.

Sabou, M., Lopez, V. and Motta, E. (2006) 'Ontology Selection for the Real Semantic Web: How to Cover the Queen's Birthday Dinner?', in Vrandecic, D., Suarez-Figueroa, M. C., Gangemi, A., and Sure, Y. (eds) *Proceedings of the 15th International Conference on Knowledge Engineering and Knowledge Management Managing Knowledge in a World of Networks*. Edinburgh, Scotland, UK: Springer, Berlin, Heidelberg (CEUR Workshop Proceedings), pp. 96–111. doi: 10.1007/11891451_12.

Sabou, M., Lopez, V., Motta, E. and Uren, V. (2006) 'Ontology selection: ontology evaluation on the real Semantic Web', in *15th International World Wide Web Conference (WWW 2006)*. Edinburgh, UK.

Saunders, M., Lewis, P. and Thornhill, A. (2004) *Research Methods for Business Students*. Third Edit. Harlow : Prentice Hall Financial Times.

Saunders, M., Lewis, P. and Thornhill, A. (2009) *Research methods for business students*. Fifth Edit. Pearson Education.

Saunders, M., Lewis, P. and Thornhill, A. (2012) *Research Methods for Business Students*. Sixth Edit. Pearson.

Schmidt, D., Basso, R., Trojahn, C. and Vieira, R. (2018) 'Matching Domain and Top-level ontologies exploring Word Sense Disambiguation and Word Embedding', in Demidova, E., Zaveri, A. J., and Simperl, E. (eds) *Volume 36: Emerging Topics in Semantic Technologies*. IOS Press Ebooks, pp. 27–38. doi: 10.3233/978-1-61499-894-5-27.

Schober, D., Tudose, I., Svatek, V. and Boeker, M. (2012) 'OntoCheck: verifying ontology naming conventions and metadata completeness in Protégé 4', *Journal of Biomedical Semantics*, 3(Suppl 2), p. S4. doi: 10.1186/2041-1480-3-S2-S4.

Seiler, M. J. (2004) *Performing Financial Studies: a Methodological Cookbook*, *The British Accounting Review*. New Jersey, USA: Prentice Hall.

Shadbolt, N., Berners-Lee, T. and Hall, W. (2006) 'The semantic web revisited', *IEEE intelligent systems*, 21(3), pp. 96–101.

Sharp, M. E. (2017) 'Toward a comprehensive drug ontology: extraction of drug-indication relations from diverse information sources', *Journal of Biomedical Semantics*. BioMed Central, 8(1), p. 2. doi: 10.1186/s13326-016-0110-0.

Simperl, E. (2009) 'Reusing ontologies on the Semantic Web: A feasibility study', *Data & Knowledge Engineering*. Elsevier B.V., 68(10), pp. 905–925. doi: 10.1016/j.datak.2009.02.002.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L. J., Eilbeck, K., Ireland, A., Mungall, C. J., Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R. H., Shah, N., Whetzel, P. L. and Lewis, S. (2007) 'The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration', *Nature Biotechnology*, 25(11), pp. 1251–1255. doi: 10.1038/nbt1346.

Smith, B. (2008) 'The Evaluation of Ontologies : Editorial Review vs . Democratic Ranking', in *Proceedings of InterOntology 2008*. Tokyo, Japan, pp. 127–138.

Sowa, J. F. (2000) 'Ontology, Metadata and Semiotics', in *International conference on conceptual structures (ICCS)*, pp. 55–81.

Sridevi, K. and Umarani, R. (2013) 'Ontology Ranking Algorithms on Semantic Web: A Review', *International Journal of Advanced Research in Computer and Communication Engineering*, 2(9), p. 17.

Stamper, R., Liu, K., Hafkamp, M. and Ades, Y. (2000) 'Understanding the roles of signs and norms in organizations - a semiotic approach to information systems design', *BEHAVIOUR & INFORMATION TECHNOLOGY*, 19(1), pp. 15–27. doi: 10.1080/014492900118768.

Stevens, J. (2009) *Applied multivariate statistics for the social sciences*. Fifth Edit.

Routledge.

Suárez-Figueroa, M. C. and Gómez-Pérez, A. (2008) 'First Attempt towards a Standard Glossary of Ontology Engineering Terminology', in *Proceedings of 8th International Conference on Terminology and Knowledge Engineering (TKE'08)*. Copenhagen. doi: ISBN: 87-91242-50-9.

Suárez-Figueroa, Mari Carmen, Gómez-Pérez, Asunción and Fernández-López, M. (2012) 'The NeOn Methodology for Ontology Engineering', in Suárez-Figueroa, M C et al. (eds) *Ontology Engineering in a Networked World*. Springer Berlin Heidelberg, pp. 9–34. doi: 10.1007/978-3-642-24794-1_2.

Subhashini, R. and Akilandeswari, J. (2011) 'A Survey on Ontology Concstruction Methodologies', *International Journal of Enterprise Computing and Business Systems*, 1(1), pp. 60–72.

Subhashini, R., Lecturer, S., Akilandeswari, J. and Sinthuja, V. (2011) 'A Review on Ontology Ranking Algorithms', *International Journal of Computer Applications*, 33(4), pp. 975–8887.

Suhr, D. D. (2006) 'Exploratory or confirmatory factor analysis?', in *Proceedings of the Thirty-first Annual SAS Users Group International Conference*. San Francisco, California.

Supekar, K. (2005) 'A peer-review approach for ontology evaluation', in *8th International Protégé Conference*. Madrid, Spain, pp. 77–79. doi: citeulike-article-id:300085.

Supekar, K., Patel, C. and Lee, Y. (2004) 'Characterizing Quality of Knowledge on Semantic Web', in *Proceedings of 7th International Florida Artificial Intelligence Research Society Conference (FLAIRS'02)*. Miami Beach, Florida, pp. 472–478. doi: 10.1.1.100.8050.

Sure, Y., Angele, J. and Staab, S. (2002) 'OntoEdit: Guiding Ontology Development by Methodology and Inferencing', in Meersman, R. and Tari, Z. (eds) *On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE. OTM 2002. Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, pp. 1205–1222. doi: 10.1007/3-540-36124-3_76.

Symon, G. and Cassell, C. (2012) *Qualitative organizational research : core methods and current challenges*. SAGE.

Talebpour, M., Sykora, M. and Jackson, T. W. (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*. Funchal, Madeira, Portugal, pp. 119–127. doi: 10.5220/0006589201190127.

Tan, H. and Lambrix, P. (2009) 'Selecting an Ontology for Biomedical Text Mining', in *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing (BioNLP)*. Boulder, CO, USA: Association for Computational Linguistics, pp. 55–62.

Tartir, S., Arpinar, I. . and Sheth, A. . (2010) 'Ontological evaluation and validation', in Poli, R., Healy, M., and Kameas, A. (eds) *Theory and Applications of Ontology: Computer Applications*. New York: Springer., pp. 115–130.

Tartir, S. and Arpinar, I. B. (2007) 'Ontology Evaluation and Ranking using OntoQA', in *Proceedings of the First IEEE International Conference on Semantic Computing (ICSC 2007)*. Irvine, CA, USA: IEEE Computer Society 2007, pp. 185–192. doi:

10.1109/ICSC.2007.19.

Tartir, S., Arpinar, I., Moore, M., Sheth, A. and Aleman-Meza, B. (2005) 'OntoQA: Metric-Based Ontology Quality Analysis', in *Proceedings of the IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources (ICDM'05)*. New Orleans, Louisiana, USA, pp. 45–53. doi: 10.1.1.131.2087.

Tashakkori, A. and Creswell, J. W. (2007) 'Exploring the Nature of Research Questions in Mixed Methods Research', *Journal of Mixed Methods Research*, 1(3). doi: 10.1177/1558689807302814.

Tashakkori, A. and Teddlie, C. (eds) (2003) *Handbook of Mixed Methods in Social & Behavioral Research*. SAGE Publications.

Teddlie, C. and Tashakkori, A. (2009) *Foundations of Mixed Methods Research Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. Edited by A. Tashakkori and C. Teddlie. SAGE Publications, Inc.

Teddlie, C. and Yu, F. (2007) 'Mixed Methods Sampling A Typology With Examples', *Journal of mixed methods research*, 1(1), pp. 77–100. doi: 10.1177/1558689819844638.

The Gene Ontology Consortium (2004) 'The Gene Ontology (GO) database and informatics resource', *Nucleic Acids Research*, 32(suppl_1), pp. D258–D261. doi: 10.1093/nar/gkh036.

Thompson, B. (2004) *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* Washington: American Psychological Association. doi: 10.1037/10694-000.

Trotter, R. T. (2012) 'Qualitative research sample design and sample size: Resolving and unresolved issues and inferential imperatives', *Preventive Medicine*, 55(5), pp. 398–400. doi: 10.1016/j.ypmed.2012.07.003.

Tudorache, T. and Noy, N. (2007) 'Collaborative Protégé', in *Proceedings of the Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007) at the 16th International World Wide Web Conference (WWW2007)*. Banff, Canada.

Tummarello, G., Delbru, R. and Oren, E. (2007) 'Sindice.com: Weaving the Open Linked Data', in *The Semantic Web*. Busan, Korea: Springer, Berlin, Heidelberg, pp. 552–565. doi: 10.1007/978-3-540-76298-0_40.

Turner III, D. W. (2010) 'Qualitative Interview Design: A Practical Guide for Novice Investigators', *The qualitative report*, 15(3), pp. 754–760.

Ungrangsi, R., Anutariya, C. and Wuwongse, V. (2007) 'combiSQORE: An Ontology Combination Algorithm', in *The Semantic Web*. Springer, Berlin, Heidelberg., pp. 566–579.

Ungrangsi, R., Anutariya, C. and Wuwongse, V. (2008) 'combiSQORE: A Combinative-Ontology Retrieval System for Next Generation Semantic Web Applications', *IEICE Transactions on Information and Systems*, 91(11), pp. 2616–2625. doi: 10.1093/ietisy/e91-d.11.2616.

Uschold, M. (1998) 'Knowledge level modelling: concepts and terminology', *The Knowledge Engineering Review*, 13(1), pp. 5–29.

Uschold, M., Clark, P., Healy, M., Williamson, K. and Woods, S. (1998) 'An Experiment in

Ontology Reuse', in *Proceedings of the 11th Knowledge Acquisition Workshop KAW98*. Banff.

Uschold, M., Healy, M., Williamson, K., Clark, P. and Woods, S. (1998) 'Ontology reuse and application', in Guarino, N. (ed.) *Proceedings of the first International Conference on Formal Ontology in Information Systems (FOIS'98)*. IOS Press, pp. 179–192.

Uschold, M. and King, M. (1995) 'Towards a Methodology for Building Ontologies', in *Proceedings of IJCAI95's Workshop on Basic Ontological Issues in Knowledge Sharing*.

Vandenbussche, P.-Y., Atemezing, G. A., Poveda-Villalon, M. and Vatant, B. (2017) 'Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web', *Semantic Web*. IOS Press, 8(3), pp. 437–452.

Vargha, A. and Delaney, H. D. (1998) 'The Kruskal-Wallis Test and Stochastic Homogeneity', *Journal of Educational and Behavioral Statistics*, 23(2), pp. 170–192. doi: 10.2307/1165320.

Visser, P. R. S. and Bench-Capon, T. J. M. (1998) 'A Comparison of Four Ontologies for the Design of Legal Knowledge Systems', *Artificial Intelligence and Law*, 6(1), pp. 27–57.

Völker, J. *et al.* (2008) 'AEON–An approach to the automatic evaluation of ontologies', *Applied Ontology*, 3, pp. 41–62. doi: 10.3233/AO-2008-0048.

Wang, X., Guo, L. and Fang, J. (2008) 'Automated ontology selection based on description logic', in *Proceedings of the 2008 12th International Conference on Computer Supported Cooperative Work in Design, CSCWD (2008)*. IEEE, pp. 482–487. doi: 10.1109/CSCWD.2008.4537026.

Xiang, Z., Courtot, M., Brinkman, R. R., Ruttenberg, A. and He, Y. (2010) 'OntoFox: web-based support for ontology reuse', *BMC Research Notes*, 3, p. 175. doi: 10.1186/1756-0500-3-175.

Xiang, Z., Mungall, C., Ruttenberg, A. and He, Y. (2012) 'Ontobee: A Linked Data Server that publishes RDF and HTML data simultaneously', *Semantic Web Journal*.

Yan, T. and Curtin, R. (2010) 'The Relation Between Unit Nonresponse and Item Nonresponse: A Response Continuum Perspective', *International Journal of Public Opinion Research*. Narnia, 22(4), pp. 535–551. doi: 10.1093/ijpor/edq037.

Yang, Z., Zhang, D. and Ye, C. (2006) 'Evaluation metrics for ontology complexity and evolution analysis', *IEEE International Conference on e-Business Engineering (ICEBE'06)*, pp. 162–170. doi: 10.1109/ICEBE.2006.48.

Yong, A. and Pearce, S. (2013) 'A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis', *Tutorials in Quantitative Methods for Psychology*, 9(2), pp. 79–94. doi: 10.20982/tqmp.09.2.p079.

Yu, J., Thom, J. A. and Tam, A. (2009) 'Requirements-oriented methodology for evaluating ontologies', *Information Systems*. Elsevier, 34(8), pp. 686–711. doi: 10.1016/j.is.2009.04.002.

Yu, J., Thom, J. a and Tam, A. (2007) 'Ontology evaluation using wikipedia categories for browsing', in *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management (CIKM 07)*, pp. 223–232. doi: 10.1145/1321440.1321474.

Zhang, Y., Vasconcelos, W. and Sleeman, D. (2004) 'OntoSearch: An Ontology Search Engine', in Bramer, M., Coenen, F., and Allen, T. (eds) *24th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, AI'04*. Springer, pp. 58–69.

Zulkarnain, N., Meziane, F. and Crofts, G. (2016) 'A methodology for biomedical ontology reuse', in *International Conference on Applications of Natural Language to Information Systems*. Springer, Cham, pp. 3–14. doi: 10.1007/978-3-319-41754-7_1.

Zuniga, G. L. (2001) 'Ontology: Its Transformation From Philosophy to Information Systems', in *Proceedings of the international conference on Formal Ontology in Information Systems*. ACM, pp. 187–197.

Appendix A: Selection Approaches Review

This section aims to review some of the selection approaches for ontologies (Table A-1). As it is seen in the literature, ontology selection is an umbrella term used to refer to the all the systems that facilitate the process of selecting an ontology for reuse, including, search engines for ontologies, ontology evaluation approaches, ontology ranking approaches and ontology recommendation approaches. There are a number of similarities between these approaches, namely, how they collect ontologies and form their repositories and how they evaluate ontologies.

Table A-1 Selection Approaches

Tool or Method	**Recommendation scenario**	**Domain**	**Web service or UI page**	**Year**
OntoSelect	Metadata	General	http://views.dfki.de/ontologies/ (dead link)	2004
Hong et al.	Keywords	General	Not Available	2005
Sabou et al.	keywords	General	Not Available	2006
Wang et al. (DL-AOSF)	Keywords	General	Not Available	2008
Tan & Lambrix	Text	Biomedical	Not Available	2009

OntoSelect. This approach was based on a dynamic crawling procedure that would monitor the web to find the newly published ontologies in RDF/S, DAML or OWL format (Buitelaar, 2004; Buitelaar, Eigner and Declerck, 2004). Ontologies in OntoSelect were stored and organised according to different characteristics, such as their format, name, and language. OntoSelect also supported semi-automatic ontology selection by using different criteria, namely, coverage, structure, and connectedness. These metrics are very similar to those used by other search and recommender systems for ontologies, namely, NCBO BioPortal.

Hong, Chang, & Lin (2005). According to Hong, Chang, & Lin (2005), the ontology selection process consists of two main phases: requirement analysis and ontology selection. Two sub levels were also suggested for each of the phases in this process, including coarse-level and fine-level. Coarse-level requirement analysis aimed to elicit the selection requirements and retrieve what they called a “fixed number of promising candidate source ontologies”. The fine level requirement analysis helped in the process of finding the "best matched ontology" by defining details and clarifying the requirements identified in the previous steps. In the selection phase, the requirements identified in the two levels of the first phase would be used to find the desired ontology. Like many of the previous approaches, ontology selection in this approach is a manual task.

Sabou, Lopez and Motta (2006). In this study, a set of selection requirements were identified, and an algorithm was proposed to support the process of ontology selection. This selection algorithm is very similar to the ones previously proposed in the literature; it starts from query expansion and ends by identifying a combination of ontologies that best covers the input requirements and that are ranked based on their level of abstraction. What makes this algorithm a bit different is the fact that the selection process can happen in different stages, in case one of them fails to identify an ideal ontology or a combination of them.

DL-AOSF. As an automated framework for ontology selection, DL-AOSF consisted of five different components, namely, need acquirement, ontology library, automated selection criteria, output interface, and parameter console (Wang, Guo and Fang, 2008). Queries submitted to this system were first converted to logic query and then compared with the ontologies in the library of this system. Two different metrics, namely, topic coverage and knowledge richness were then used to select and retrieve the potential ontologies. Lastly, the only non-automated component of this framework would allow users to set different console parameters and to control the output.

Tan and Lambrix (2009). This framework aimed to help the process of selecting what they call the "most appropriate" ontology for a text mining application and consisted of three main components, each of which was responsible for addressing different selection requirements. The first component, for example, aimed to retrieve the initial set of ontologies by analysing their content and supporting technologies. The second component would verify and evaluate different aspects of the previously identified ontologies. Finally, in the third component, ontologies were used in an application to test how well they satisfy the selection requirements.

Appendix B: Review of Libraries and Repositories for Ontologies

Ontology repositories aim to collect, manage, publish and provide access to ontologies from different resources (Naskar, 2014). A list of some of the well-known repositories for ontologies is presented in Table B-1. As it is seen in the following part of this section, ontology repositories are different from each other in a number of important ways, namely, the domain they cover, how they collect the ontologies in their repositories and the type of metadata that they keep about each ontology.

Table B-1 Libraries and Repositories for Ontologies

Tool or Method	**Recommendation Scenario**	**Domain**	**Web service or UI page**	**Year**
EBI Ontology Lookup Service	Keywords	Biomedical	https://www.ebi.ac.uk/ols/index	2005
OBO Foundry	Metadata	Biomedical	http://www.obofoundry.org/	2007
BioPortal	Term & Metadata	Biomedical	http://bioportal.bioontology.org/ontologies	2009
Ontobee	Keywords	Biomedical	http://www.ontobee.org/	2011
OKFN Linked Open Vocabularies	Keywords	General	http://lov.okfn.org/	2011
Fairsharing (previously known as BioSharing)	Metadata	Life Sciences	https://biosharing.org/	2011
Ontohub	Metadata	General	https://ontohub.org/	2014
AgroPortal	Term and Metadata	Agronomic Data	http://agroportal.lirmm.fr/ontologies	2016

EBI Ontology Lookup Service. OLS was first introduced in mid-2005 as a user-friendly access point for publicly available biomedical ontologies; it aimed to address one of the main challenges in ontology domain, which is the availability of different ontologies that are scattered all over the web. To address this issues, OLS integrates publicly available biomedical ontologies into a single repository or a "point of query" (Côté *et al.*, 2006). OLS interface offers keyword(s) based search using a "suggest-as-you-type" form. The results of this system include a set of ontologies that can be filtered using different metadata. Dynamically generated tree structure and term history are also available for each of the ontologies in this system.

OBO Foundry. The Open Biological and Biomedical Ontology (OBO) Foundry is not only a very popular library for ontologies, but it also provides models of good ontology development practises. One of the notable differences between OBO Foundry and the other proposed repositories is that it only accepts ontologies that meet a specific set of criteria, including, being open and available, being expressed in a common shared syntax, and being well documented (Smith *et al.*, 2007). Moreover, each of the ontologies in this repository is evaluated by two different groups of editors, namely, coordinating editors and associate editors, whose primary responsibility is to make sure that ontologies meet the requirements mentioned above (Smith, 2008).

OBO Foundry, as a very well known example of closed rating systems, plays a significant role in the ontology domain; however, some have challenged the peer review method used by this system. Noy, Musen, & Guha (2005), for example, argued that reviewers in closed ranking systems like OBO Foundry only deal with a few sets of ontologies; therefore, they are only able or capable of providing reviews for those very few ontologies that they have previously used.

BioPortal & AgroPortal browse. The browse function of both NCBO BioPortal[49] and IBC AgroPortal[50] helps the users to navigate through the list of ontologies that are available in these libraries, both through their web interfaces and web services that they provide (Martínez-Romero *et al.*, 2017). In these systems, ontologies can be explored using different metadata, namely, their format (e.g. OBO, OWL, UMLS), the natural language they have been built in, their type (e.g. core ontology, domain ontology), and the group they belong to. Unlike OBO Foundry, no editorial process is used in these two repositories and therefore, every user is allowed to submit their ontologies.

Ontobee. Ontobee has been defined as a linked ontology data server that aims to facilitate the process of browsing biomedical ontologies. It stores different types of information and metadata about the ontologies retrieved from OBO Foundry repository and saves them in two different databases: RDF triple store, and MySQL database (Xiang *et al.*, 2012). Ontobee

[49] https://bioportal.bioontology.org/ontologies
[50] http://agroportal.lirmm.fr/

architecture is formed of three different tiers, including presentation tier, logic tier, and data tier. It also supports a wide range of activities, including ontology browsing, ontology query, and keyword-based search, ontology visualization, and linkage of ontology terms.

Ontobee is similar to many of the repositories reviewed in this section, as it stores biomedical ontologies. However, its developers claim that their system is different from other popular repositories for biomedical ontologies, namely, BioPortal, AberOWL, and OLS in a number of respects, namely, because it is the only system that is able to dynamically dereference and present individual ontology term URIs in different formats, e.g. HTML web pages for users or RDF source code for semantic web applications (Ong *et al.*, 2017).

OKFN Linked Open Vocabularies (LOV). LOV, as a catalogue of reusable vocabularies, aims to provide access to available and online linked data and to help data publishers in finding vocabularies that can be used to describe their data. LOV architecture is composed of four main components: Tracking and Analysis, Curation, Data Access, and user interface and application program interface (Vandenbussche *et al.*, 2017). Different activities, namely, keyword-based ontology search, ontology browsing, ontology assessment and ontology mapping, are supported by this system.

LOV is different from the other reviewed repositories in this section in several respects, namely, because it is not limited to a specific domain, e.g., biomedical. LOV interface is also very interesting and looks very different, compared to other repositories for ontologies. Like OBO Foundry, adding vocabularies to LOV is not automated, and each submitted vocabulary is first evaluated by a group of curators to make sure it meets what they call "LOV quality requirements" (Vandenbussche *et al.*, 2017).

BioSharing/FAIRsharing.org. FairSharing[51], previously known as BioSharing, has been identified as a community driven portal for different types of registries in the life sciences, namely, standards, databases and data policies. FAIRsharing is more concerned with linked information and data management rather than ontology management (McQuilton *et al.*, 2016). However, it provides its users with some functionalities that are very similar to what ontology repositories offer, namely, search and metadata-based browsing.

[51] https://fairsharing.org/

Ontohub. Ontohub[52] has been described as a "repository engine" that helps in managing heterogeneous distributed ontologies (Mossakowski, Kutz and Codescu, 2014). Some of the distinct features of this system include the availability of multiple repositories, Git interface, and modular architecture. Ontohub developers argue that this system is not only a library for ontologies, but it is also a collection of ontology languages, their underlying logic and their translation (Mossakowski, Kutz and Codescu, 2013). Ontohub is comparable to other repositories, as it supports different functions, such as ontology search and browsing using different metadata, including ontology type, project, formality level, license model and task.

[52] https://ontohub.org/

Appendix C: Review of Search Engines for Ontologies

Search engines, as one of the most popular type of selection systems in ontology domain, aim to facilitate the process of ontology exploration and retrieval (Naskar, 2014) by finding an ontology, a module in an ontology (Suárez-Figueroa and Gómez-Pérez, 2008) or a set of ontologies (Martínez-Romero *et al.*, 2017) that are the most relevant to users' queries (Alani *et al.*, 2007). A list of some of the most popular search engines for ontologies are presented in Table C–1. Despite all the similarities, ontology search engines can be significantly different from each other in the way they collect ontologies and also how they evaluate and rank ontologies.

Table C-1 Search Engines for Ontologies

Tool or Method	Recommendation scenario	Domain	Web service or UI page	Year
OntoKhoj	Keywords	General	http://sice527.ddns.umkc.edu/ontokhoj (not Available)	2003
Swoogle	Keywords	General	http://swoogle.umbc.edu	2004
OntoSearch	Keywords	General	http://www.ontosearch.com	2005
Alani et al.	Terms	General	Not Available	OCT 2007
Sindice	Keyword, URI, ifp	General	http://sindice.com (Not Available)	NOV 2007
NCBO BioPortal	Keywords	Biomedical	https://bioportal.bioontology.org/search	DEC 2007
Watson	Keywords	General	http://watson.kmi.open.ac.uk/WatsonWUI/	July 2008
AberOWL	Term, Phrase or Description Logic query	Biomedical	http://aber-owl.net	2015

OntoKhoj. As a Semantic web portal, OntoKhoj aimed to simplify ontology engineering process by providing a context-oriented query interface for ontologies. The key process in this system consisted of crawling the web for ontologies, classifying ontologies, ranking them, and providing ontology search and visualisation facilities. OntoKhoj differs from the other search engine for ontologies in the way it pre-processed its input; this system used WordNet (Miller,

1995) to provide a list of different senses associated with each input, that was a keyword and allowed its users to select the right sense. A set of synonyms and hypernyms that were associated with the selected sense of the input would then be retrieved.

OntoKhoj did not only try to facilitate the process of searching for ontologies, but it could also be considered as one of the very first systems to talk about the issue of trust on the semantic web. Developers of this system argued that there are no restrictions on the online information and human or the machines and agents that work on behalf of them are responsible for evaluating the validity, quality and trustworthiness of information on the web (Patel *et al.*, 2003). To tackle this problem, they proposed and developed an algorithm called OntoRank; this algorithm was very similar to PageRank (Page *et al.*, 1999) and would rank ontologies based on the number of other ontologies that refer to them.

Swoogle. Swoogle's architecture consists of different components, namely, discovery component, metadata creation component, data analysis component and interface component (Finin *et al.*, 2005). Discovery component is responsible for crawling the web, identifying online and accessible semantic web documents (SWD), and forming a collection of ontologies. The metadata creation component helps the search process by collecting different types of metadata, namely, basic, relations, and analytical about each semantic web document (Ding *et al.*, 2004). Other components include data analysis that is responsible for SWD ranking and classification and also the interface component that provides different types of data services to the semantic web community. Despite all the facilities it offers, Swoogle has not been actively updated for many years; therefore, it might not be able to address many of the recent challenges in the ontology domain, especially when it comes to finding a reusable ontology.

OntoSearch2. According to Pan, Thomas and Sleeman (2006), OntoSearch2 is a query-based search engine for ontologies that not only allows its users to submit ontologies to it, but it also allows them to query its repository using a restricted subset of SPARQL. OntoSearch2 is based on its predecessor version, ONTOSEARCH (Zhang, Vasconcelos and Sleeman, 2004), that would get keywords as input, use google to perform the search and return RDF files in the results. There is currently no live version of this search engine available; the last available and live version of OntoSearch2 would use query autocomplete to populate a set of options for each keyword in the input and would then return three different types of output, namely, HTML, RDF and Graph. However, it did not seem to be based on the architecture proposed by (Pan, Thomas and Sleeman, 2006) or (Zhang, Vasconcelos and Sleeman, 2004).

Alani et al. Besides all the different search engines and frameworks for ontologies, there are some other approaches proposed in the literature that have never been developed and implemented, but worth mentioning and exploring; the approach proposed by Alani et al. (2007) is one of them. To address the challenge of finding a reusable ontology, these authors monitored how users tend to search for ontologies; they found out that users usually tend to use the domain name as the query term(s). Therefore, to enhance the search experience, they proposed a novel query expansion technique that could be used in the process of searching for ontologies. In this approach, query term(s) entered by users were considered as a domain name and were extended by finding different web pages that were relevant to that particular domain using Google search or Wikipedia pages. It would help in identifying the top 50 related terms that could be used as the query input.

Sindice. Sindice was proposed as a lookup service for semantic web documents and resources and aimed to facilitate the process of locating data sources as well as integrating them. This system worked by collecting RDF documents and indexing them by using resource URIs, keywords, and Inverse Functional Properties (IFPs). The user interface (which is not currently available) would allow human or semantic web agents to look for and find an indexed resource. Sindice has been blamed for not allowing its users to exploit the located resources; it would make them download and process the resources locally (d'Aquin *et al.*, 2008).

BioPortal & AgroPortal search. Search is one of the very widely used services (Martínez-Romero *et al.*, 2017) offered by both NCBO BioPortal[53] and IBC AgroPortal[54]. These systems allow their users to search for different components of an ontology, for an entire ontology or across multiple vocabularies and ontologies (Jonquet *et al.*, 2016), using both their websites and web services. The output of these systems includes a list of ontologies that match the users' requirements. The ontologies in the output are ranked by different criteria, namely, ontology acceptance (Martínez-Romero *et al.*, 2017). Despite being very advanced and popular, these two search engines are limited to two very specific domains and therefore, cannot be used to find and select general ontologies.

Watson. Watson was developed in 2007 and aimed to work as a gateway to the online semantic information and to support application developers in exploiting what they called a large amount of heterogeneous distributed data (d'Aquin *et al.*, 2007). Like Swoogle, Watson architecture

[53] https://bioportal.bioontology.org/search
[54] http://agroportal.lirmm.fr/search

consisted of different components that were responsible for collecting, analysing and indexing ontologies (d'Aquin *et al.*, 2008). A web interface was also provided to facilitate ontology search; however, it is not available anymore.

AberOWL. AberOWL is a library and also a semantic search engine for biomedical ontologies. This search engine accepts term(s), phrase(s) or a description logic query. Its output can either be different classes of an ontology or a set of ontologies that have the search query terms as a part of their class description. AberOWL developers have compared their system with some of the other popular repositories and search engines for biomedical ontologies, such as BioPortal and OLS and claim that it differs from those systems in a number of different ways, namely, because it provides different reasoning infrastructure and reasoning services for ontologies (Hoehndorf *et al.*, 2015).

Appendix D: Evaluation Approaches Review

Ontology evaluation is a fundamental property of ontology domain, as the process of selecting an ontology for reuse highly depends on it. A list of metric-based ontology evaluation approaches is presented in Table D-1; this section provides a review of them.

Table D-1 Evaluation Approaches

Tool or Method	Recommendation scenario	Domain	Web service or UI page	Year
(ONTO)2Agent	Query term(s)	General	Not Available	2000
ONTOMETRIC	Metric-based evaluation	General	Not Available	2004
Supekar	Metadata	General	Not Available	2005
Lewen et al	Not applicable	General	Not Available	2006
Maiga & Williams	Task selection	Biomedical	Not Available	2009

(ONTO)^{2}Agent. One of the very early and also interesting ontology evaluation approaches was proposed by Arpírez, Gómez-Pérez, Lozano-Tello, & Pinto (2000) to address what they claimed to be an significant problem of ontology reuse in their time, which was the lack of standard features that could be used to describe and characterise ontologies from the users' point of view. To address this issue, they reviewed a set of ontologies available on the web and identified three different sets of features that could be used to characterise them; those sets included: (1) features that provide some information about the ontology and its developers, (2) features that aims to describe the form and content of an ontology, and (3) features that are mostly about how an ontology functions and how it can be used in an application.

Moreover, to support ontology search and selection, (ONTO)^{2}Agent was proposed and was described as a broker for the ontology domain that could help in searching and retrieving a set of ontologies that totally or partially meet and satisfy a set of constraints and requirements. One of the major drawbacks of this approach was that while aiming to "characterize the ontologies from the user point of view" (Arpírez *et al.*, 1998), no user was asked what they thought the main characteristics of a reusable ontology were.

ONTOMETRIC. As is one of the most popular metric-based evaluation methods in the literature, ONTOMETRIC proposed 160 different metrics that could be used to evaluate five different dimensions of ontologies, namely, their content, language, methodology, tools and costs (Lozano-Tello, 2002). ONTOMETRIC is believed to be helpful in the process of selecting the most appropriate ontology among various alternatives and also while making a decision about the suitability of a particular ontology for a project. However, applying metrics identified in this approach is very time consuming and will require a team of analysts (Lozano-Tello and Gómez-Pérez, 2004).

Supekar, K. As it is seen in the literature, very few studies have investigated the role of qualitative rankings and reviews in the ontology evaluation process. In 2005, Supekar proposed a new evaluation method which would allow users to provide qualitative reviews and ratings on different aspects of ontologies (Supekar, 2005). Ontology users were also asked to provide two different sets of metadata, namely, source metadata and third-party metadata for each ontology. It was argued that these sets of metadata are helpful in capturing quality features of ontologies. Despite its uniqueness and significant features, this method has not been used or implemented in any of the selection systems for ontologies.

Lewen, Supekar, Noy, & Musen (2006). In 2006, Lewen et al. proposed what they called an open rating system for ontology evaluation. In this approach, users were not only allowed to provide reviews for different ontologies, but they could also rate other users' reviews. This research was also one of the few examples of using the notion of trust in the ontology domain; it was argued that users might trust each other's reviews differently based on the topic or the area of the review.

This approach was implemented by developing what they referred to as "Knowledge Zone", which was an environment that would allow ontology developers and users to submit ontologies to a repository as well as annotating them with different metadata and reviews. To the best of the researcher's knowledge, the implemented version of this approach is not available on the web. Moreover, the open rating approach used by Lewen et al. (2006) has been challenged and criticized by the developers of OBO Foundry (Smith, 2008).

Maiga & Williams (2009) reviewed and identified different evaluation requirements for ontology selection in the biomedical field and designed a tool that was able to cover those requirements. Their proposed tool was based on three main activities, namely, ontology summarization, task determination and matching and update. The matching components of this tool would get a summary of ontologies and users' selection requirements as input and would try to make a match between the inputs and provide and recommend an ontology in the output. Each ontology in this tool would be evaluated using different metrics, such as granular density, scope, and biomedical ontology structure integration.

Appendix E: Ranking Approaches Review

There are a number of similarities between different ranking algorithms proposed in the literature. Firstly, there is no live version or a developed system available for any of them. Secondly, none of them would work as a stand-alone system, and they all need a search engine like Swoogle to retrieve a set of ontologies that would best match their users' queries and requirements. Some of the well-known ranking algorithms are reviewed in this section; as it is seen in Table E-1, CombiSQORE is the only approach based on the semantic query.

Table E-1 Ranking Approaches

Tool or Method	Recommendation scenario	Domain	Web service or UI page	Year
AKTiveRank	Keyword	General	Not Available	2006
OntoQA	Keyword	General	Not Available	2007
CombiSQORE	Semantic query	General	Not Available	2008

AKTiveRank. As an experimental system, AKTiveRank aimed to assess ontologies based on how well they would represent users' requirements or different concepts of interest (Alani, Brewster and Shadbolt, 2006). This system consisted of different components, namely, a Java Servlet, that was responsible for dealing with different HTTP queries from users or agents. Like most of the other ranking algorithms, AKTiveRank did not function as a stand-alone system and relied on Swoogle for the process of ontology search and retrieval. AKTiveRank would use four different criteria, namely, Class Match Measure, Density Measure, Semantic Similarity Measure, and Betweenness Measure to assess different representational aspects of each of the retrieved ontology and calculate their ranking and relevance to the users' queries (Alani, Brewster and Shadbolt, 2006).

OntoQA. As an evaluation and ranking system for ontologies, OntoQA was initially developed in 2005 and then enhanced in 2007. According to Tartir & Arpinar (2007), OntoQA supports three different types of selection scenarios and input: (1) An ontology, (2) An ontology and keywords, and (3) Keywords. Metrics related to two different dimensions of ontologies, namely, Schema dimension and Instances Dimension are defined and also used to evaluate, rank and calculate an overall score for each ontology. Similar to AKTiveRank, OntoQA uses Swoogle to retrieve a set of relevant ontologies for each input query.

CombiSQORE. Ungrangsi, Anutariya, & Wuwongse (2008) introduced combiSQORE as an ontology retrieval system that could cover users' requirements by finding and integrating a set of ontologies that fulfil those requirements. This system was based on five different components, namely, a semantic query, combiSQORE retrieval engine, an ontology database, a semantic lexical database and semantic web gateways, such as Swoogle or Watson (d'Aquin and Motta, 2011).

After receiving a semantic query as input, combiSQORE would use semantic web gateways to retrieve URIs of the potential ontologies that somehow match with the input query. This step was supported by the use of semantic lexical databases, such as WordNet (Miller, 1995). combiSQORE would then download the candidate ontologies into its database and would try to find the ontologies that were semantically similar to the input query. The output of this system was a single or a combination of ontologies that would fulfil the users' requirements.

Appendix F: Recommendation Approaches Review

Some of the very well-known recommendation approaches in the ontology domain are presented in Table F-1. The term "recommendation" has been used in the ontology engineering literature to refer to the systems whose primary focus is on ontology evaluation. However, there is no live selection system in the ontology domain that implements recommendation algorithms, such as content-based filtering or collaborative filtering.

Table F-1 Recommendation Approaches

Tool or Method	Recommendation scenario	Domain	Web service or UI page	Year
WebCORE	Text	General	Not Available	2007
Falcons	Keywords	General	Not Available http://ws.nju.edu.cn/falcons/ontologysearch/index.jsp	April 2008
NCBO Recommender v1	Text	BioPortal	Not Available	2010
BIOSS	Keyword(s)	BioPortal	Not Available	2014
NCBO Recommender v2	Text	BioPortal	https://bioportal.bioontology.org/recommender	2017
Agroportal Recommender	Text	Agriculture	http://agroportal.lirmm.fr/recommender	2017
RecOn	Keyword(s)	General	Not Available	2016

WebCORE. As the first, and maybe the only system to apply collaborative filtering algorithms to the ontology domain, WebCORE would allow its users to describe their selection requirement(s) by a set of initial terms or a text (Cantador, Fernandez and Castells, 2007); it would then pre-process and extend the users' input using NLP techniques and WordNet (Miller, 1995). The other unique feature of WebCORE was that it would present its users with different evaluation techniques and allow them to choose the ones they wanted to be applied in the ontology selection and ranking process. Finally, it would present users with a list of retrieved ontologies and would allow them to manually evaluate and re-rank them using five different criteria, namely, correctness, readability, flexibility, level of formality and type of model (Cantador, Fernandez and Castells, 2007). One major drawback of WebCORE was that the process of ontology evaluation was mostly manual; moreover, this system is obsolete, and there is no published work or evidence available on how it would work.

Falcons. Falcons was a keyword-based search engine for ontologies and also one of the very few systems to use the notion of “recommendation” in the ontology domain. According to the developers of this system, Falcons could provide two different types of recommendation, namely, ontologies for concept search and classes for object search (Cheng, Ge and Qu, 2008). The recommendation method used in Falcons was based on different criteria and techniques, such as coverage, popularity and TF-IDF.

BiOSS. Martínez-Romero *et al.* (2014) tried to address the challenge of finding ontologies by proposing BiOSS, which was a multi-criteria selection approach for biomedical ontologies. Ontology selection in BiOSS would start by accepting a set of biomedical terms as input and end by showing a list or a combined set of ranked ontologies in the output. The evaluation and ranking process in BiOSS were based on three different metrics, including domain coverage, semantic richness, and popularity.

NCBO Recommender. BioPortal applies the notion of recommender systems to the ontology domain by proposing an updated version of the recommender service previously introduced in (Jonquet, Musen and Shah, 2010) and (Martínez-Romero *et al.*, 2014); similar service and technology are also applied to IBC AgroPortal. The NCBO BioPortal Recommender works by getting a list of keywords or a corpus of text as input; it then identifies and presents a list of ontologies or ontology sets that best cover the input query in the output. The evaluation component in NCBO BioPortal Recommender is based on four different metrics, namely, coverage, acceptance, specialization, and the level of details of the ontology classes (Martínez-Romero *et al.*, 2017).

RecoOn is an ontology recommendation approach recently proposed by Butt, Haller, & Xie (2016). RecOn architecture includes four main components, each responsible for different processes, namely, query pre-processing, ontology retrieval, ontology evaluation and ontology ranking. This approach works by getting a query string as input and showing a ranked list of ontologies in the output. Ontology evaluation and recommendation process is based on different metrics, such as matching cost, informativeness, popularity of an ontology and relevance score (Butt, 2016). This system is very similar to other recommendation systems proposed for ontologies, especially BiOSS (Martínez-Romero *et al.*, 2014).

As was discussed in this section, the main focus of most of the ontology recommenders is on identifying a set of metrics that can be used in ontology evaluation process. Apart from Cantador et al. (2007), there is a general lack of research on recommendation algorithms and how they might affect the evaluation and selection process. A possible explanation for this might be that none of the selection systems in the ontology domain would allow their users to provide reviews and ratings for ontologies; therefore, no data that can be used by the recommendation algorithms exist.

Appendix G: Interview Questions of the First Phase

Table G-1 First Phase Interview Questions

N	Question	Detailed Questions	Purpose
1.	**Could you tell me a bit about yourself and your experiences in ontology domain?**	▪ How long you have been building ontologies for? ▪ What domains have you worked in? ▪ How big were/are the ontologies you've built? ▪ Do you usually build domain ontologies or purposive ontologies?	The aim here was to get general information about interviewees, which could then be used to classify the result or trends according to the interviewees' experiences and domain.
2.	**How do you build ontologies?**	▪ Do you follow a specific methodology? If yes which, if not, why? ▪ While building ontologies, do you consider/care about building ontologies that are reusable for others? If yes, how do you make sure that your ontology is reusable? ▪ What are the characteristics of a reusable ontology?	Different methodologies are proposed in the literature to help in the process of ontology development. The aim of this question was to get some general information about how people build ontologies and how do they make sure that the ontologies they are building is reusable (RQ1). It would also help in understanding the role and importance of methodology in ontology domain.
3.	**Do you reuse ontologies? if yes, how?**	▪ How popular is ontology reuse in your domain? ▪ How do you find the ontologies that you want to reuse? ▪ How do you evaluate, select and reuse ontologies?	This was one of the most important questions of this interview and aimed to get some idea about the status of ontology reuse in different domains and the ontology reuse process. This question is relevant to RQ1 and RQ4.
4.	**How do you find ontologies for reuse OR how do you search for ontologies?**	▪ Which search engine do you use? ▪ How do you search? (e.g. keyword(s), corpus, etc?) ▪ How do you evaluate the ontologies that you have found? ▪ How do you think ontologies should be ranked? ▪ Have you ever thought of searching for people with similar expertise or research interests as yourself?	The focus of this research is on ontology evaluation for reuse. So, this question will help in finding an answer for RQ1 and RQ2.
5.	**what are the properties that you tend to look at when judging the general quality or suitability of an ontology?**	▪ Which evaluation metrics do you use? ▪ what do you think makes a good (well-designed) ontology? ▪ Do you think that the number of times an ontology is used can be considered as a factor? ▪ How important is community and social factors in ontology evaluation?	The aim of this question was to identify some of the quality metrics used in the ontology evaluation process (RQ1). This question also helped in finding an answer for RQ3 by comparing evaluation metrics and talking about the role of communities.
6.	**How would you compare your domain with other domains e.g. biomedical?**	▪ What is the status of ontology reuse in your domain?	This question will help in finding an answer for RQ4 and will clarify how using and building ontologies differ in different domains.

Appendix H: A Sample Interview Conducted in the First Phase

Table H-1 Interviewee Details

Interviewee Code	Job Title & Domain	Interview Date
Bi3	Ontology Develop r | Biomedical	23/01/2017

Question: **Can you tell me a bit about yourself and your experience in this field?**

Answer: I got involved first with using ontologies probably 15 years ago, when I started to working for [project name…the rest of the answer is removed for anonymity]

Question: How do you build ontologies? do you use methodologies? can you tell me the steps that you take?

Answer: So I started off in business OBO format I still a kind of use that but interpretation I use is as a translation to owl and I develop in protégé mostly. [Software name] software developed initially by [person name], actually he works with us, I do not use the web based one cause I am very reliant on reasoning software and I tried to follow to as a kind of a [aim?] or rough approximation Allen Rector normalisation which is the try to have only one as a parent and then it automate rest of the classification, but it is pragmatic, you do as much as you can in that direction to make it all sustainable, I am very focused on trying to like a development system well I guess is ..

Question: **Do you use any of those famous methodologies that are out there like METHONTOLOGY?**

Answer: I am not quite sure what METHONTOLOGY is but we use, we do a little bit of a upper ontology where it is a basic formal ontology but we are not, there is not too much that rely on that and we have a centralized ontology repository relations and that so in terms of the upper ontology interpretation, BFO basically drives what we do in the same; so, for example the our approach to time , we have this idea of continues, so process so I am hold on the time and they are other approaches that try to fold the time in a different way.

Question: **One of the main questions of this research is if people use any kind of methodology? Are methodologies useful?**

Answer: I mean what counts as methodology. I mean I think if you are building anything more than a, if you are building very small and then you can probably get away with doing a lot of stuff by hand, but as these things grow, then some level of automation becomes essential; so, the basic idea of try to define equivalent, how familiar you are with OWL, so if you try to use equivalent class expressions to come up with necessary definitions for classes, it is not always possible but where you reasoner comfortable then you can do so, then you do so and you instead of asserting classification, as much as possible you enumerate properties, you enumerate restriction, I guess if only use the owl ontology, and then use the reasoner to automate classification, and then try to sprinkle and disjoint where possible where will allow you to debug where you are really clear that two classes should be disjoint, and import slight reuse ontologies wherever possible so I think loosely you can say that is rector normalization approach, I can point you to [paper name] from the [year], as well and the layout the basic approach, so it is not a very new approach, but I think it is something that has become much more prominent in past few years.

I should say also within the attempt to reuse ontologies that have been developed within the OBO world not exclusively, but so the […] developed around this OBO Foundry, I mean it is not, it was not attempted to bring some standard, I would not say it has been massively successful, but then if you take something like the NCBO, National Centre of Bio Ontology in Stanford, but they BioPortal is a repository for every kind of ontology that any student make so the big problem there is what do you use? you come across and you just use the first one on a list so it worth having some kind of sense of what the quality is, whether it has been used already, so whether it's [worth deducting as standard?], whether it is funded, whether they are responsive, and these kind of things,

Question: **You are a kind of answering my next question, but before going to reuse, I want to ask you something that is related to building ontologies. When you are building ontologies,**

how do you take care of the quality of the ontology? how do you make sure that it is reusable for other people? do you consider this kind of things while building ontologies?

Answer: Yes, I mean I think certainly when working on [project name] and the [ontology name] ontology then when someone constraint by a set of use cases, you know, there are a particular task and queries that you want to be able to do. So, we build in order to, so we have test these cases to say ok we should be able to query to this, and then we use disjoint-ness to check, to check the least rid out the most […] types of errors, we probably should do more of that; so I guess disjoint-ness axioms are really our main , well also we tend to, all the ontology I work on, we use continues integration, and so the continue integration will test the consistency but then we have a lot of syntactical checks that we put in as well. So, and those are often sort of constraints that [refer back to us?] from databases, they need to make sure that we fit within some syntactical constraints that people can load stuff into their databases. In terms of reusability, I guess we do not have anything formal, but we tend to work with curators and request where you know somebody comes across and you and name for something that we don't have yet then I may go in the synonym list, so I mean we could always do more automate this things, more reusable useful but we have quite a lot of input to help pushes in that direction.

Question: **You mentioned BioPortal. do you have any experience using it for ontology search?**

Nowadays, I tend to use the Ontology Lookup Service, [information removed for confidentiality and anonymity] but it is a selected set, it is the OBO Foundry library ontologies plus a few more, so I tend not, it does not have every […] thing that I have ever loaded onto and […] it is the way they set up the API, much better constructed.

Question: **What are the main characteristics of the search engine that you use?**

Nice sort of complete, nice API, nice GUI, I mean it is just pretty well designed, and, I mean actually the previous version of this was really usable, but it was a kind of got very out of date as well, so here is the [link]

Question: **Ok can you tell me how it works? is it keyword-based?**

Yes, I mean it will do; it has a reasonably well-tuned auto complete the works of the labels so I guess all the ontologies here use RDFS label as their [unclear], so usually there is a concept of a label and then synonyms and so search works […] and it is a good way to find if there are other ontologies that are potentially relevant, I have to say, I tend to, in reusing thing, there is often politics and connections are as important as anything else, you know I need to know there is somebody responsive if I want to reuse an ontology and they are reasonably […] and you know there might be constraint in terms of, I may not like a particular ontology but because a bunch of other people are using it and I want to standardize with them, I might use it anyway. So, it is not always the best one that wins, but I mean it is an issue with standard anyway, isn't it? some standard is better than no standards very often.

Question: **After searching, what happens?**

Answer: well you find the term and you go there, and you can browse around, what it does deal is that the fact that ontologies are imported each other. That is only important advances over the previous version I think advance BioPortal as well it does not use them directly for the cross links. Right now, I think they are working on a new site that has not released yet; it does cross referencing between ontologies, as well, but actually sometimes if you want to look for cross references I use, for all the things in the OBO world, I use Ontobee so this is just an auto complete thing in GO in ontology look up service, so I can choose anything but if I am just going to check one term, [the interviewee shared their screen with me and showed me how they do it].

Question: **is there anything else that you want to add about your search experience for ontologies?**

Well visualization is always, I don't know I mean visualization is a hard problem so here we have got, you can scroll down a lot and you see all this paths through ontology or you could look and somewhere here is a connected graph, I don't really like the graph view here but then I am not sure, you know it is a hard problem how you deal with you know when these ontologies get more complex, then I would really ideally view this graphs except may be for particular in order to answer particular question. So, I don't know if there is a good answer in how to make a generic graph viewer

Question: **So, you think that is the problem?**

Answer: well I think a better graph viewer would go in this but it, the problem is that you have to really take care that it does not explode. I will show you using examples [the interviewee shared their screen with me and showed me how that system works].

Question: **My next question is about reuse, you have been reusing and editing ontologies, can you tell me how you do it?**

Answer: Well the reuse, we actually use, we have standard system for generating import modules so we don't reuse the whole ontology; we reuse slice of, you know dynamically generated slices of other ontologies so the files are reasonable size and we used to use some pseudo codes to do that, partly developed by me, when I was at [project name], that was all [confidential information] anyway [confidential information].

Question: **So, how do you evaluate those modules? and how do you pick the module you want to reuse**?

Answer: Well, we would do that before starting making modules really. So, I would say does the ontology, does it look sane, does it look reasonable constructed? is it funded? do the people involved in building it other they responsive? if they are not building something very well, can they take some advice on how to improve? so you know it can be, with the [ontology name] ontology , we are knocked into using [name] ontology or [name] ontology, and bunch of other things, and if they screw up, it screws us up. You know so I mean actually I should say I do a bit of development in the [name] ontology, and there are strong connections between us and [organisation name] on this. So, it is a reasonably close set of groups mostly, but not completely.

Question: **You talked about well-constructed ontology; how do you check? what are the characteristics of well-constructed ontology?**

Answer: You just look at it and see if it makes sense and do they use reasonably parallel naming? or do they mix different adjectives and verbs and nouns sort of randomly, all over the place so you have to normalization in it. So, because when you work with these things, you want to be able to present them in the way that it makes sense to users and if it jumps around too much it is a problem, you know look down at the classification, hierarchy, does it make sense? Go to next to the next, you know for transitive relations, or for classification, is it does it look kind of complex and tangled and disorganized? does it, you know [unclear]. Mainly, I am a [interviewee's job title], I will look at the ontology in [interviewee's domain] and say does it make sense to me? Does it look maintainable?

Question: **You mentioned being funded, how does it affect the quality of ontology?**

Answer: Well, because you may want to request a change; so, it may be high quality but if you never change it, you got have editors, right, then you should have somebody on the other end who can change it themselves. It does not necessarily have to be funded if there is a reasonable group of people with some motivation to carry on working on it, it will do the edits. It is like open source software; why would you reuse a piece of open source software well, actually having an active community around it is really important, when you find a bug, can you go and ask to fix it?

Question: **So next question is what are the challenges and benefits of reuse?**

Answer: Well, I mean one of the reasons we tend to use things from the OBO world is that the annotation properties line up. So, if the OBO translation to OWL or any ontologies that are sort of came from that background using standards to annotation properties. If you use an ontology, that does not come from that background, then it is going to use different annotation properties and it is a kind of hard to mix and match. So, you know I don't think, you know, there are some criticism of OBO standards and stuff, but I don't think it's that important; in the end, you just need a standardization. So, you something can entirely [developed in OWL?] use those annotations properties, but I think the fact that OWL, there was never an OWL recommendation, through annotation property use. I think that causes problems to compatibility between, you know if something uses entirely DC annotation properties, then it is hard for us to work with it.

Question: **What can be done to improve the reuse experience?**

Answer: It is like software, document. I guess the other thing is, [unclear], they need to have a good stable identifier policy, so identifiers must never disappear, things that can [absolutely be replicated?], but there needs to be a clear migration root when they are, automated if possible, or with documentation if not. So, that is really essential for reuse. and again, sticking within the OBO world, we have kind of standard ways to do that, so but that sort of complete …

Question: **You have already mentioned different metrics that you tend to look at when judging the general quality of the ontologies, can you rank the quality metrics that you may use to evaluate ontologies?**

Answer: No, not really, because I am not somebody out there choosing lots of different ontologies all the time. I think you better talk to somebody who does a lot of annotations, so may be talk to some of the [group name] people at the [organisation name], [group name]. I can give you [contact information of people] who are doing sort of more annotating quite diverse things. And whereas, you know, I only reuse a handful and it is hard to rank. You know ideally every ontology I reuse would be really nice structured and well documented, and funded and fast but then you have to a kind of balance, and you have to balance those things and make a decision. But I think being, if it is easy to me to reuse because of the way they set out the annotation properties and everything and there is a same responsive person at the other end and there is a bit of [money?] so they going to keep working on it, that can even trump having a really good quality ontology; but it is a judgment call.

Question: **My next question is about how you compare biomedical field with other fields?**

Answer: I don't know, maybe it is a part of historical accident. The particular group that came together tried to standardize things they maintain relationships with each other. I think also people trying to build this very broad ontologies as well. You know, I have not come across ontology in other fields that much, but I have come across these sort of little very application specific ontologies that may be you wouldn't want to reuse and I think the biomedical ontologies for better or worth have been a kind of bottom up effort, there has been a need for these things so people started to making it before they really understood how to do them. That cause problems, but that does mean that it has been driven by needs, and it generally been driven by biologist. You know the whole sort of computer science OWL world is full of people building these things that has never get reused but they don't get people easy entry points when they don't really get it, they are very computer science oriented.

Question: **It is interesting; so, you think people in industry start building ontologies, then those ontologies will be reused?**

Answer: I think you need computer scientist, so that you can build something, or you need reasonable standards for development for development, so you know I mean the rector normalization well normalization full stop. What is a good normalization scheme so you can go in and just change your thing and error in one place and work? You know, those kinds of things, the things that computer scientist can really help with. So, I think, it is, I mean to be honest, I am [interviewee's job title] by background and I taught myself these stuffs. The more people that just a kind of stand in middle of that I think the better. If the technical side is completely separately from the content side, I don't think that it works.

Question: **You emphasised the importance of community and having someone responsive at the other end, would you put it as the first factor for evaluating ontologies?**

Answer: I would say it is definitely high up; I mean having someone at the other end of line that you feel that you can trust is definitely very important even that or having the editor right yourself.

Question: **What about when they are building ontology? are you interested to know more about the building process and if they have reused any ontology?**

Answer: Yes, I mean if they have got good development practices, that is definitely a big plus. So, you know the other thing is that can they persuaded to adopt better development practices, I guess.

Question: **How do you contact the people who you want to reuse their ontologies?**

Answer: I mean, I would also just check to see if they have a tracker, is it active. If you are going to reuse a piece of open source software you will do the same thing, you will open the GitHub website and say, you know if you looked in it and nobody updated it or anything in three years and no body updated […] three years you might think as if it looks like there is an active ongoing community, you will think yes if I have problems I can ask people and I can get bugs fixed.

Appendix I: Survey Questionnaire Used in the Second Phase

Survey Flow

1. Introduction (1 Question)
2. Building an Ontology (1 Question)
3. If you wanted to build an ontology... (2 Questions)
4. Building a Reusable Ontology (3 Questions)
5. Ontology Reuse Experience (1 Question)
6. Finding a Reusable Ontology (4 Questions)
7. Challenges in Ontology Reuse (2 Questions)
8. Ontology Evaluation (6 Questions)
9. Importance of Community in Ontology Engineering (2 Questions)
10. Demographic (5 Questions)
11. Block 10 (4 Questions)

Introduction

This survey aims to determine the steps ontologists and knowledge engineers use to build reusable ontologies. The sections of the questionnaire determine the metrics that are considered important in ontology engineering and other evaluations before selecting an ontology for reuse.

Q1 - How often do you build an ontology?

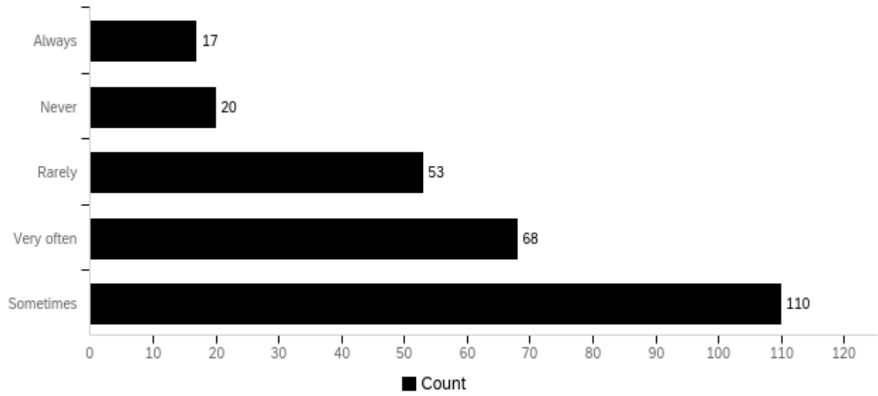

Figure I-1 Ontology Development

Building a Reusable Ontology

Display This Question: If Q1 != 1

Q2-R - While building an ontology, you make sure that it is reusable by:

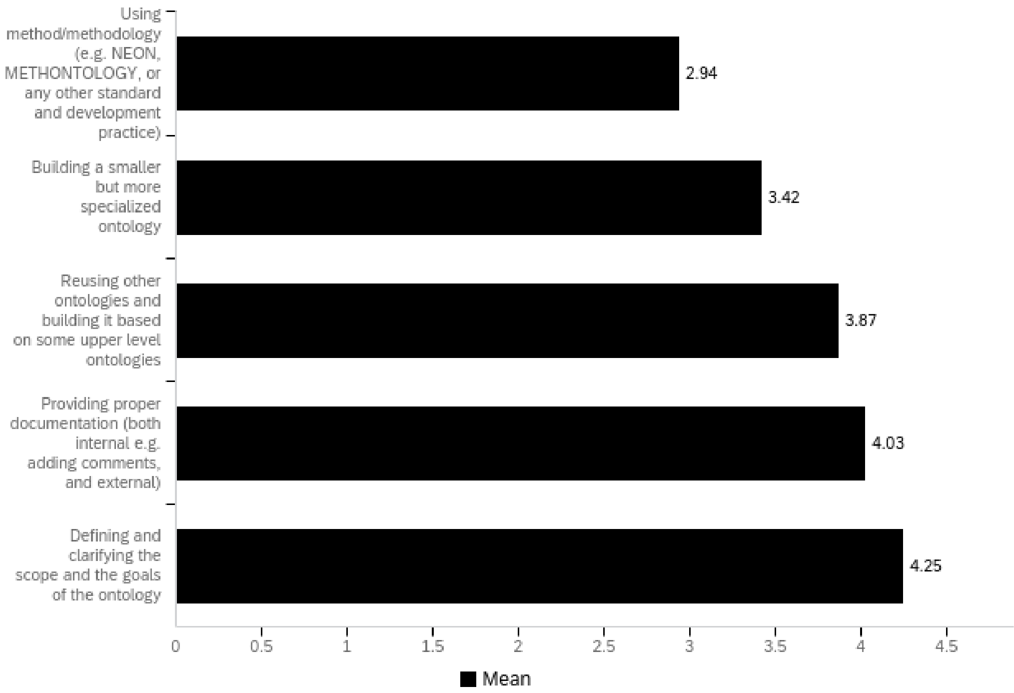

Figure I-2 Reusability Considerations

Q2-2R - If there is any other, please specify:

Sample Answer(s):

- Ensuring consensus by the contributing experts

Q24 - In your opinion, what are the main characteristics of a reusable ontology?

Sample Answer(s):

- One that is developed and well-documented for possible reuse, and that has been vetted by other knowledgeable users as being useful and readily reusable.

Display This Question: If Q1 = 1

Q2-NR - If you ever wanted to build an ontology, you would make sure that it is reusable by:

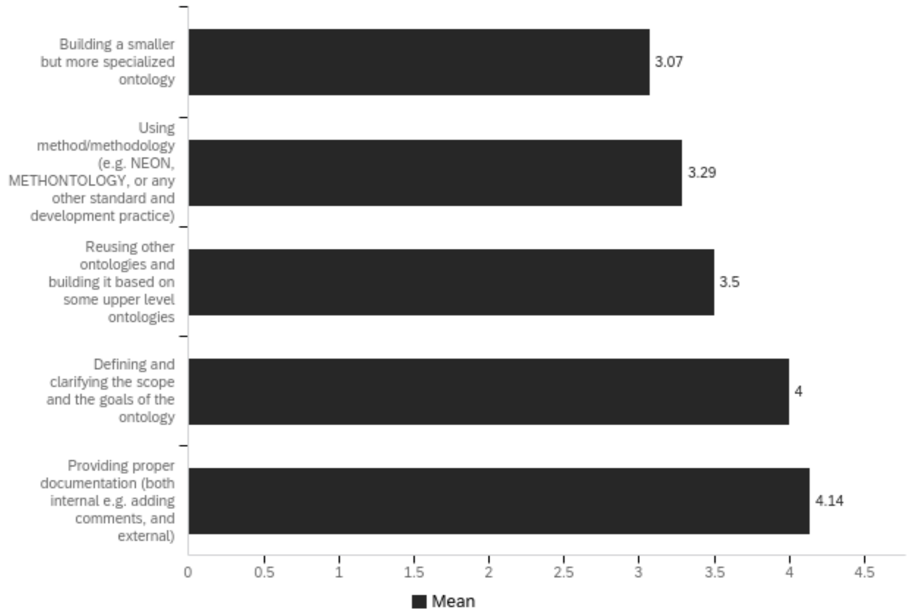

Figure I-3 Reusability Considerations (no Prior Experience)

Q2-2NR - If there is any other, please specify:

Ontology Reuse Experience

Q3 - How often do you consider reusing existing ontologies?

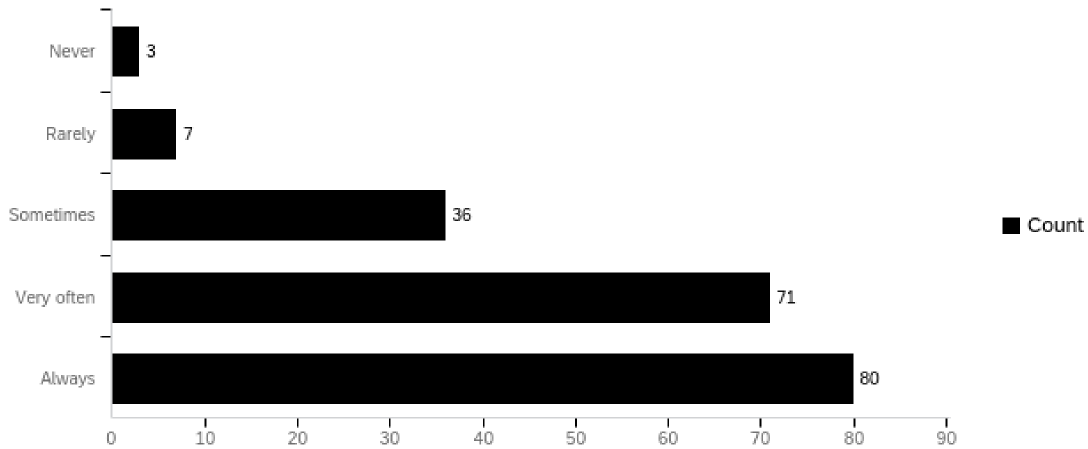

Figure I-4 Reuse Frequency

Finding a Reusable Ontology

Display This Question: If Q3 != 1

RQ4-1 - To find an ontology for reuse:

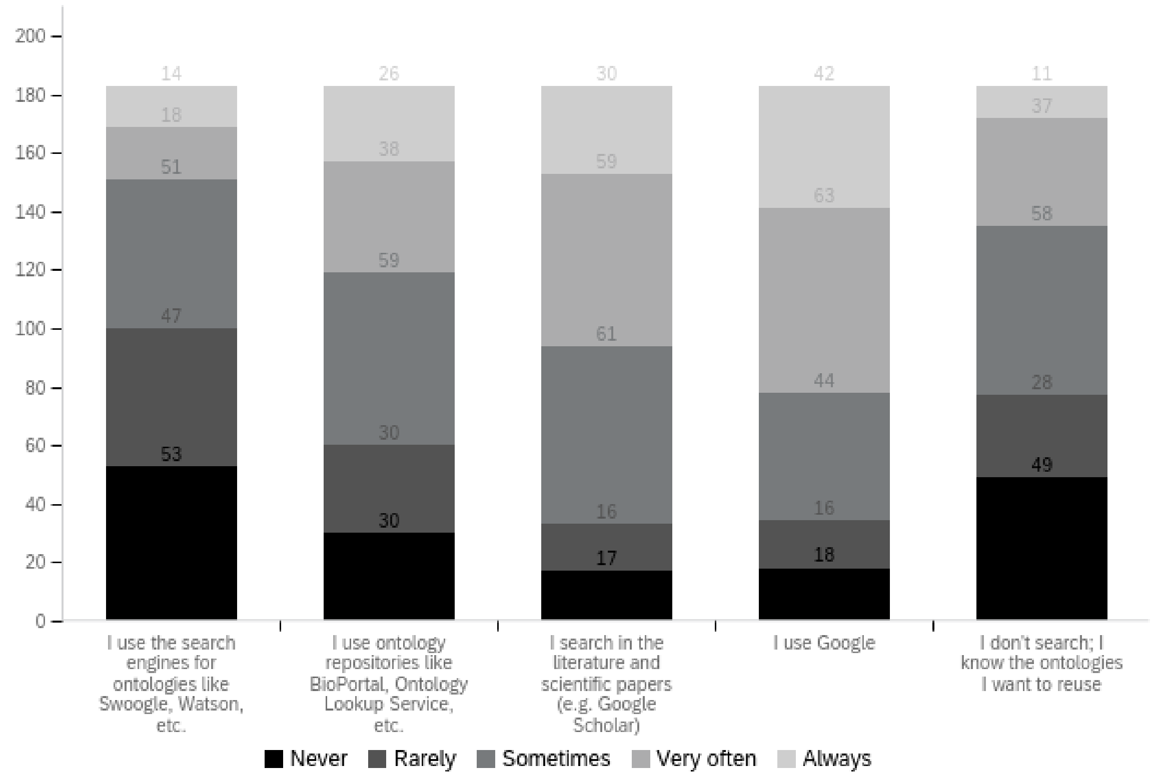

Figure I-5 Finding Ontologies for Reuse

RQ4-Other - If there is any other search engine, repository, etc. you would use, please specify:

Sample Answer(s):

- LOV

Display This Question: If Q3! = 1

RQ4-2 - What are the main challenges faced while searching for a reusable ontology?

Sample Answer(s):

- Ontology is underdeveloped and is not clearly evaluated
- Multiple ontologies that have similar purposes but different models that might overlap, yet contain incompatible notions

Display This Question: If Q3! = 1

Q25 - What is the best ontology you have reused and why?

Sample Answer(s):

- FALDO. Documented for the perspective of a data provider and its limited size made it easy to use.
- Schema.org, because it is easily applicable in a lot of cases.

Display This Question: If Q3 = 1

NRQ4-1 - Why have you never reused an ontology? (e.g. challenges)

Display This Question: If Q3 = 1

NRQ4-2 - Have you ever tried to evaluate the quality of an ontology? (e.g. before selecting it for reuse, annotation, etc.)

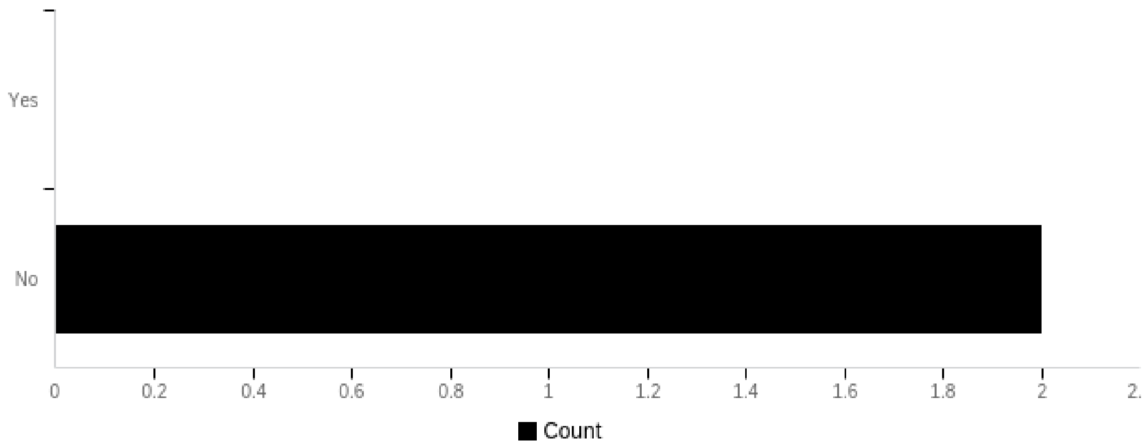

Figure I-6 Evaluation Experience

Ontology Evaluation

Display This Question: If Q3! = 1 Or NRQ4-2 = 1

Q26 - How do you evaluate an ontology before selecting it for reuse? (e.g. evaluation criteria, evaluation approach, etc.)

Sample Answers:

- Is it being used, does it express the concepts well, can the ontology classify instances or is it just a vocabulary?
- Standard (W3C, DC, etc.) or not, then popularity (by default standard), then relevance and quality.
- Developed/endorsed by a standard body or formal interest group, well-documented spec., registered on an Open Vocabulary Registry, references in scientific papers.

Display This Question: If Q3! = 1 Or NRQ4-2 = 1

Evaluation

The aim of the following sets of questions is to survey the importance of the different evaluation factors used before selecting an ontology for reuse.

EQ-1 - When assessing the quality of an ontology, how important do you think the following factors are?

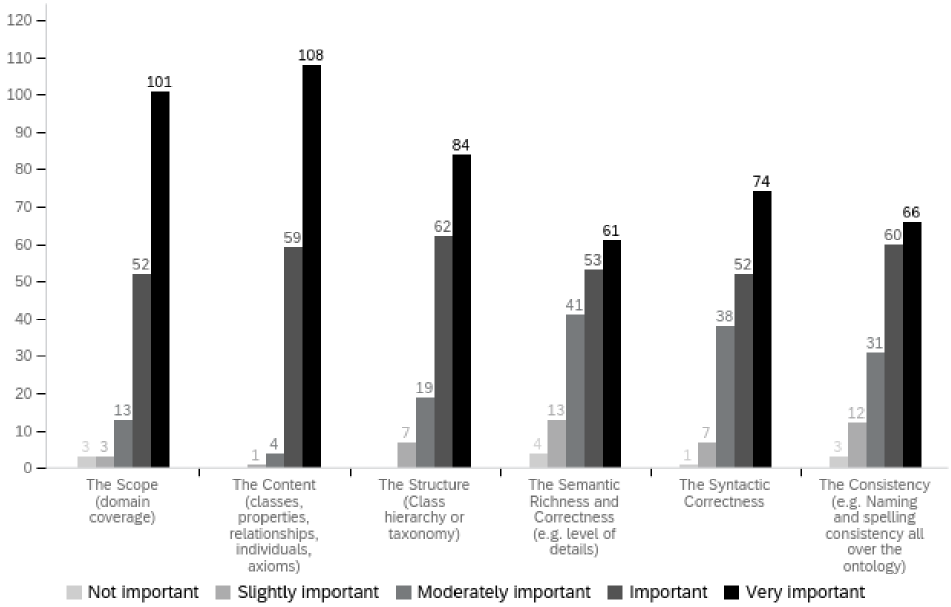

Figure I-7 Importance of Evaluation Factors-Internal Metrics

EQ-2 - When assessing the quality of an ontology, how important do you think the following factors are?

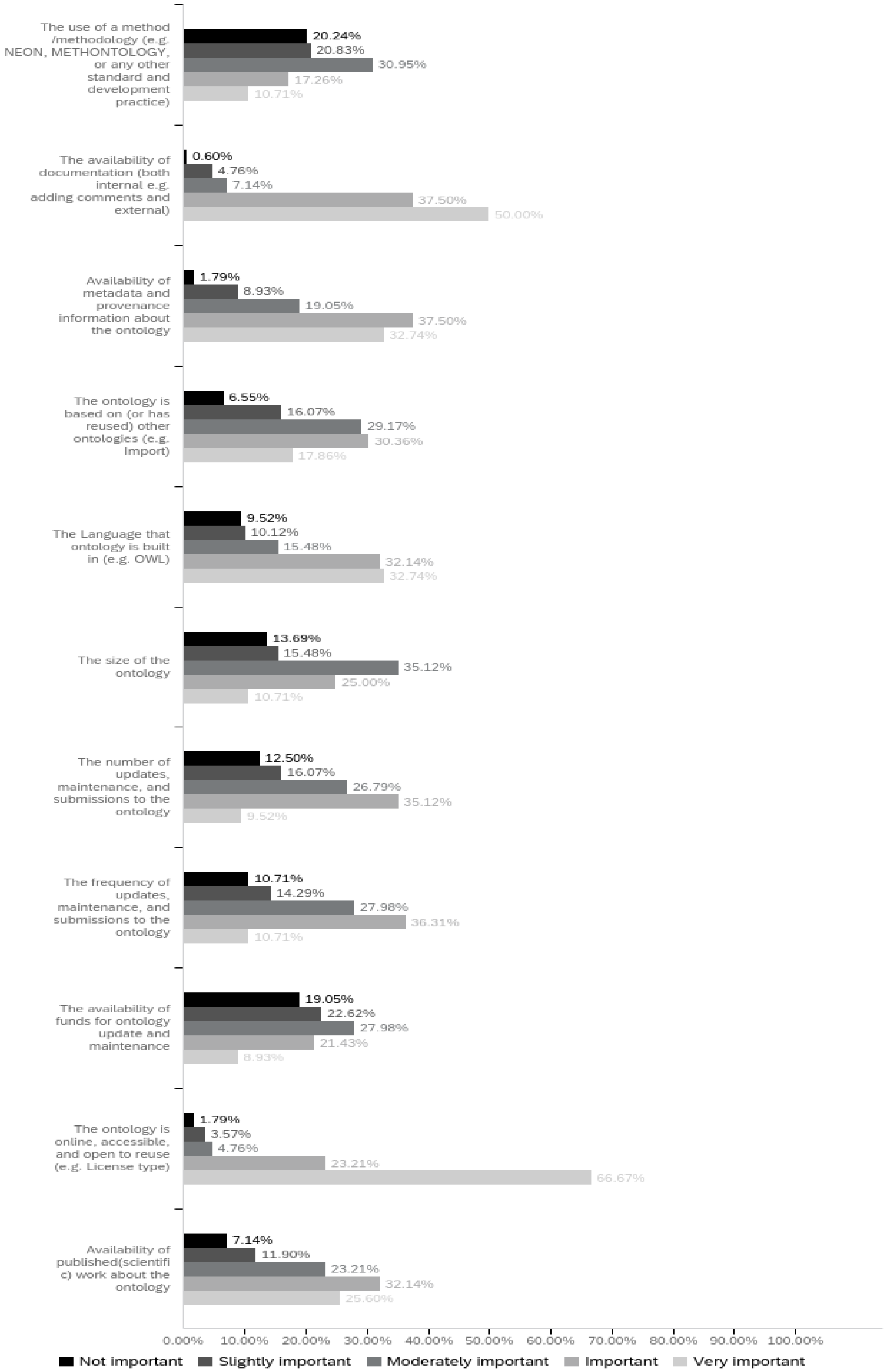

Figure I-8 Importance of Evaluation Factors-Metadata

EQ-3 - When assessing the quality of an ontology, how important do you think the following factors are?

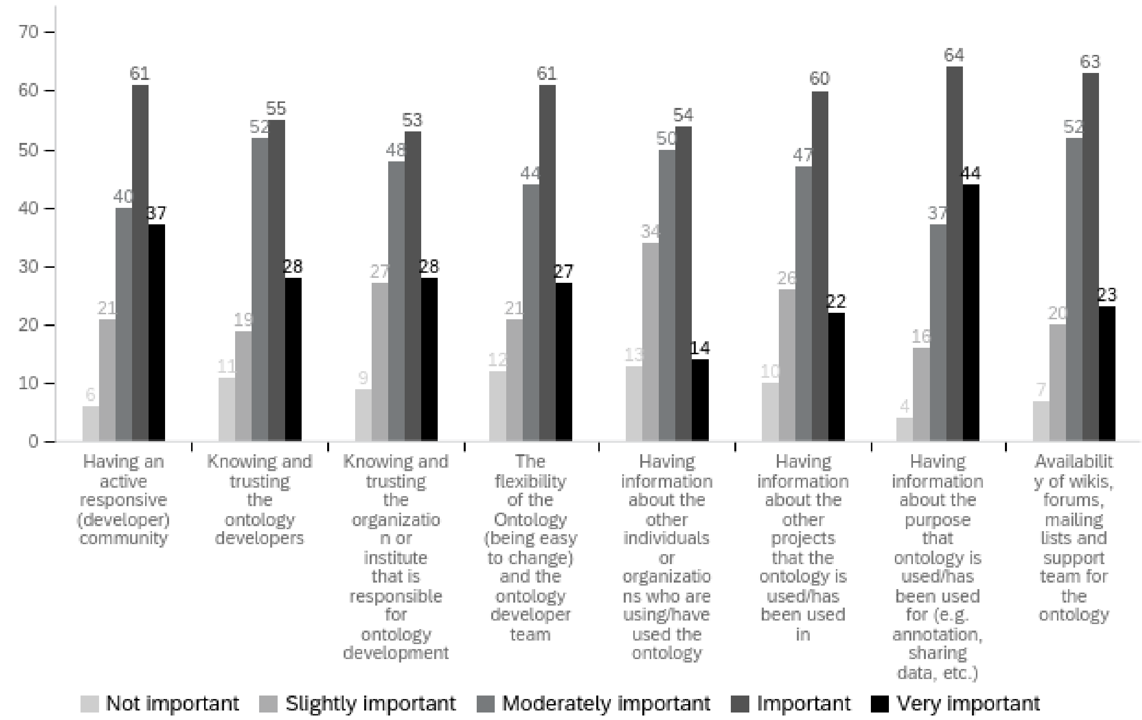

Figure I-9 Importance of Evaluation Factors-Community Aspects

EQ-4 - When assessing the quality of an ontology, how important do you think the following factors are?

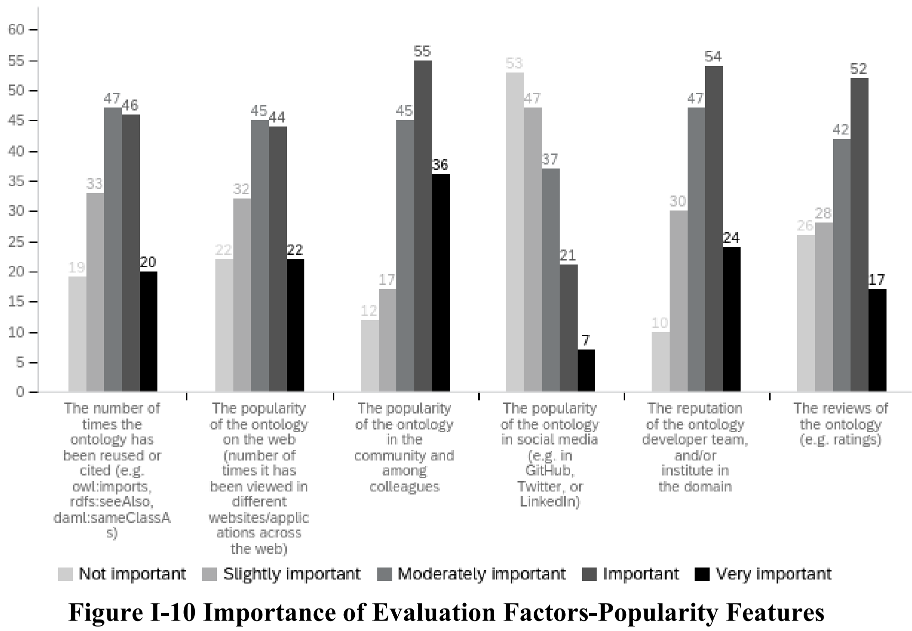

Figure I-10 Importance of Evaluation Factors-Popularity Features

Importance of Community in Ontology Engineering

CQ5 - To what extent do you agree or disagree with the following statements?

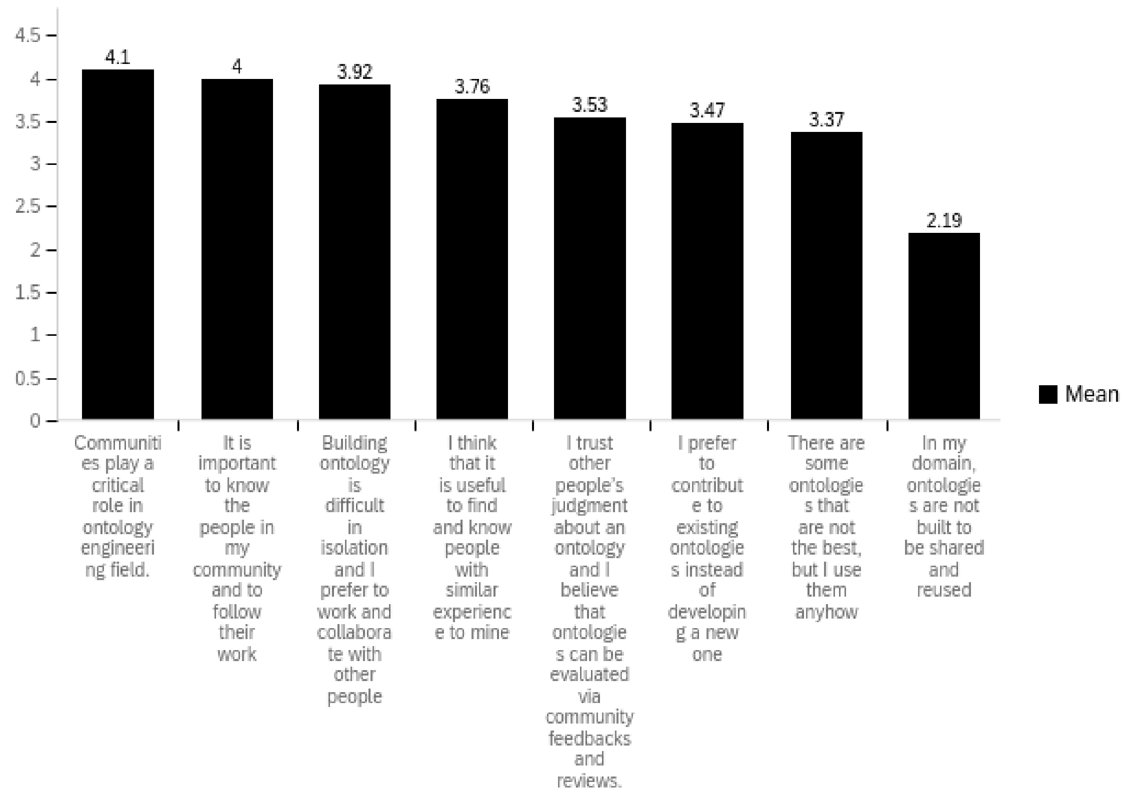

Figure I-11 Community Features

Q33 - Do you have any further experiences you would like to share with us? (e.g. issues reusing ontologies, problems building them, success stories, etc.)

Sample Answer:

- It is critical to work with user communities when building an ontology, to ensure that their needs are met. At the same time, ontology development has to be led by a small group, that makes the decisions. As there may be many (and not always agreeing) use cases/needs. To build a community ontology, it is important to have a clear vision of what you are building, the scope of content while at the same time being open and willing to integrate input from the community.

Demographic Information

DQ-1 - What is your job title?

Figure I-12 Participants' Job Titles

DQ-2 - What type of organization do you work for?

Figure I-13 Participants' Organisation Types

DQ-4 - How many years of experience do you have in the ontology domain

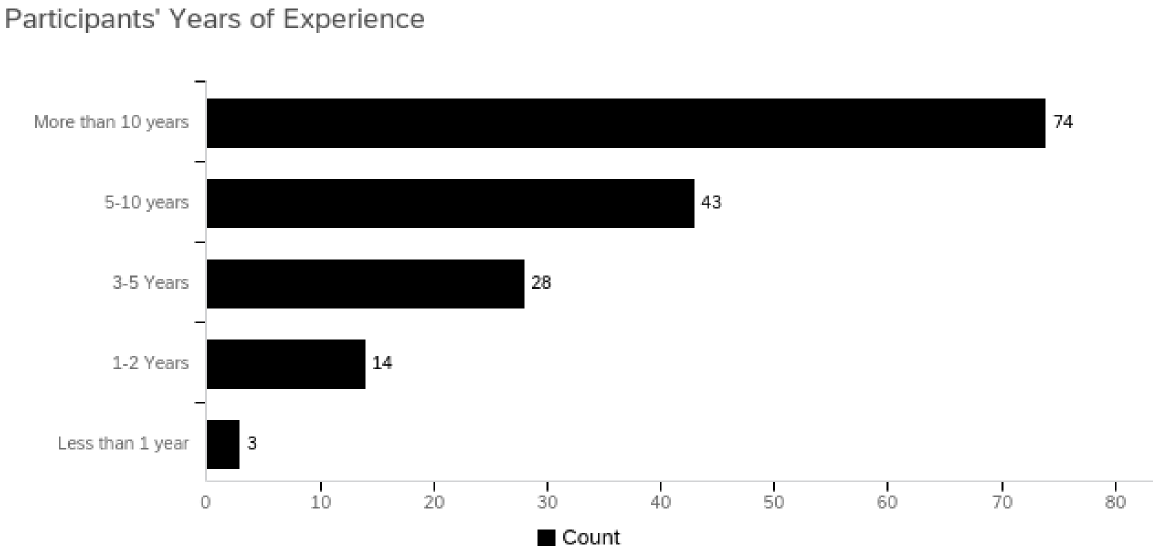

Figure I-14 Participants' Years of Experience

DQ-3 - What are the main domains that you have either built or reused ontologies in?

Figure I-15 Participants' Domains

DQ-5 - What is the primary language that you build ontologies in?

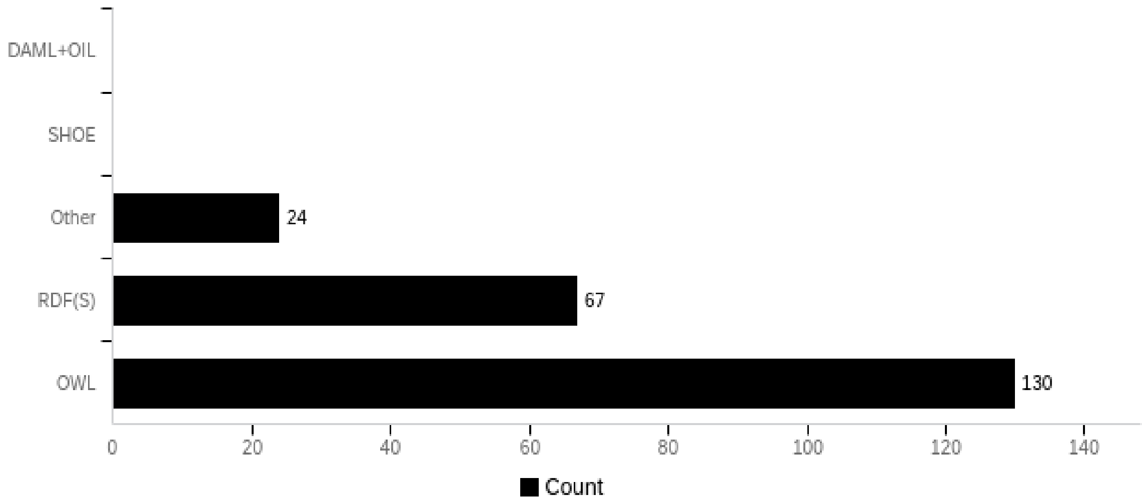

Figure I-16 Main Language Used for Ontology Development

Appendix J: Interview Questions of the Third Phase

Table J-1 Third Phase Interview Questions

N	Question	Detailed Questions	Purpose
1.	**What do you think about the new ranking?**	▪ How would you rank the ontologies in the search result? ▪ Which of the recommended ontologies would you select? ▪ Does the ontology you would select for reuse have a higher rank in the new ranking? ▪ How useful do you think the new ranking is compared to the one by BioPortal?	This question aimed to determine how well the findings of this study can predict the ontology knowledge engineers would select for reuse.
2.	**How important do you think the metrics identified in this study are in the evaluation process?**	▪ Do you find the factors proposed by this study useful in the selection process? ▪ Do you find the explanations and the extra information helpful in the selection process? ▪ Would you like to have access to the information this study is providing, while selecting an ontology for reuse?	This question investigated whether ontologists and knowledge engineers find having information about the factors proposed by this study helpful in the selecting process or not.
3.	**What do you think about the weights assigned to each selection criteria?**	▪ Do you think metrics used by selection systems should have different weights? ▪ What do you think the most important metric in the evaluation and selection process is?	This question helped in exploring the importance of different criteria used in the selection process.
4.	**General Questions**	▪ What do you think about the current selection systems in ontology domain?	This question explored interviewees opinions about the current selection systems for ontologies.

Appendix K: Transcripts of the Interviews Conducted in the Third Phase

Interviewee A

Interviewee A—Query used in the experiment

The DNA damage response (DDR) involves a complex network of signaling events mediated by modular protein domains such as the BRCA1 C-terminal (BRCT) domain. Thus, proteins that interact with BRCT domains and are a part of the DDR constitute potential targets for sensitization to DNA-damaging chemotherapy agents. We performed a pharmacologic screen to evaluate 17 kinases, identified in a BRCT-mediated interaction network as targets to enhance platinum-based chemotherapy in lung cancer. Inhibition of mitotic kinase WEE1 was found to have the most effective response in combination with platinum compounds in lung cancer cell lines. In the BRCT-mediated interaction network, WEE1 was found in complex with PAXIP1, a protein containing six BRCT domains involved in transcription and in the cellular response to DNA damage. We show that PAXIP1 BRCT domains regulate WEE1-mediated phosphorylation of CDK1. Furthermore, ectopic expression of PAXIP1 promotes enhanced caspase-3-mediated apoptosis in cells treated with WEE1 inhibitor AZD1775 (formerly, MK-1775) and cisplatin compared with cells treated with AZD1775 alone. Cell lines and patient-derived xenograft models expressing both PAXIP1 and WEE1 exhibited synergistic effects of AZD1775 and cisplatin. In summary, PAXIP1 is involved in sensitizing lung cancer cells to the WEE1 inhibitor AZD1775 in combination with platinum-based treatment. We propose that WEE1 and PAXIP1 levels may be used as mechanism-based biomarkers of response when WEE1 inhibitor AZD1775 is combined with DNA-damaging agents.

Interviewee A—Ranking comparison

Table K-1 Ranking Comparison for Interviewee A's Query

Ontology Name	New Rank	BioPortal Rank	Comments
National Cancer Institute Thesaurus	1	1	Same
SNOMED CT	2	2	Same
Medical Subject Headings	3	3	same
Logical Observation Identifier Names and Codes	4	5	External documentation and funds available
Neuroscience Information Framework (NIF) Standard Ontology	5	4	
Online Mendelian Inheritance in Man (OMIM)	6	8	Some documentation available and a website and mailing list
Read Codes, Clinical Terms Version 3 (CTV3) (RCD)	7	10	Availability of external information, have been reused before
Computer Retrieval of Information on Scientific Projects Thesaurus	8	6	
Interlinking Ontology for Biological Concepts (IOBC)	9	7	
Regulation of Transcription Ontology (RETO)	10	9	

Interviewee A—Transcript

Question: **What do you think about the new ranking?**

Answer: Information is king here. Transparency of metric is important; tell people why something is third. BioPortal gives you some transparency; it is a bit confusing, but they try to give you some info. Ontology ranking is more helpful when all the information is available.

Question: **What do you think about the metrics proposed by this PhD? Do you think they are important in the selection and evaluation process?**

Answer: Yes, I think they all are.

Question: **What about the comments and the extra information that is provided?**

Answer: It will be useful to see information, to see the updates per year. If you are new to the field, it is very helpful.

Question: **How important do you think each one of the following metrics are in the evaluation process? Authority and who has built an ontology?**

Answer: Someone's PhD, maintainanance or group of people or organisation or an professional organisation definitely to consider. Accurate and maintenance shows something; authority.

Question: **How important do you think the language that an ontology has been built in is?**

Answer: Is not a huge deal but if you are using pipeline it can hurt a bit; will consider UMLS outliers; (language) It's a metric, not the first. Coverage is first.

Question: **Ontology usage info?**

Answer: Usage info is very important; it's second.

Question: **What about documentation and metadata?**

Answer: Versioning, process, ticketing system; that's important too.

Question: **What about the size of an ontology?**

Answer: It is useful to know how big the ontologies are. Wikipedia can always win but it doesn't mean it is the best. Spectorum tiny to huge; two exterems are not useful. Very small ontologies or very huge ontologies are not useful.

Question: **What do you think about the frequency of updates?**

Answer: Ontologies should be updated more than once a year. Actively developed ontologies are updated every month.

Question: **What about the ontology acceptance?**

Answer: Acceptance, meaning that some people are using it in BioPortal maybe not that useful. What database are using an ontology and who uses it is more imp than the number of times.

Interviewee B

Interviewee B—Query used in the experiment

Gene, protein, mutation, tissue, location

Interviewee B—Ranking comparison

Table K-2 Ranking Comparison for Interviewee B's Query

Ontology Name	New Rank	BioPortal Rank	Comments
NCIT	**1**	**1**	NCIT was reused 16 times while LOINC was reused 8 times. NCIT is a part of OBO Foundry and is updated monthly while LOINC is not in OBO Foundry and has been updated less than NCIT. However, there is a wiki/forum for LOINC!
LOINC	**3**	**2**	LOINC has been reused 8 times; external documentation is available for it and has wikis & forum
CRISP	**5**	**3**	Has a very high acceptance score; however, has never been reused before in BioPortal and is not a part of OBO; is built using UMLS; is better than HUPSON (because it was only once uploaded/updated in 2014)
SNOMEDCT	**2**	**4**	One of the most popular ontologies; has been reused 23 times; is frequently updated and has wikis/forums. Higher acceptance and number of reused compared to the ontologies that were ranked 2nd and 3rd.
NIFSTD	**4**	**5**	Has previously been reused (5 times compared to CRISP, which hasn't been reused before); the acceptance score is really lower than CRISP, but it has external documentation and is built in OWL
HUPSON	**9**	**6**	Very similar to AURA; but is smaller and has a slightly higher acceptance; so, it is ranked higher.
CHEAR	**6**	**7**	Very similar to HUPSON but has been updated (3 times) in 2018 so it goes higher
GRP	**7**	**8**	Has been reused twice but was last updated in 2016; was built based on OBO; however, it is marked as obsolete there. Is small and is built in OWL
AURA	**10**	**9**	Has never been reused before. Is larger than the ontologies that are ranked similarly with it.
PANDA	**8**	**10**	Is smaller than HUPSON and was updated twice in 2017 (which is better than HUPSON that was last updated in 2014)

Interviewee B—Transcript

Question: **What do you think about the new ranking?**
Answer: The ranking in itself may not be of the greatest value. May be the fact that one ontology is reused a lot or has been updated multiple times is not the most important for a user (at least this is dependent on the end task). What is VERY useful is to have the comments provided that can in a glimpse provide more confidence for a user to make an informed decision while picking an ontology of his choice.
Question: **How useful do you think the new ranking is?**
Answer: It is moderately useful.
Question: **The new ranking is based on a set of metrics; how useful do you think each one of them is? The first one is Documentation.**
Answer: It is very helpful
Question: **What about metadata about the ontology?**
Answer: It is moderately important
Question: **Is it important if the ontology is built based on a methodology or a standard?**
Answer: For me, it is very important.
Question: **What about the size of an ontology?**
Answer: It is slightly important
Question: **And the language?**
Answer: Language is moderately important
Question: **Information about the other projects that have reused the ontology?**
Answer: It is slightly important
Question: **And the people or organisations that have reused the ontology?**
Answer: [Having access to] this information is moderately helpful and important.
Question: **What about the purpose that the ontology has been reused for? E.g., annotation?**
Answer: It is slightly important.
Question: **What about the popularity of an ontology in the domain?**
Answer: It is very important
Question: **What about the reputation and popularity of the ontology developer team?**
Answer: It is slightly helpful.
Question: **Do you care about the frequency of the updates of an ontology?**
Answer: No, this factor is not at all helpful.
Question: **If you wanted to rank the importance of the mentioned metrics, which one would be the most important for you and which one the least important.**
Answer: Documentation is the first and the most important for me. Popularity of the ontology in the domain is second. Following a standard or methodology is the third. Language is fourth for me. Metadata is the fifth important.
Question: **Which one is the least important?**
Answer: Frequency of updates is the least important.
Question: **Do you think different weights should be assigned to the metrics in the evaluation process?**
Answer: I think what would be useful is to let the user change the weight he wants to assign for each of the three categories.
Question: **Is there anything that you would like to add?**
Answer: No.

Interviewee C

Interviewee C—Query used in the experiment

Trait

Interviewee C—Ranking comparison

Table K-3 Ranking Comparison for Interviewee C's Query

Ontology Name	New Rank	BioPortal Rank	Comments
NCIT	**1**	**1**	
VT	**3**	**2**	Just because it was reused less than PATO; but it has github
UPHENO	**6**	**3**	A part of OBO and has github
AURA	**10**	**4**	
FAST-TOPICAL	**8**	**5**	Has wikis, that's why it is higher than the remaining 2
CRISP	**5**	**6**	Has a very high acceptance score; has been updated 3 times in 2018
HUPSON	**9**	**7**	Smaller and has a higher acceptance than AURA
CARO	**4**	**8**	Has a lower acceptance score compared with CRISP, but is a part of OBO Foundry, is very small
PCO	**7**	**9**	Has a higher acceptance than UPHENO, but has been last updated in 2017
PATO	**2**	**10**	Has been reused 14 times before; has been updated 4 times in 2018; follows OBO Foundry; higher acceptance (than VT) in BioPortal

Interviewee C—Transcript

Question: **What do you think about the new ranking?**

Answer: I mean, ahh, it moves things slightly, but I am not familiar with all the ontologies that are here. Some of them I am. In terms of doing stuff, just don't know them, but I mean in the list there are things that I would expect to show up and they show up in both ways.

Question: **So, what about the order? Does the order really matter? Like which one comes first and second?**

Answer: In some respect for this one I am not sure why PATO would have dropped. No, I take it back. No, in the new ranking, its 2. That would make sense and some of the others.

Question: **Like CARO, it was ranked 8th, it is now number 4?**

Answer: That one I am not familiar with. The ones I knew about in this area were PATO, VT, and then the PCO are the ones I knew the most about. Well, UPHENO as well, take that back, I would have known that one too. But having the PATO and VT toward the top make sense. I mean why PATO was so lower ranked in BioPortal rank, I am not exactly sure. Some of that may be just semantic. It is in the difference between what is in Trait and what phena type.

Question: **If you want to compare PATO with VT, PATO is ranked higher because it has been reused more and has a higher acceptance in BioPortal, which is not the best way of doing it, but we have no other option because there is no way of finding acceptance other than in BioPortal.**

Answer: It makes sense; PATO has been used a lot.

Question: **So, it makes sense?**

Answer: Yes

Question: **Would you consider my raking, and would you like to see all the information that I have provided while you are evaluating an ontology?**

Answer: Having those things available, if in [unclear] I am looking for, is very useful, yes.

Question: **So, you find them useful?**

Answer: Yes

Question: **How important do you think each one of the following metrics are in the evaluation process?**

Answer: Well, for some of them like size it's probably going to depend upon the domain which they are operating in. You are comparing multiple in the same domain, then it should be reasonable. But if they are across two different domains, that I think would probably be not as important.

Well the other would be is the ontology precomposed or decomposed may make a difference in what they are. So, if you look at the difference between PATO and VT one of them is precomposed, the other one is decomposed and that would probably make a difference and there will be other examples.

How you get after that reputation and popularity, it is pretty subjective portion but if there is way of doing it, which I think to some extend your other, I mean, how often is it reused or how many times is it, that probably gets that question to some extent. So, I think there are probably some redundancies in some of things which you have cause you could probably take your different metrics that you have here and put them in some respects, larger sub categories in which they are trying to get after.

Question: **Yes, I have done factor analysis and they each belong to different groups**.

Answer: Yes, make sense

Question: **What about being based on any standard or methodology like OBO foundry? when you are evaluating something, so you want it to be based on OBO or something?**

Answer: When we use things ourselves, we had them in both formats, so we are able to use them. If they were all similar, yes probably would make life easier but there are so many things that's not the case. I am not too hung up on that personally.

Question: **Do you think different weights should be assigned to the metrics?**

Answer: For the most part, I would say yes, the answer is yes. What I don't know is what's the best way about going, about figuring out what those weights should be. Because in my own field, the genetics, people do a lot of works, specifically, in order to identify what the weighting should be but that becomes a research project in [inaudible] itself.

Question: **Is there anything that you would like to add?**
Answer: No, I mean it make sense in terms of what you provided. And hope that this is helpful for you

Interviewee D

Interviewee D—Query used in the experiment

Natural Language Processing, health informatics, ontology

Interviewee D—Ranking comparison

Table K-4 Ranking Comparison for Interviewee D's Query

Ontology Name	New Rank	BioPortal Rank	Comments
National Cancer Institute Thesaurus	1	1	Same
EDAM	2	2	Same
Medical Subject Headings	3	4	Has been reused 15 times. Has been updated in 2018 while SWO was last updated in 2017. Has a very higher acceptance than SWO
SWO	4	3	Is smaller than NIFSTD. Is in OBO foundry. I know the developer and I know he is responsive
Neuroscience Information Framework (NIF) Standard Ontology	5	8	Has publication (IOBC hasn't). Has GitHub. Has a higher acceptance compare to number 6. Has been reused 5 times (number 6 hasn't been reused yet; maybe because it is new)
Interlinking Ontology for Biological Concepts (IOBC)	6	6	Same
PLOSTHES	7	5	Is ranked lower than the ones on top of it because it has never been reused, it is built in SKOS, has last been updated in 2017. Has been ranked higher than the ones below it because has publication, has GitHub
BRO	8	9	Seems like a good ontology as it has 73 notes, average acceptance score (24.8), has been reused 4 times, has publication, very small size (488 classes) and is built in OWL. But, it was last updated in 2010!!
RH-MESH	9	7	Has been last updated in 2014 and has a slightly higher acceptance (28.1). But is very large compared to number 8 (305,249), has been reused 3 times, has no notes, has no publication.
BRO-ACRONYM	10	10	Same

Interviewee D—Transcript

Question: **What do you think about the new ranking?**

Answer: Yes, so I know about, I work with MESH. I think for my query, I don't know the MESH is the most suitable ontology, but I am sure that, for example, the NCIT I do not think it is the best ontology for my query. My query was about natural language processing, [unclear] and the third keyword was [unclear] but I think for such a query, the MESH is the first ontology. Because it is very, because the mesh is used in PubMed is used in many biomedical databases and I think for me, I know little bit about the NCIT, I know what the MESH is because I work with the MESH. For example, if we compare this ontology, the best ontology will be the MESH compared to NCIT. For the other ontologies, I don't have any idea about. But so, I think your new ranking compare to the BioPortal rank for example, for me, MESH is good. In your rank, the MESH is third in BioPortal is 4^{th}. So, compared to it is good. I don't know the second ontology maybe it is good ontology.

Question: **So, you think the new ranking is a bit better, not for the first 2 which I kept the same but for the third and fourth place, it is better because you would choose MESH?**

Answer: Yes, exactly

Question: **The problem with BioPortal is that NCIT is shown as the first ontology for many of the queries. Do you find my ranking more helpful or less helpful?**

Answer: If we suppose that the NCIT is given by default as the first place, I don't know about EDAM, the second ontology, I don't know which domain it is applied, but I think when we see your ranking for mesh compared to BioPortal, I think your ranking is better than BioPortal.

Question: **So, I want to talk a bit about the metrics I used to come with this ranking and ask you how useful you think the metrics are and ask if you would consider these metrics while evaluating ontologies or not. The first metric I am using is documentation; is it important if some kind of external documentation e.g., website is available for an ontology? do you want an ontology to have documentation when you want to reuse it?**

Answer: Yes

Question: **It is important?**

Answer: Yes

Question: **What about metadata about an ontology?**

Answer: Yes, I think the best metadata for me which is very important when we deal with ontologies is the version of the terminology. MESH for example, I think there is a new version, so it is important. If each and I think there are many changes between two versions, so I think version as metadata is very important.

Question: **Do you want the ontology to follow any standard? OBO foundry, W3C?**

Answer: This is important for me. Now I deal with RDF semantic Data. It is very helpful we know which format the ontology exist and if the N3 or using RDF XML. So, it is important.

Question: **What about the size of an ontology, number of classes?**

Answer: It is very important

Question: **Do you prefer smaller ontologies or the bigger ones?**

Answer: You know this question depends on why we want to use an ontology? for example, if we use an ontology, for example, indexing, it will be useful that the ontology must be not very big because when we want to look for any term or any concept we need the system to be faster. Nevertheless, if you want to use some tools and show how [unclear] of the tool is going to deal with the ontology, so yes, we need that the ontology be the most [unclear].

Question: **So, you care about the size; and the language that it is been built in? OWL, RDF/ do you prefer an ontology that has been built in owl or you don't mind?**

Answer: For me it is not important, but I think for many research projects it will be important because there are some tools which prefer RDF because it is lighter in terms of reasoning and in term of W3C. For me it is not important. But I think there are some research projects which the difference between OWL and RDF is very important.

Question: **What about who else has refused an ontology? or in what project the ontology has been reused? Do you want to have some information?**

Answer: No, it is not important for me

Question: **What about for what purpose the ontology has been reused? (e.g., annotation)**

Answer: Purpose is very important. For example, we know that MESH is very useful in term of indexing biomedical databases but NCIT we know that is used for cancer. The purpose is very important.

Question: **But who and what project is not?**

Answer: Who I think it is not. Since we know that the ontology is useful for some purpose, I think who uses it is not that much important for us but yes who uses it is not as important.

Question: **What about popularity of an ontology in the domain? and how would you define popularity?**

Answer: So, I think that popularity is depends who use the terminology. As in example of MESH, it is used PubMed search engine, so I think it is very popular. I think SNOWMED CT is very good ontology is a very big ontology but in France we never use it but because it is not anywhere in FRANCE, so we use another terminology for medical domain, but I think for me, the popularity of an ontology depends on who is using it. If it is a big one or a big company that is using the terminology, I think yes, the terminology will be very popular.

Question: **So, it is not about the number of reuses, e.g., how many times an ontology has been reused? it is about who has used it?**

Answer: Yes exactly. Popularity does not depend on the number of reuses.

Question: **What about reputation and popularity of the ontology developer team or organisation? do you care who has built the ontology?**

Answer: Yes exactly, we care about this because sometimes we need to define if we are you know authorised to use the ontology or not. Yes, for example, if it is you know private company who developed the ontology, we know that it is very difficult to use it in public hospital or any public product. So, I think it is very important

Question: **So, I guess the next one is also relevant; maintenance and availability of contact information? do you want the ontology to be maintained and do you want to be able to contact the people who have built the ontology?**

Answer: Yes, I think it is very important. So, we have this problem with MESH and we needed to contact the developer of MESH which is [unclear] and yes, I think it is very important if we use the [unclear] in our product, so we need at least a contact information to have some, to ask some questions.

Question: **What about frequency of updates?**

Answer: Yes, it is very important. The frequency of update is very important. I think the MESH is updated every year, so I think it is very … So, I remember when we developed with MESH so it was some problem because every year we need to update our system but I think if the update is, for example, if the terminology is maintained so it is ok but if the frequency is very very large in year it will be ok.

Question: **Which one of the metrics do you think is the most important for you? when you are evaluating an ontology, what is the first quality related metric that you would look at?**

Answer: I think the best thing, it is different id we need indexation, we need that the ontology in the context of indexation of any resource [unclear] , so we need the ontology much more [unclear] for our purpose for me for example if we need to test some tools to load some ontology or to query an ontology the size is very important. So, there is two parameters for me purpose of the terminology: what I need to do with this ontology or the size and the number of classes and properties and the format if we need to use the ontology in any system so test in any system.

Question: **So, you care about if the ontology is good for indexing or is good for annotation or what the ontology is good for so that is the first thing?**

Answer: Yes exactly.

Question: **So, I have another question. Every selection system is based on so many different metrics. Do you think each metric should have different weight and importance?**

Answer: Yes, I think so. I think if there is a metric the score is given to any metric will be different for example, I will give more score for purpose than the popularity.

Interviewee E

Interviewee E—Query used in the experiment

Quitting smoking is hard. Some people who have quit say that it was the hardest thing they have ever done. But most smokers eventually are able to quit smoking. And you don't have to do it alone. Ask your family, friends, and doctor to help you. Get what you need to help you quit for good. Get ready. If you're ready to quit right now, go ahead. Medicines and support can help you stay on track. But if you want to plan ahead, you don't have to stop right away. Set a date to quit. Pick a time when you won't have a lot of stress in your life. Think about cutting down on smoking before your quit date. You can try to decrease the number of cigarettes you smoke each day as a way to quit smoking. Get rid of ashtrays, lighters, or spit cups before you quit. Talk to your partner or friends about helping you stay smoke-free. Don't let people smoke in your house. Change your routine. For example, if you smoke after eating, take a walk instead. Use medicine. It can help with cravings and stress, and it doubles your chances of quitting smoking.

Interviewee E—Ranking comparison

Table K-5 Ranking Comparison for Interviewee E's Query

Ontology Name	New Rank	BioPortal Rank	Comments
SNOMED CT	1	4	Has the highest number of reuse and the highest acceptance score. Is frequently updates and is developed by a very known organisation (NHS).
National Cancer Institute Thesaurus	2	1	Is very similar to MESH but is ranked higher because it is smaller, it is developed in OWL, it is a part of OBO Foundry and is updated every month
Medical Subject Headings	3	2	Has a lower acceptance than RXNORM but has been reused 15 times compared to 7 times of RXNORM.
RXNORM	4	10	RXNORM is also very similar to LOINC and has a higher acceptance than LOINC. LOINC has been reused more (8 times) than RXNORM (7 times). LOINC has a community but the organisation responsible for RXNORM (National Library of Medicine serves) is more popular than the one responsible for LOINC (Regenstrief Institute); national level vs institute level. Moreover, RXNORM is smaller than LOINC
Logical Observation Identifier Names and Codes	5	3	
Neuroscience Information Framework (NIF) Standard Ontology	6	7	Has a very lower acceptance compared to LOINC (28.9). Compared to the ones below it: has been reused more, is developed in OWL
MEDLINEPLUS	7	8	Very similar to RCD; both have been reused twice and are updated similarly and are built in UMLS. MEDLINEPLUS has a slightly lower acceptance score but it is very smaller than RCD (2238 classes vs 140065).
Read Codes, Clinical Terms Version 3 (CTV3) (RCD)	8	6	Is ranked higher than the two ontologies below it because has documentation, has been reused more and has a higher acceptance score than them.
CHEAR	9	9	Both CHEAR and CRISP are developed by national level institutes. CRISP has a very higher acceptance score compare to CHEAR however, it has never been reused while CHEAR has been reused once. Plus, CHEAR is developed in OWL.
CRISP	10	5	

Interviewee E—Interview Transcript

Question: **What do you think about the new ranking?**

Answer: I guess I agree that *SNOMED CT* is probably the most useful terminology for the kinds of concepts that are discussed in the text that I used, because it is the broadest of all the clinical vocabularies out there really in terms of coverage. It also does a pretty good job of representing what I would call consumer health concepts as you may have inferred from the text that I use the kind of work that I do is often from a consumer perspective on health or I'm often developing models that attempt to bridge or cross walk between consumer and clinical models and *SNOMED CT* does a fairly good job of that; I mean of course it does contain very clinical concepts as well but it also contains concepts that are much more general or that are phrased in sort of a consumer friendly way so in that sense *SNOMED* is pretty relevant for this text.

Question: **Yes, so in BioPortal it was ranked fourth, but it is ranked first in the new ranking.**

Answer: right

Question: **Do you think it is a better way of ranking it? do you find it more useful or not?**

Answer: In the case of that particular ranking, yes. I think having it ranked higher is probably correct as far as my, the way that I would evaluate the suitability of this vocabularies for classifying this test. But it is a sort of a weird problem, when you put some full text from an article as input to try to find an ontology, that can potentially be useful, but I do not know that it is useful over time. If you notice the mark-up of the input text, the way that recommender returns it with the blue words highlighted, there is obviously a high number of words in there that are pretty irrelevant like have and able and ask; you know they don't represent any sort of concepts that are important for this text so I think that if you were to develop an algorithm to do this better, you would need to use some NLP techniques that are little more examed that what this uses where you consider the frequency of terms and prefer terms that are less frequent because these are more likely to be semantically (…).

Let me look at your ranking again. So, in this case *NCIT* is actually a pretty good resource here because this text is about quitting smoking. Smoking is a common cause of cancers, so concepts that are represented here are probably all very core terms for *NCIT* and I think having it ranked number two, as your ranking has it, is probably pretty good. *MESH* you have it third, BioPortal had it second; *MESH* is fine, it is designed for all kinds of medical tasks. *RXNORM*, there are couple of drugs mentioned in the text and so for that reason, it is probably pretty relevant.

Question: **BioPortal had ranked it 10th. what do you think about it? Do you think it is better ranked higher or at the end?**

Answer: I don't know, I mean there are a lot of different concepts that are mentioned in this text and medications which is the only thing that's in *RXNORM* are I suppose only one part of what is being mentioned there. So, it is highly relevant for some of the keywords in my text but not for others. So, for that reason, I think having it in the middle of the list is probably appropriate. I don't know if I would have arbitrary ranked these myself manually, maybe I would put it somewhere in between 4 and 10.

Question: **Would you bring LOINC higher or the other ones?**

Answer: LOINC is hard for me that it would, it is hard for me to see how *LIONC* could be relevant here. *LIONC* is primarily about labs and other clinical observations and there is nothing on this text about that at all so I would, for this content, I would rank *LIONC* very low.

Question: **But you see it was rank third in BioPortal!**

Answer: Yes, that is not appropriate. I would put it at the bottom of this list, maybe not at the bottom I don't even know what *ChEAR* and *CHRISP* are! I don't know what, I am not familiar with *NIF* standard ontology but if it is about neuroscience, then it is not very applicable for this text at all because, again, this is a consumer text. *MEDLINE*, I think actually would be quite useful, because that *MEDLINEPLUS* describes a lot of consumers facing (...) So I probably rank it higher. I would probably rank it like number 5 or something. *RCD*, I don't even know what that is *CHEAR* and *CHRST*; I am not familiar with. So, there is only 5 of this that I think (...) not really applicable at all for this text and those that would be *LOINC* and *RDF* and *NIF* and *CHEAR* and *CHRIP*. it could be just a gap in my knowledge, but I wouldn't consider using those ontologies for classifying a text like this.

Question: **So if you want to compare the new ranking with BioPortal ranking, the ones that find less useful are ranked lower in my ranking compared to BioPortal.**

Answer: Yes, I think if we look overall, then your ranking is more useful.

Question:	**This ranking is based on a set of metrics. I want to know if you use any of them while evaluating an ontology for reuse and how important do you think they are. The first one is availability of documentation and Metadata.**
Answer:	I think that is extremely important. If something, if a published ontology didn't have fair documentation, it would be pretty useless to me because I would not feel like I was able to use it in the intended way.
Question:	**Same with metadata?**
Answer:	Give me an example, what kind of metadata?
Question:	**Like version, different kind of information that is there about ontology?**
Answer:	It depends on what metadata. Who has developed is certainly relevant. I do think it is important what version it's on, I mean I suppose that could be relevant if we are talking about like you know is this something new, a very new ontology, or has it been around for a long time. So, I guess those things can be important.
Question:	**And if the ontology is based on any methodology or is following any standard like OBO Foundry?**
Answer:	Yes, those are important, because just the domain of an ontology is not enough to know whether it is usable for the project I am working on; also need to know the approach they have taken with it.
Question:	**Size and the number of classes?**
Answer:	Number of classes? I mean I suppose that is helpful information to know. I mean you can, [if you make the decision to you know download] something or start working with it you are going to get a sense for that, but it is nice to have that.
Question:	**What about language?**
Answer:	Oh yes that is very important.
Question:	**Do you want to know if the ontology has been reused in any other project?**
Answer:	Yes, I have used that in BioPortal before, gives definitely a good sense of whether something has been well adopted within the community or not and whether the type of organisations that are using it have goals that are aligned with what I am trying to do.
Question:	**What about why the ontology was reused? Purpose?**
Answer:	Yes, that would be nice to know; I do not think that information is in BioPortal right?
Question:	**No, it is not. What about popularity of an ontology? how do you define popularity?**
Answer:	I would just define it as the number of projects that are using it or the number of people who regularly use the ontology for their work. I would say it is somewhat important, but of course if an ontology is designed for a very specific purpose, then it might not be broadly used but that does not necessary mean that it is not a good piece of work and that it wouldn't be useful for something that I am doing. So, it is one factor to take into account but not the only one.
Question:	**What about reputation or popularity of the ontology developer team or organisation?**
Answer:	Yes, I mean that is got to matter a little bit, I suppose. I will be more likely to trust something that was developed by a group that I heard of compared to one that I have never heard of. But I suppose even more important it is whether it is made standard for something like you know, SNOMED, being a sort of international WHO standard for example
Question:	**What about maintenance and availability of contact information?**
Answer:	Yes, both of those [are] super important. In terms of contact information, I basically never reused a model directly from BioPortal without you know trying to see if it exists, if I can get it directly from the source; like if there is some organisation that maintain the ontology on their website, I rather get it there, because I know that I am getting the most recent version and I know that that's gonna have the contact information that's gonna lead me to the right people. So, I would say these are things that are important.
Question:	**Frequency of updates, how important that one is?**
Answer:	I do not care whether if it is updated every three months or every 6 months, but if it is, if something is just put up there and then never updated that is a big red flag for me. I would think that that's you know nobody gets everything perfect on the first try; so, if you putting something out there for the world to see it and you never update it, that would be very concerning to me, [if I was thinking about ...?] partner with those developers to use that model in my work.
Question:	**So, if an ontology is updated in 2016, would you consider it?**
Answer:	Yes, if it has previously been updated before that and they stopped updating in 2016, yes, I mean I consider using it. I guess I would want to figure out the reason why they have not been updated since then and then kind of go from there.
Question:	**Which one of this metrics is more important for you?**

Answer: The most important factor is domain coverage; does it contain terms that represent the appropriate subject matter.

Question: **Is there anything else that you would consider after that?**

Answer: I guess after that would be the methodological approach of it, like that would include both the formal language that it is constructed in and also the I guess the granularity of it like how general or how specific concepts and classes are.

Question: **You know in literature, when talking about methodology they mean the set of steps that are taken when building an ontology**

Answer: Oh, I guess I wasn't using the word in the same way

Question: **Do you think metrics in the evaluating process should be weighted differently?**

Answer: If you are coming up with an algorithm to do it automatically? Yes, I do think that they should be weighted differently, but I am not sure how you would know for a given use how to, obviously you have to make some universal decisions about what is most important. But probably it is safe to say that universally, domain coverage is the most important thing. I think that is probably that's BioPortal current rules; I suspect that's what they prioritize as well but as I say they have room to improve their ranking algorithm by ignoring certain very common words.

Interviewee F

Interviewee F—Query used in the experiment

Query was a part of the interviewee's research and is removed for confidentiality reasons.

Interviewee F—Ranking comparison

Table K-6 Ranking Comparison for Interviewee F's Query

Ontology Name	New Rank	BioPortal Rank	Comments
GO	1	10	The most well-known ontology. Has been reused 63 times. Covers all the metrics that we are looking for.
SNOMED CT	2	4	One of the other well-known ontologies/terminologies. Has been reused 23 times
National Cancer Institute Thesaurus	3	1	Well known, is ranked third because it has only been reused 3 times
Logical Observation Identifier Names and Codes	4	7	Has been reused less than MESH, but has wiki and a slightly higher acceptance score compared to MESH
Medical Subject Headings	5	3	Is ranked higher than the ones below it because the number of times it has been reused and the good acceptance score.
Read Codes, Clinical Terms Version 3 (CTV3) (RCD)	6	9	Has been reused less than NIFSD but has documentation, is funded by NHS, has a higher acceptance score.
Neuroscience Information Framework (NIF) Standard Ontology	7	5	Has a lower acceptance score compared to MESH. Does not have external documentation
EDAM-BIOIMAGING	8	2	Is very smaller than NIFSD but has only been reused once. Has Wikis and website and documentation
CRISP	9	6	Very similar to the one below it because it has never been reused but has a very high acceptance score
Interlinking Ontology for Biological Concepts (IOBC)	10	8	It is new and under development. Has never been reused and the number of classes is not clear.

Interviewee F—Transcript

Interviewee: I got a bit different order in recommender because I have configured the weights differently which is actually a nice thing. I find that their metrics, these four things, that they have or actually three things plus the coverage they are very limited but at least what is good is that they user can change the way it is ordered and [unclear]. So, I think that's a very good thing. So, if you ever made an application, that would be a nice feature. Because in some cases I may not be interested in some of those things whereas in other I may like for example, this acceptance that takes how much it is used on BioPortal but it also takes into account whether it is a part of the UMLS, but if I am doing something that is not medical then it is irrelevant whether it is UMLS or not. So, also, when I pick this input as text, I notice that it is better because when I put keywords, then it only matching the whole thing between the commas but for example here is something I added as a test so then it also matches part of the keyword which is nice so it is much better coverage.

Question: **Do you think different weights should be assigned to different metrics in a selection system?**

Answer: Yes definitely. I think this is very good that the users can set it. I suppose maybe I can also order it by a given metric so that is very nice.

Question: **So, you have said the GO first**?

Answer: Yes, because GO has been reused the most and had everything that I was looking for. Yes, I think this is excellent. So, I would really pick the things which are important in the go context; so, it is good. The second one SNOMED CT; yes, it is a kind of mix of everything. Most of these big medical ontologies [unclear] so yes, I am not sure if that is useful for reuse or not. These medical ones maybe, but then which one, because they also big, I would say definitely in all the projects that I am working on then the size of the ontology is an important metric. But would say the smaller it is the better not the opposite.

Question: **Ok and that's how I rank the ontologies; the smaller ones are ranked higher**

Answer: Yes, I think that is very good. So that is in NCBO, they are trying to have specialization score. That feature caters for it a bit.

Question: **Yes, that is the same for them; the smaller ontology gets the higher score. So, the next one is NCIT. NCIT that is the first in BioPortal ranking**

Answer: Yes, but it is also like a dictionary of almost everything so I suppose this can be very useful in some text mining context, may be, but if you use it for the database when we want to somehow categorise the things then it is definitely kind of good coverage but in the same time small as possible the ontology probably the best thing.

Question: **So next was [ontology name]. It is a bit similar to SNOMED; then MESH.**

Answer: Interesting but it is also a mix of everything but at least not so, not really all the things. So, this RCD, they are very similar to each other these medical things. NCIT and SNOMED they cover almost everything or not almost everything, so many things and they overlap a lot and then also the MESH, RCD and LOINC, they also overlap. NIFSD; ok this is interesting this is a lot of overlap by the EDAM ontology. That is interesting. CRISP but this the specialization score in BioPortal that is something weird still because the NCIT has 58 and then GO has 2.2 so it is not correlated with the size. It somehow takes the size into account but not only.

Question: **Maybe because GO is larger?**

Answer: NCIT is bigger. Their specialization score is some kind of a weird metric that [unclear]. Yes, ok it is more how deep in the hierarchy it is the match. Is that relevant, I am not sure it is working to some extent, but I would say not very well. So certainly, your metrics can do much better than this.

Question: **So which ranking do you prefer?**

Answer: Well I would say just like from the little I have seen and I know about this query, because I have made it, so I would say that the first one should be the EDAM ontology but that is I mean you said it is an outlier of course in your case because it is so small, but it is also under development so this will a part of the whole EDAM probably next year so all these score will be calculated differently. So that would in this, that would be the first one for me and then the GO as the second I think it is good; those two I find most relevant for what

Question: **GO was ranked 10 in BioPortal.**

Answer: Yes, that is because it didn't have so many hits, but the important thing is that it had kind of many of those important hit's that EDAM did not have so kind of it covers what is left out. And that might be what they do when I choose an ontology set. But it didn't not work for me somehow. So, I think that is also an important for me and in the project that I am involved and I think that is the important measure is whether the ontology together with the other ontologies that we would like to reuse a kind of complement each other in a good way so we would always rather use a small set of small ontologies. But together they cover the keywords that we need. It doesn't have to be super small ontologies but a kind of reasonably specialized not the very big medical ones that have kind of almost random collection of concepts.

Question: **You know there is a thing with BioPortal and that's the fact that NCIT is the first ontology recommended most of the time.**

Answer: I think that is because it is one of the things that is so big and covers almost everything so it always has big coverage [unclear] for queries that I tried and at the same time it has a very big specialization score it is weird but maybe it is because how is the hierarchy? maybe it is a very flat hierarchy? that could be the reason.

Interviewee G

Interviewee G—Query used in the experiment

> ***Query was the abstract of a paper the interviewee had recently published and is removed for confidentiality reasons.***

Interviewee G—Ranking comparison

Table K-7 Ranking Comparison for Interviewee G's Query

Ontology Name	New Rank	BioPortal Rank	Comments
SNOMED CT	**1**	**2**	One of the most well-known ontologies/terminologies. Has been reused 23 times.
National Cancer Institute Thesaurus	**2**	**1**	One of the other well-known ontologies; has been developed in OWL, reused 16 times and is a part of OBO Foundry. Is frequently updated and has all the extra information (e.g. documentation) that this PhD looks for.
Logical Observation Identifier Names and Codes	**3**	**5**	Has been reused less than MESH, but has wiki and a slightly higher acceptance score compared to MESH.
Medical Subject Headings	**4**	**3**	Is ranked higher than the ones below it because the number of times it has been reused and the good acceptance score.
Read Codes, Clinical Terms Version 3 (CTV3) (RCD)	**5**	**4**	Has been reused less than NIFSD but has documentation, is funded by NHS, has a higher acceptance score.
Neuroscience Information Framework (NIF) Standard Ontology	**6**	**8**	Has a lower acceptance score compared to RCD. Does not have external documentation
CRISP	**7**	**6**	It has never been reused but has a very high acceptance score. Is developed and funded by US gov.
Interlinking Ontology for Biological Concepts (IOBC)	**8**	**9**	This ontology is new and under development. Has never been reused and the number of classes is not clear. But it is getting updated very frequently and has a team working on it.
ONTOAD	**9**	**7**	Has never been reused and was last updated in 2013! Is ranked higher than the one below it because it has publication.
suicideonto	**10**	**10**	Very small ontology that has been reused once. Was last updated in 2013!!

Interviewee G—Transcript

Question: **Starting from the ranking, what do you think about the new ranking?**

Answer: Yes, I mean there is not that much difference, right?

Question: **Yes, some of them, not for all of them. Many of them are similar but I have moved some of them around.**

Answer: So yes, the main thing is that actually none of these ontologies is really suited for it, so the one that would have been most close would be the basic formal ontology. I do not know whether that one is in the BioPortal.

Question: **Aha ok. I do not see it in the first 10 ontologies; you know, it is not even in the list. It is not in the first 22 ontologies, so it has not been recommended at all. Ok if you want to compare these ontologies, query independently, without a query, do you think that the ranking is good? do you think SNOMED should be ranked higher than NCIT? and then higher than LOINC and MESH? how do you think this ranking is?**

Answer: So yes, SNOMED CT is the only one which I could accept as being an ontology and as being any quality, so I don't know whether you know my history, but I have written papers about the NCIT and how bad it is, about MESH and how bad it is, and I compare a couple of other ontologies, and my overall impression of the BioPortal itself is that it is a collection of s***, pardon the word, but as soon as something is in the BioPortal, my recommendation would be don't use it.

Comment: **Ok that's interesting.**

Answer: But that has nothing to do, of course, with your criteria. but there are couple of perspectives. So, you say SNOMED CT has been reused 23 times but that is based on your information in the BioPortal, right? nobody gets it from the BioPortal.

Question: **Yes, that is one of the problems with the BioPortal; acceptance and number of reuse and all that. But compared to the other ontologies in BioPortal, like SNOMED compared to the other ones, having all this information, maybe SNOMED is better than NCIT.**

Answer: Oh, it is way better but the point that I am trying to make is that if you use reuse from within the BioPortal as a criterion, so they are missing all those uses that are made of those systems outside the BioPortal. SNOMED CT has it its own distribution, NCIT has its own distribution through the NCI, LIONC the same thing. I get all of my ontologies directly through the national library of medicine which are in the UMLS; I do not use the UMLS, but I use the source terminologies so that's a little bit of a pity that you have used BioPortal. But I understand it that you get all the information together, right?

Question: **Yes because of the ranking and the four set of information that they provide for each ontology and I would not get them from anywhere else and that's why I had to use them. Why do you think NCIT is not good? for all the, most of the, queries used in my experiment, NCIT was the first recommended ontology.**

Answer: That is because people are not aware of the bad quality. So, there are very few people who are aware how bad they actually are, and they use it because they are only interested in a couple of terms; they use it for a bit of annotation and that's it. So usually they won't do any reasoning on the basis of those ontologies and so if they do not do that, they will never experience the problems with it.

Question: **So how would you define ontology quality? because that is the main topic of my PhD?**

Answer: Oh, you should look in the paper that I wrote in 2006 [the rest of this answer has been removed for confidentiality reasons].

Question: **The set of metrics that I am using, most of them are query dependent and they are more about the social aspects and the metadata that can be provided about an ontology and how they might affect the quality. The first one is availability of documentation, availability of external documentation. How important or useful do you think it is for an ontology to have external documentation when you want to evaluate it for reuse?**

Answer: Oh of course important, are you aware of the OBO Foundry.

Question: **Yes, I am, OBO is one of the other things that I check; when I was re ranking ontologies, I took a look at OBO to see if the ontologies are a part of OBO. I used it as a kind of standard.**

Answer: No no, within OBO you have the OBO Foundry. It is only those within the OBO Foundry that are checked.

Comment: **Yes, that is what I have checked.**

Answer: And it's those principles that are important and documentation is one of the requirements for being accepted in the OBO Foundry.

Question: So, you think documentation is important?

Answer: Yes of course.

Question: The next question is about being a part of a standard or a bigger body standard e.g., W3C or OBO. How much do you care about that one?

Answer: Not at all.

Question: But doesn't it show quality, because they follow some principles.

Answer: Yes, but for instance, the OBO Foundry principles are [unclear] besides the OBO Foundry does not say anything about quality. They specify only a couple of criteria to make the ontologies reusable. So, if you have a very bad ontology and you document it, OBO Foundry takes it off. If you have a very bad ontology and you make it publicly available, OBO Foundry says good so that isn't anything about quality. So, what is important for me is that you can objective quality analysis have been done or that it is possible to do an objective quality analysis. So, that's why I say that documentation is important because on the basis of documentation, you can do a quality analysis to a certain extend.

Question: Ok, but following something like methodology, following a set of steps or principles, doesn't it affect the quality of an ontology? the way an ontology is built?

Answer: It depends on the criteria. So, most ontologies are concept-based right and [person's name] has shown that a concept paradigm is [fault?] so as soon an ontology is concept based, I ignore it.

Question: What about the size of an ontology?

Answer: Doesn't matter

Question: The language that it has been built it?

Answer: You mean the formal language?

Comment: Yes, like OWL

Answer: Yes, so if it is built in OWL, I ignore it, because OWL is not expressive enough. If you follow the realism based principles, where an important distinction is between a current and continuant, and these are the two words which are in my abstract, but it wasn't [unclear] by any of those ontologies, the important distinction there is that whenever, for an individual, some relationship is made with respect to a continuant, you require time indexing. OWL cannot deal with time indexing, OWL work only with triples. But you must [relationship?] which a is related to b during time period; in OWL you cannot do it. So, you never can have a good representation in OWL.

Question: Ok so if it is in OWL, you won't use it?

Answer: No, even the formal ontology, so when I use the basic formal ontology, then it is the way how it is originally being designed and there are versions of the basic formal ontology in OWL; I consider them wrong.

Question: What about having some information about the other projects that have reused the ontology?

Answer: That might give me some idea about in what domain it is useful to use it, but again it doesn't say anything about the quality.

Question: What about the reusability? like if you want to reuse it. If you have two, three ontologies that you know have the same quality, but you want to choose which one you want to reuse. If you have a set of ontologies, all with similar level of quality and you want to pick one to reuse, will this information help you, will you take a look at that information and say ok I reuse it because it has been reused in that project. So, I will trust it?

Answer: No.

Question: What about the other people, do you want to have information about the other people or organisations that have reused an ontology?

Answer: No, I would like to have information about the people who have made the ontology.

Question: What about the purpose that an ontology was reused?

Answer: Yes

Question: Next is more social aspect; popularity of an ontology. I want to know how you define popularity, what do you think a popular ontology is and how important do you think it is?

Answer: I think it is number of times that it has been used in one or other way or written about and so on and my [unclear] is the more popular ontology, the worse it is.

Question: [Ontology name] is very popular; you don't like [ontology name]?

Answer: [Ontology name] is the cream of the c***. So, it is the only one that worth analysing, not the ontology, I think the [ontology name] is [unclear] and of course ontologies that are direct decedents of BFO they are worthwhile. SNOMED CT is an old-style ontology, but they are doing a very good job in trying to make it better. The problem is so many people, actually [removed for confidentiality reasons], have their saying in the entire process that it takes a long while. So over time, I see a positive evolution but with every new version, I see, I mean so stupid mistakes that have been made and even changes in the model. So, actually I am preparing a paper for [journal name], which is about the last version of [ontology name], and they change their model for [domain name] products. So, they changed the way in which they model products and drugs and so, and they violate their editorial principles by doing so. It is amazing.

Question: **Ok what about the reputation and popularity of the ontology developer team or organisation? does it matter?**

Answer: The more popular, the more caution I would be!

Question: **What about maintenance and contact information? when you see an ontology, do you want it to be maintained?**

Answer: Yes

Question: **Frequency of updates?**

Answer: That depends on []. So, the larger it is, the more frequent it should be updated. So, imagine an ontology is really perfect, then it should never be updated. So, the amount of time that is updated, is rather a lack of quality than a criterion of a quality.

Question: **So how many times do you think is normal for an ontology to get updated? because when I used to do my interviews, some used to say that no ontology is perfect and if something wants to stay relevant, then it should be updated so how frequently should the ontology get updated?**

Answer: Let me rephrase the question and specifically going back on the way that the other person phrased it. So, an ontology describes or should describe what is generic in reality, right? it describes types, how often do you think that types change in reality?

Question: **It doesn't change, it depends on the type of ontology**

Individuals change, but you do not represent them in an ontology. So, even ontology is updated, so that for 99.9999% cases if you call something was missing which existed already, there are a few exception; I do not think that AIDS could have been an ontology 150 years ago or [unclear] flue maybe 50 years ago, that really did not exist, it is not that it was not discovered yet, it didn't exist. So only those new things, that requires actually updating an ontology that at a specific moment in time is perfect. I am giving you a totally [unclear] look.

Question: **How do you search for the ontologies that you want to reuse? do you know them?**

Answer: So, I do not use any; I only evaluate ontologies. That has been my research agenda over the past [number of] years so why I look at originally terminologies and then it is in the mid 19s that suddenly that term ontology became fancy, right? and people started to call terminologies, ontologies which I find it already a problem. So, I have been studying that and how to optimally use them in information system like [names]. So, that is what I have been doing. So, using them no, but evaluating them, see how good their model is, what are their update policies, how many can you track over time, what the changes are in the form of way, so those kinds of things.

Question: **So, what is the most important metric or criteria for you when evaluating an ontology?**

Answer: Whether you can generate on the basis of the changes that they describe. Ok what kind of mistakes have been made in the past. So, for instance suppose that on ontology is used for annotation in electronic health caring, ok, and the ontology in the next version obsoleted certain terms, ok? what are you going to do with those annotations that were made with the previous version, if the new version is not there anymore. There should be a mechanism that they say oh and, for instance, [ontology name] does that. For instance, they say that concept that we obsoleted it, because we discovered that it was ambiguous, and we created two new concepts and they [unclear] outside of the ontology as a part of the documentation but in a formal way. They specify that the concept that the concept that is deleted is now replaced by two new ones. So, it is useful information because whenever you have a new version, you can check what is deleted, in electronic health records and then you can reannotated it by which of the two alternatives. So, that is a very concrete example and there are many other examples. So that I find the most important aspect for the ontology evolution over time.

Question: **So, a kind of transparency, what is happening, why is it happening?**

Answer: Yes, so the why and even in a formal way not just reading in a documentation. So that example for [ontology name] is for instance one of the reasons why they give is [unclear]. That's not follow the editorial conventions, ok, and then they change maybe the term of the concept, but they keep the concept active. Now in the last version, I noticed that there were 270 concepts for which the original terms which was of the form medicinal products containing exactly X were changed into medicinal product containing at least X. That's wrong, so they did not change the concept. They say in the new version, it is still the same thing, it denotes the same thing, the term is different so people who selected that concept in the previous version in the understanding, that it has to be only that drug, they are put on the wrong foot as [ontology name] itself does not give any indication there that the meaning actually changed that is a bad example.

Question: **My metrics are query independent. Do you think there is any place for query independent reuse metrics?**

Answer: I never thought about it. So, I mean if I wouldn't have found it interesting, I would not have responded to your request. But I do not know yet, what the impact is going to be. So, you become famous about this work, I do not dare to make a prediction.

Question: **I have also been thinking to use the term non-ontological aspects of ontologies. I am thinking that not only the internal characteristics of ontologies matter, but how they have been reused also matters.**

Answer: Yes, I agree. Are you familiar with the FAIR principle? it is another initiative; it is like OBO Foundry. I do not agree with all the principles that they put forward and they do not go far enough, but nevertheless it is an initiative that you should not overlook in your thesis.

Question: **How useful do you think this set of principles are? if the conclusion of my PhD is a set of principles or set of steps or a set of information that people better provide if they want their ontology to be reusable, how useful do you think that is?**

Answer: I think in principle that is a good thing, it depends how good your principles are.

Question: **How important do you think that there is publication about an ontology?**

Answer: Oh, very important.

Interviewee H

Interviewee H—Query used in the experiment

Measurements, units, traits, phenotypes, disease

Interviewee H—Ranking comparison

Table K-8 Ranking Comparison for Interviewee H's Query

Ontology Name	New Rank	BioPortal Rank	Comments
SNOMED CT	1	1	
Logical Observation Identifier Names and Codes	4	2	Very similar to MESH; however, has been only reused 8 times (compared to 15 times that mesh has been reused)
Medical Subject Headings	3	3	
National Cancer Institute Thesaurus	2	4	OBO Foundry, smaller than the other 2, has been reused more, is updated monthly, is built in OWL
Gene Expression Ontology (GEXO)	7	5	**GEXO, REXO and RETO** are all a part of a bigger project; all have been active in 2015 and haven't been updated since then. The only factor that was different among them was the acceptance score; so, they are ranked by their acceptance score
REXO	9	6	
Neuroscience Information Framework (NIF) Standard Ontology	6	7	Has been reused 5 times and was updated in 2018 so it goes higher than the other ontologies
Regulation of Transcription Ontology (RETO)	8	8	
ONTOAD	10	9	Last time updated in 2013; low acceptance score
Read Codes, Clinical Terms Version 3 (CTV3) (RCD)	5	10	Has been reused twice before; has been updated in 2018; has a very high acceptance score; is developed and funded by NHS; has external documentation

Interviewee H—Transcript

Question: **What do you think about the new ranking?**
Answer: So, I, I gave some fairly high-level terms, I think there is, which might not typically be how someone would use this, right? because I said I want an ontology about disease, measurements, units, trades, phono types. They were fairly high-level things which I think is, and both sets of rankings of ontologies that cover those things, now you know you rank the NCI thesaurus higher, but I did not mention cancer, and that is an ontology specifically developed for cancer studies.
Question: **So, it is not good for your query, if you had the list of ontologies in front of you, and you wanted to pick one, which one would you pick?**
Answer: Well, if I was interested in disease and it was clinical data, I would go for SNOMED. If I was interested in like laboratory measurements, then I would go with LOINC, I would never pick MESH. NCIT is interesting because it has a very good representation of cancer, but it also covers lots of other things. Not only did they just define cancer types, but they also define diseases and anatomy, laboratory instruments. So, given the keywords that I send to you, I understand why NCIT came up second because of all those things but at the same time.
Question: **In BioPortal it was fourth, do you think it better be fourth or second?**
Answer: I don’t know. So, I think it is interesting what you've done, and I think it is interesting to try and take these extra information about, try to assess some sort of level of quality of these ontologies to change the ranking but I think the thing that is missing is what am I trying to do? so I have given you set of terms but it does not tell you what I am trying to, the context within which I want these ontologies, right? I can make up some context, I can think of different scenarios that would change the ranking of these ontologies. If I was, if I really cared about having full logical OWL axioms and I wanted to do something, you know some smart semantics with the ontology and integrating it with other data, then I really would not want to use MESH cause I know that MESH has no semantics, right? it is not really an ontology.
Question: **One question, as far as I know, none of the search and selection systems for ontologies, they do not allow you to choose the purpose or context, they do not ask you why do you find an ontology?**
Answer: No, exactly, and this is the problem. You know the other thing about SNOMED CT, actually I would not use SNOMED CT ever because we are in academic institute working with public open data and SNOMED CT has a license. So the license restriction on SNOMED CT is like if you have public open source data, then you cannot use SNOMED CT, and that is not reflected here. So while SNOMED CT is a very good clinical terminology, it comes with a license and a cost, so if you are a student and you are trying to use an ontology, then you probably shouldn’t be using SNOMED CT because you probably do not have the license to use it. And BioPortal is very bad at showing you that, it does not make it clear that you have to have a license to use this. So, you have to have a license to use it through it, to access it from BioPortal. So, I think there is other factors as well.
Question: **Ok, so can I ask you about different metrics I have used and how important you think they are? how important do you think the following metrics are when selecting an ontology for reuse? having access to external documentation?**
Answer: Very important.
Question: **Metadata, you mentioned license information, is it important?**
Answer: Yes, that is important.
Question: **What about being a part of a standard or having followed some principles? e.g., OBO Foundry?**
Answer: Quite important, I mean the reason that I say that is we do use ontologies that are not as a part of OBO Foundry. We prefer that ontologies are part of OBO Foundry. It makes things easier but there are also some very good ontologies being built in the life sciences that are not a part of OBO Foundry; one of which is [ontology name] which we built here. So, we are not a part of OBO Foundry. We try to follow their standards but there are other reasons why we are not a part of Foundry.
Question: **Ok, is there any standard that many ontologies follow, or methodology?**
Answer: No, I think that in the life sciences the closest we have is kind of what OBO suggests that are in using BFO and some of their standard annotation properties, but I do not think there is any well recognised standard, no.
Question: **What about the size of an ontology or the number of classes it has?**

Answer: Yes, I mean that is less important, that depends, depends what it is about. Obviously, if I need an ontology of disease and there are only 50 classes, then I will probably be a bit suspicious that is going to give me the coverage that I need. So, I am a kind of incurious about the size but again it depends what area and how big I expect it to be.

Question: What about the language that an ontology has been built in?

Answer: I would expect it to be in OWL. So, it is important that it is in OWL.

Question: Someone told me that OWL is not expressive enough, what do you think about it?

Answer: That is true, the formalism that are more expressive, but I think you have also got to kind of weight expressivity with sort of scalability and how much you can actually compute with these languages. So yes, I could take some other first order language, but I might not be able to find anything that can make the scale for the ontology sizes that we work with. So at least with OWL it is not just about the expressivity. It is the fact that it is a well-known standard there are tools that I can use with OWL, there are reasoners, there are [unclear] reasoners that scale very well. So, it is not just about how much can I say in the language. There are other benefits to using a language like OWL because you've got things like Protege and you have got the OWL API. So, it is not just how expressive it is, and I am happy to give up a bit of expressivity in order to get access to tool that works.

Question: So, next is about the extra information about an ontology. Do you want to know what are the other projects that an ontology has been reused in?

Answer: Yes, we care about that. Typically for u,s we are looking to use the ontologies that describe our data here. If we want to use someone else’s ontology, then our assumption is that ontology is being used somewhere else to describe similar data so that means that we can integrate with our data.

Question: Is it more about why an ontology was reused before? or does it matter why an ontology was reused before?

Answer: It might matter, it depends. I think if you have a good ontology that just describes a domain and it is not too tight to a particular application, then that is quite nice because it means that people can reuse and readopt the ontology for different purposes and that is quite true with a lot of OBO ontologies. But the problem, you have always got to strike the balance between the more generic you make an ontology, the less it fits the need of any particular use case that you have. So what happens is that people will build a fairly generic ontology that will claim to solve provide a general description of anatomy or disease and then when you come along and try to use it, what will happen is that you often end up having to say well actually I need to change this. Because you know in my domain we call things slightly differently and we need, you usually need to extend it and the real thing that you see happening more often is that if someone else's ontology doesn’t quite fit what you are trying to do, people will often go away and just write another one. This is why many [unclear]. Even though OWL is quite a nice language and you can extend it, does not seem that people really do that.

Question: Maybe because it is easier to build your own ontology?

Answer: Yes.

Question: Do you care to know who else has used it or what organisation; if your colleagues are using an ontology, will it be a thing for you to use it?

Answer: Well no I would not use it just because someone else is using it. I mean it all comes back down to if we need it. If we need, we have got data and we need an ontology for it, I will first go, not necessarily used it but who built it. So, I mean, there is certainly a reputation around who has built a particular ontology and that might influence. If I had to decide between two, and I knew some of the developers of one and that can mean two things, that can mean one that I trust them, and I think that they build good ontologies and the other is that I know them and I know that they will be willing to collaborate and we could work on extending it together, so it is good.

Question: So, it is important to know the people around an ontology?

Answer: I think it certainly helps in reality to be able to [unclear].

Question: What about maintainability and frequency of updates?

Answer: Yes, so we have a kind of done studies on this in the past. So, I think that again it is, so if the ontology is changing a lot, on the one hand it is a good sign because it is telling you someone is actively working on this ontology. It is also potentially a bad sign because it might mean that this ontology is changing and therefore it is unstable and, therefore, I should not be using it because so you have to understand the nature of what is being change; are they just altering labels all the time or are they actually adding new classes? are they deleting classes? are they moving things around the hierarchy? So, it is actually an ontology changing and being updated regularly is not necessarily a good thing; it depends on the type of changes.

Question:	**What type of changes do you prefer?**
Answer:	I preferably just like to see additions. You do not want to see stuff getting deleted. And, you do not really want to see big hierarchy rearrangements. That suggest that they have done something wrong or they are changing their modelling. But then there is a flip side that you sometimes, some ontologies that we use that have not changed a lot in the last 5 years and they never updated, but that is because they are almost finished you know they are good. The [name] ontology is quite a good example of something that has not needed to change a lot in the last 5 years, and it is very stable, and we all use it. Seeing change in an ontology does not really mean anything one way or the other; you need to understand the context of that change.
Question:	**How frequently do you think it is ok for an ontology to be updated? what do you think the normal rate of update is for an ontology?**
Answer:	I think a healthy project you would expect to see updates every month at least a couple of updates a month.
Question:	**Someone told me that if an ontology is good enough, it never needs to be changed. if they are adding things every month, doesn't it show that they did not get the whole thing the first time they created an ontology?**
Answer:	It depends which domain, so you know we work in science and our understanding of science changes every month, so our knowledge does change quite regularly. I think it might be other domains and areas where things can be fairly stable and understood, but if you look at something like the Gene ontology, our knowledge about what proteins do changes every month or not necessarily changes, but we are adding to it. This is what I mean about the types of, so you are not just adding knowledge to an ontology. That is good because if you are deleting classes or you are moving stuff in the hierarchy, then that suggests like you say, that what you understood before is not true or you are in a domain that is changing quite a lot. So, I think you would have to look at those, the types of changes over a pretty long time to understand whether this ontology is stable or whether this ontology is about an area that is unstable and there is a lot of change going on.
Question:	**The last one is about popularity; how would you define the popularity of an ontology and how important do you think it is?**
Answer:	It is not important, but I think popularity should be measured on usage, and I mean usage as in the data that I can point to that has used this ontology not, [unclear], and the problem with something like BioPortal with the usage is that what just people claim, it is not, there is no proof and it is not complete so they popularity or the usage, the popularity is quite useful in BioPortal, every time someone has visited that website, but the way that people report to say they use the ontology, I think that is dangerous and that is a sort of misleading.
Question:	**So, what is your favourite selection system?**
Answer:	We use our tools.
Question:	**How do you compare it with BioPortal?**
Answer:	So, we have the [the rest of this answer has been removed for confidentiality reasons].
Question:	**How do you compare open systems with closes systems and which one you think is better for ontologies? (a part of this question has been removed for confidentiality reasons)**
Answer:	So, [removed for confidentiality reasons] say you just created an ontology for fun, and you put it in there, and you finish your PhD and you never used it again. Because people do not know your problem that you are trying to solve that people do not know which ontology to pick which is the best one, what if they search for something and the best thing is your ontology, and then they annotate that data and they submit it back to us, the data and in the public archive then, there is this ID that is actually meaningless because it was a toy ontology. We started seeing this already that you see the data from, you know I would say these toy ontologies making it into the public archives. So, I think that you have to be, you need both, when people come here and people expect when they use [the tool name], that they should be confident they can use any of those ontologies. When you use something like BioPortal, it is not clear if you should be using all of those ontologies for annotating data.
Question:	**What is the most important for quality assessment?**
Answer:	I think in our scenario, we are usually interested in does this ontology describe my data. You know do the concepts. It is really the matter of coverage for us. So, we are looking for things that have coverage. So, if I am interested in a specific area like specific disease area, I want an ontology that gives me a good coverage for those concepts and I can always fix the semantics later but I want is actually coverage and that is not necessarily everyone's use case, but I think a lot of the time with what we are trying to do, we are working with very messy data that is not

map to any standard and we want to clean this data and align it to at least one standard. So, we usually care about coverage and then we will fix and work on the semantics later.

Question: **Coverage might be the first for so many people; is there anything else?**

Answer: If think the next thing is, can I get it in owl, and does it actually have any semantics behind it. So, it is thesaurus or is it a real an owl ontology?