A Survey to Identify Factors for Vocabulary Reuse and Requirements for Vocabulary Recommendation Tools

Johann Schaible

Knowledge Technologies for the Social Sciences
GESIS - Leibniz Institute for the Social Sciences
Unter Sachsenhausen 6-8
50667 Cologne
johann.schaible@gesis.org

Abstract: The choice of appropriate vocabularies is one essential aspect in the guidelines that guide a data engineer when modeling Linked Open Data (LOD). In general, this leads to an easier consumption of the data by LOD applications and users. However, making decisions considering the adequacy of various vocabularies is not straightforward and a well known challenge; the same applies to the engineer's decision-making regarding the total number of vocabularies used in one dataset. Therefore, it is not surprising that according to some LOD data provider studies, there is still an insufficient compliance towards this particular best practice. In this paper, we examine the current importance of the best practice "vocabulary reuse", as well as the factors that influence the engineer's decision whether to reuse a specific vocabulary or not. We provide results of an online survey comprising an aggregation of knowledge, practices, and design motivations of several LOD publishers and practitioners with respect to the reuse of vocabularies. These results show that the insufficient compliance considering vocabulary reuse is not because of its lack of importance, but most likely because of deficient tool support for deciding which and how many vocabularies to reuse. We address the increased need for such tool support, and based on the results of the study, we derive several requirements for future vocabulary recommendation tools.

1 Introduction

The Linked Open Data (LOD) cloud comprises data from diverse domains. To publish LOD, Bizer et al. and Heath et al. provided a set of Linked Data best practices [BCH08, HB11]. When modeling LOD, the data engineer has to—among many other tasks—decide which vocabularies to use. Hereby, he should rather reuse classes and properties from existing vocabularies than reinventing them. This generally leads to a dataset that has an increased interoperability, as various LOD tools have tailored support for widely used vocabularies.

The empirical analysis by Hogan et al. [HUH+12] examines 4 million RDF/XML documents on their compliance to several best practices. Hereby, the compliance of data providers with respect to reusing vocabularies is described as "non-trivial", which is quite insufficient. Obviously, there are some specific factors that have major influence on the

engineer's decision whether to reuse vocabularies or not. Is vocabulary reuse is no longer considered to be as important as it used to be or is the procedure simply getting more complex? The main reasons for such a complexity is that the number of reusable vocabularies is constantly growing. This results in an deficient overview and knowledge over all vocabularies. In this paper, we examine the influencing factors and point out that vocabulary reuse is still considered to be a very important best practice; the insufficient compliance to vocabulary reuse is more likely to be based on the lack of proper tools that support the engineer in his modeling decisions.

To our best knowledge, there is no study so far that examines such factors. To aggregate the knowledge, practices, design motivations, and experience from LOD practitioners and experts, we address the LOD community to participate in a survey regarding the reuse of vocabularies when modeling Linked Open Data. The survey comprises questions about the participant's viewpoint on the reuse of vocabularies and his/her reuse strategies, as well as two ranking tasks of several example schemata implementing different reuse strategies. In detail, it covers the following aspects: (i) how important do leading LOD researchers consider the reuse of vocabularies in comparison to other best practices, (ii) what are the specific factors that influence the participant's decision whether to reuse vocabularies or not, (iii) what is the participant's general strategy to reuse vocabularies, and (iv) which example schemata do the participants consider best with respect to an appropriate vocabulary reuse?

We have obtained feedback from 23 participants acquired through the public LOD mailing list. We have performed quantitative measures to calculate the relevance of each answer in comparison to the other answers. Free comments of the participants provide further detailed information on the asked aspects. The results of our survey are of interest and valuable for developers of software tools to support the Linked Data engineers. In the latter case, the results provide requirements such a tool regarding the aspects *how should such a tool support the engineer* and *which reuse strategies should it comply to*. For example, the implementation of the novel LOVER [SGSS13] approach, which recommends classes and properties from existing vocabularies when modeling Linked Open Data, is most likely to benefit from these resulting requirements. LOVER is subject of the Ph.D. thesis, and the requirements resulting from this survey are considered crucial for its implementation.

The remainder of the paper is structured as follows. In Section 2, we present the survey and illustrate the questions. Section 3 shows the results of the survey. Hereby, we display quantitative measures such as the average value on a 5-point Likert scale or the percentage distribution of the answers. Furthermore, we discuss the participant's comments to identify further reasons for a particular percentage distribution. In Section 4, we describe the LOVER approach in more detail and explain which impact the results of the survey have on the implementation of such a vocabulary recommendation service. We conclude our work in Section 5 and discuss the future work.

2 Structure of the Survey

In order to prepare the survey, we have followed a two-step approach. First, we have conducted a pre-survey with a small group of participants with the goal to identify the most relevant questions and potential answers. Subsequently, the results of this pre-survey have been implemented in an online survey.

Pre-Survey: The pre-survey consisted of several questions in form of a qualitative interview; every question had to be answered in text form. Asking to provide answers in text form guaranteed the answers to come straight from the participant's experience and knowledge. Additionally, the participants were able to comment on each question. These comments helped us to decide which questions were the most important ones to ask the LOD community. They also improved the quality of the questions, so that they were as easy as possible to understand. For example, the question about how data engineers search for vocabularies was derived from comments of several participants. The answers to each question were analyzed according to the frequency of their occurrence. We gathered the most frequent answers and provided these as response options for the questions in the final online survey, such as the various reasons why an engineer should reuse vocabularies. Hereby, the comments and answers to the questions were manually collected and analyzed. The participants of the pre-survey were handpicked scientists with a high to an expert knowledge in the LOD context.

Final Online Survey: The survey[1] begins by presenting the participant a scenario, where one is supposed to publish some structured LOD dataset. First, the participant has to provide information on the domain of his data, his knowledge in LOD in general, and on the LOD publishing process. Further, each question of the survey is designed to gain information from every participant about his general procedure of publishing LOD with respect to the reuse of vocabularies. The survey covers four areas which we describe in more detail.

1) **Importance of reusing vocabularies**. This includes the importance of vocabulary reuse in general as well as in comparison to other best practices. We raised questions considering what the participants think would characterize "good" Linked Open Data and how they would rate various best practices on a five point Likert scale. Hereby, the participants were able to select multiple response options. This way, we were able to investigate whether the lack of importance of vocabulary reuse is the reason for the insufficient compliance of this best practice.

2) **Factors for reusing or not reusing vocabularies**. To investigate this aspect, we have provided three questions with several response options each. The first two questions covered the the participants' position on why they would or would not reuse existing vocabularies. In the third question, we asked the participants how they search for specific classes and properties from actively used vocabularies. Again, we provided a possibility to add further reasons as well as general comments regarding this aspect of the survey. This way, we intended to examine the benefits of vocab-

[1] `http://bitly.com/LODsurvey`, accessed: June 30th, 2013

ulary reuse, but even more importantly, the downsides of it, such as the time consuming search for appropriate vocabularies. Such results could exhibit that the lack of tool support for vocabulary reuse might be the reason for its insufficient compliance. Furthermore, the results could express the first requirements for a vocabulary recommendation approach, such as LOVER.

3) **Strategies for reusing vocabularies**. Every Linked Data engineer has his preferred strategy to reuse vocabularies when modeling Linked Open Data. We asked the participant to share his specific strategy by providing various response options, of which the participant was able to choose only a single one. Additionally, we asked the participant to name their personal viewpoint on the (dis-)advantages of minimizing and maximizing the number of different reused vocabularies. Once again, the results of this aspect could give insight into the requirements for a vocabulary recommendation system, as they provide information on how many vocabularies are considered appropriate for a LOD dataset that is most likely to achieve a high interoperability.

4) **Assessing Implicit Reuse-Strategies for LOD-Modeling**. Here, we provided two ranking tasks: one, that covers the domain of Social Sciences; and the other, that covers the music domain. In each task, the participant was given several example schemata that were each modeled according to different strategies of vocabulary reuse. These different example schemata had to be ranked from "best" to "worst" considering the participant's definition of good Linked Open Data. In addition to ranking the example schemata, each participant had to explain his decision. Goal of this task was to obtain knowledge on the participantsâ preferences regarding the different strategies for vocabulary reuse. This information allows for (i) a rank-based comparison of the different options and (ii) check for the overall consistency of the provided answers in the survey. All in all, the rankings provide the final aspect of a first set of requirements for a tool that supports an Linked Data engineer in reusing vocabularies by recommending appropriate classes and properties.

3 Results of the Survey

At the time of writing this paper, we counted a total of 23 participants who filled out our online survey. This included several Ph.D. students, research associates, system and software developers, as well as senior researchers all dealing with data from about nine different domains. The most prominent domains included data from life sciences and biology, data from the social sciences, as well as government data. Based on the 5-point Likert scale, most participants judged themselves as experts in Linked Open Data and highly experienced with the LOD publishing process. 20 participants (86%) stated that their knowledge on LOD is good or very good. Regarding the publishing process, 17 participants (73%) feel familiar to very familiar. Consequently, we obtained expert knowledge and experience from individuals who probably work with Linked Open Data on a daily basis in real life scenarios.

In the following, we will describe the results along the four aspects we have introduced in the previous section.

3.1 Importance of reusing vocabularies

Examining the importance of reusing vocabularies shows that this best practice is still considered to be very important when modeling Linked Open Data. In total, 20 participants (87%) consider vocabulary reuse to be a major characterization of *good* LOD. It is indeed the second most important best practice, after the response option "Publishing LOD according to the 5- Star Principles" (95%). Further answers are: "Producing a semantically correct dataset on schema level" (82%) and "Providing human-readable documentation" (78%). The participants' additions include "no or little use of blank nodes"and "ensure the availability of the data (dereferenceable data)". In addition, one participant with expert knowledge and experience in publishing LOD explained why he did not choose vocabulary reuse as a characterization of good LOD with "There are [...] vocabularies that are worth using and others are not". Thus, reusing some vocabulary does not necessarily mean that it is good choice.

A second question about rating various characterizations of good LOD underline the outcome of the first question. Figure 1 displays the average value of these characterizations on a 5-point Likert scale, where 1 corresponds to *absolutely not important* to 5 of *absolutely important*. A human-readable documentation is considered the most important characterization (M=4.1, SD=0.79). Again, reusing vocabularies is seen as the second most (M=3.95, SD=0.76) together with having a semantically correct schema representation. The other aspects have an average value from 2.4 to 3.95 and a standard deviation of about 1. In the free comments, it is mentioned that a reused vocabulary should have to be agreed on by a large community. One expert mentioned that a good LOD dataset should balance "correct but complicated formalism vs. less accurate but usable data", where the usability of a dataset is determined by how easy it is for client applications to process the data as well as for humans to understand the data.

3.2 Factors for reusing or not reusing vocabularies

For each of the questions in this aspect, the participants were able to select multiple response options. The results show that the reuse of vocabularies is acknowledged to be very advantageous in many cases, but there are also various reasons that lead to the decision not to comply to this best practice. Table 1 shows the distribution of all the provided factors that influence the engineer's decision whether to reuse vocabularies or not. About 83% of the participants acknowledged that vocabulary reuse is more likely to lead to better integration in the LOD cloud and hence will help the interlinking process, whereas 74% of the participants responded that this good practice makes the data easier to be consumed. Regarding the publishing process, 61% of the participants consider that reusing vocabular-

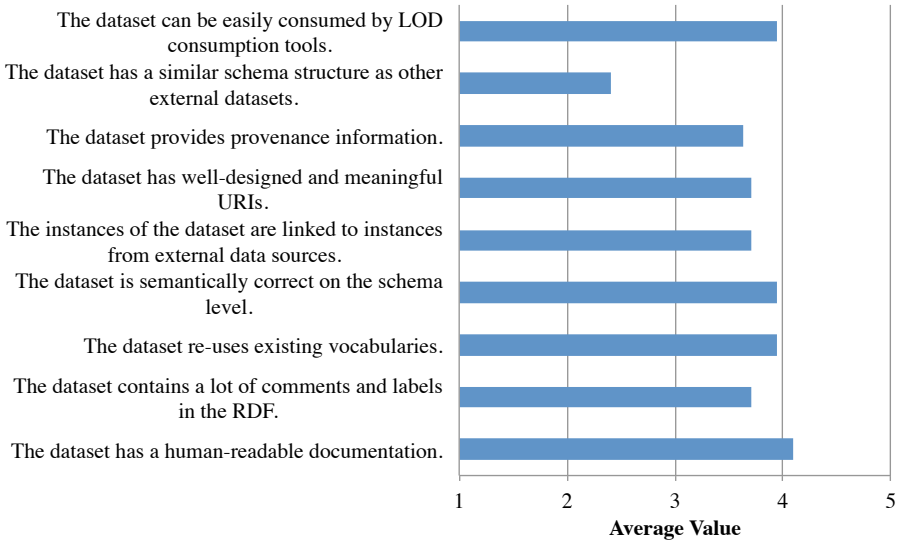

Figure 1: Rating of the characterizations for a good LOD dataset

ies makes this process easier and faster. Several participants—all with a high to an expert knowledge—commented that reusing ontologies establishes a mutual acceptance of a data description (*convention over configuration*), makes data easier to understand, as well as prevents reinvention. On the other hand, arguments for not reusing vocabularies included that a search for appropriate vocabularies is too time consuming (35%), or that there are no existing vocabularies for various domains (35%). For example, one participant stated that there is no existing vocabulary for the domain of "library holdings data". Further comments mentioned several other reasons why one would not reuse vocabularies. These include "Existing vocabularies might be of very bad quality [...]" or "it is obvious that the designer of the vocabulary did not have real-world experience". One interesting comment mentioned the `rdfs:domain` and the `rdfs:range` often to be show stoppers, as they force the engineer to reuse the exact same semantics of the vocabulary. However, the most interesting comments considered the visibility of existing vocabularies. It is claimed by five participants with high or expert knowledge that the reuse of vocabularies is lacking vocabulary search infrastructures. It is also considered important to have a hub for vocabularies and their description or documentation, since, according to the participants, this would enable finding the better choices. Another participant stated: "All those reasons for not reusing vocabularies will fall, if we have proper vocabulary services and a good global vocabulary governance by W3C etc.".

Regarding the search for appropriate vocabularies, we found out that 56% of the participants use the Linked Open Vocabulary index LOV[2] and 35% of the participants use

[2] `http://lov.okfn.org/dataset/lov/`, accessed: June 30th, 2013

Factors for reusing vocabularies	% of participants
It leads to better integration in the LOD cloud and hence helps the interlinking process.	0,826086957
It makes your data easier to be consumed.	0,739130435
It makes publishing Linked Open Data easier and faster.	0,608695652
There are consumption tools that have support for standard vocabularies.	0,608695652
Existing standard vocabularies are backed by a standards organization or a reputable community.	0,565217391
Existing standard vocabularies are being updated and maintained regularly.	0,347826087
It gives confidence that the modeling is sound.	0,347826087
It is an indicator for a high-quality dataset.	0,304347826
Factors for not reusing vocabularies	
Search for appropriate vocabularies takes too much time.	0,347826087
There is no vocabulary for my domain.	0,347826087
It does not improve the quality of my dataset.	0,173913043
It is not useful for my use case.	0,130434783

Table 1: Factors for reusing or not reusing vocabularies

Swoogle[3]. About 56% discuss the situation with colleagues or other experts, but the majority of 65% simply use Google. Some comments regarding the search for vocabularies stated the "Follow the nose" approach. This means the engineer starts at a specific entry point by reusing one vocabulary, and then refines the model with terms from the same namespace or searches for further ones.

3.3 Strategies for reusing vocabularies

We assumed that most Linked Data engineers would have different strategies to reuse vocabularies as various conformance studies such as [HUH+12, HHP+10] show that most LOD models representing similar data differ in very many ways. However, the outcome of examining this aspect proves otherwise. The clear favorite strategy to reuse vocabularies (75% of answers, where it was not possible to select multiple choices) is that one should reuse as many vocabularies as needed to best cover the domain under consideration and only then introduce own properties and classes (cf. [BCH08, HB11]). Only 4 participants consider the strategy to reuse only a few standard vocabularies as their preferred approach, whereas the others, such as reusing exactly one vocabulary and express the remaining entity types and attributes with a self-defined vocabulary, was not chosen at all.

Regarding the total number of different vocabularies within one dataset, there was a fair balance between the (dis-)advantages of maximizing or minimizing this number. Maximizing the number of different vocabularies is considered advantageous regarding the easier consumption by LOD tools, as they are more likely to have tailored support for

[3] `http://swoogle.umbc.edu/`, accessed: June 30th, 2013

well-known and widely used vocabularies. This could lead to an increase of interoperability of the data. However, maximizing is also more likely to result in a description and structure of the data that might become way more difficult to understand for humans. Furthermore, the data might comprise potential conflicts in its modeling, and loading all namespaces from the web when consuming the data might be too time-consuming. These arguments state that maximizing the number of vocabularies would decrease the interoperability of the dataset. On the other hand, minimizing this number is more probable to provide a clear structure within the data as it usually generates less complexity and thereby increases the readability and the interoperability of the entire dataset. However, the central disadvantage is that in most cases a minimal number of vocabularies cannot express the entire data. This might result in: (i) fitting several entity types into classes and properties that do not express the entire semantics; or (ii) the ontology engineer has to invent a lot more self-defined terms, which increases the probability to redefine an already existing vocabulary term. Both are consequences that might decrease the interoperability of the dataset.

3.4 Assessing Implicit Reuse-Strategies for LOD-Modeling

A total of 19 participants ranked four example schemata from the social sciences domain, and 16 participants ranked four example schemata from the music domain. The results are shown as percentage distribution in Figure 2. Regarding *Task 1*, Schema-I reuses a minimum number of vocabulary, whereas Schema-II maximizes this number. Schema-III reuses only standard vocabularies, and Schema-IV reuses a minimum number of vocabularies per concept. Hereby, each participant was able to rank multiple schemata the same way. The results for this task show that 89% chose the schema that reuses primarily popular and standard vocabularies as the best modeling. The average rank of the schema that reuses only popular and standard vocabularies is 1.16 with a standard deviation of 0.5, where 1 denotes the "best" and 4 denotes the "4th best". The other schemata have an average value that is severely below (M=2.63 - 2.98, SD=1). The results from *Task 2* underline the outcome of *Task 1*. The participants were given an initial schema and were asked to decide which final example schema is considered to be the best. Final Schema-I reuses a minimum number of vocabularies in total, whereas Final Schema-II reuses a minimum number of vocabularies per concept. Final Schema-III reuses only popular and standard vocabularies and Final Schema-IV reuses only vocabularies that have already been reused in the initial schema. Again, the strategy that reuses only popular and standard vocabularies was chosen as the best modeling by the majority of the participants (over 60%) (M=1.56, SD= 0.81). The other schemata have an average value that is about one point below (M=2.53 - 2.87, SD=1).

The participants were also asked to explain their decisions. Most explanations for choosing the schema that reuses popular and standard vocabularies included reasons such as clarity, simplicity, reusability, and understandability of the dataset. Furthermore, there were several explanations, that existing LOD tools have tailored support for vocabularies such as

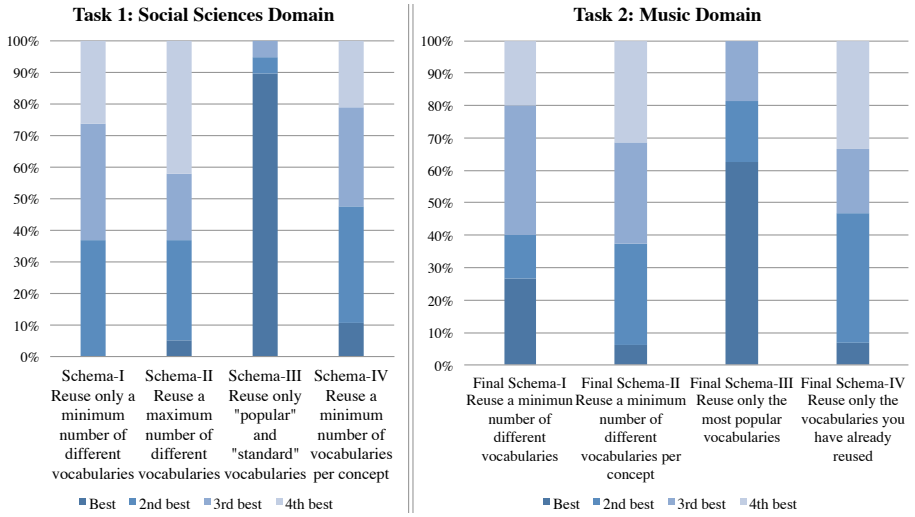

Figure 2: Ranking tasks of the different schemata

FOAF[4] or Dublin Core[5]. As mentioned before, this is considered to make the dataset to have an increased interoperability.

3.5 Discussion of the results

Regarding the first aspect, the results of this survey have shown that reusing vocabularies is considered a very important best practice when modeling Linked Open Data. Thus, it is not the reason why there is such an insufficient compliance towards this best practice. This goes along with the results from the second aspect, where it was exhibited that there are several factors influencing the data engineer in reusing vocabularies; the most important factor is the increase of interoperability. However, it is also shown that complying to this best practice is quite difficult. The data engineer is provided little to no tool support; he thereby ends up in a situation where he has to solely rely on his personal experience and gut feeling when it comes to the decision of vocabulary reuse. To alleviate this situation, the engineer should be provided tool support that recommends him classes and properties from vocabularies based on several aspect, such as "is it a widely-used vocabulary?", "is the semantic use of `rdfs:domain` and `rdfs:range` correct", and others. Many vocabularies seem to be not quite so visible and this fact demands something like hub or repository for reusable vocabularies. A matter of particular importance is that there might not be a vocabulary for some domains at all. Because of the insufficient visibility and

[4] `http://xmlns.com/foaf/spec/`, accessed: June 30th, 2013
[5] `http://dublincore.org/documents/dcmi-terms/`, accessed: June 30th, 2013

the time-consuming exploration of the web for reusable vocabularies, there is a highly increased need for a proper search service for suitable vocabularies.

Considering the strategy for reusing vocabularies, we explain the result by the fact that our survey has addressed experts in the area of LOD and that those experts actually have already agreed on the common strategy to reuse as many vocabularies as *needed* before introducing a new one. However, as the conformance studies [HUH+12, HHP+10] illustrate the lack of compliance towards vocabulary reuse, there must be some differences in the decisions of which vocabularies to reuse. One explanation could be the distributed viewpoints on maximizing and minimizing the number of vocabularies, respectively. Regarding their (dis-)advantages, in general, there should be a fair balance between these two approaches, which also goes along the statement of balancing the correct but complicated formalism and the less accurate but usable data. This could result in a dataset that has a clear structure, is easier to understand and to be consumed by applications with support for wide-spread vocabularies. Furthermore, it is less likely that the Linked Data engineer has to fit several entity types into classes and properties that do not express the entire semantics. The ranking tasks second this argument. By reusing many well-known and wide-spread vocabularies, such as FOAF or Dublin Core, the data engineer makes the data easier to understand and to be consumed by LOD applications, as the terms from these vocabularies are most likely to be known or supported. At the same time, the engineer should try to keep an appropriate number of vocabularies in order to maintain a clear structure of the data. One good example for such an approach would be to introduce a class from an own vocabulary, if such a class does not already exist in another wide-spread vocabulary, and establish a `rdfs:subclass` relationship with a parent class from a well-known vocabulary, e.g., `ex1:MalePerson` is `rdfs:subclass` of `foaf:Person`.

4 Impact of the survey on LOVER

In general, it seems to be difficult to reuse appropriate vocabularies when modeling LOD in a real world scenario. The survey has provided us with the following results: (i) vocabulary reuse is still considered to be very important, (ii) there is an increased need for a proper search service for adequate vocabularies, and (iii) the requirements for a vocabulary recommendation tool comprise providing a lot of extensive meta information on vocabularies and their terms, keep a fair balance between maximizing and minimizing the number of vocabularies, and prefer to reuse well-established and widely used vocabularies. It seems to be obvious that the Linked Data engineer has to be supported in (ii) and (iii) to make appropriate modeling decisions. Existing vocabulary search engines like Swoogle [DFJ+04][6] and LOV [VVR11][7] provide a first support, as they allow to search for appropriate classes and properties by keyword search. However, a huge part of the modeling process still has to be done manually, as it demands additional efforts such as measuring the correctness of a retrieved class or property and incorporating the terms into a modeling system.

[6] `http://swoogle.umbc.edu/`, accessed: June 30th, 2013
[7] `http://lov.okfn.org/dataset/lov/`, accessed: June 30th, 2013

To alleviate the situation and to address the increased need for tool support in this area, we present the novel approach LOVER [SGSS13] (short for: Linked Open Vocabulary EngineeRing). LOVER is a generic approach, which enables the Linked Data engineer to select appropriate classes and properties from vocabularies in the LOD cloud. We use the results from this survey in order to implement LOVER in a way that it complies with the identified requirements and thereby supports the engineer in reusing vocabularies in the best possible way. For example, when searching for a class or property, LOVER should provide extensive meta information on both the vocabularies and their terms, and also support the engineer to mix an acceptable number of different and preferably widely used vocabularies. Taking these requirements into account, LOVER incorporates the SchemEX index [KGSS12] for a concept based search of appropriate classes and properties. This index contains a comprehensive directory of the use of all properties and concepts that appear in some dataset from the Billion Triple Challenge data from 2012. Such a concept based search is able to identify a best match for a concept or property within a RDF triple. For example, having the concepts `swrc:Publication` (from the SWRC[8] Ontology) and `foaf:Person`, LOVER is most likely to recommend the object property `dc:creator` between them. Such a selection can be optimized by reflecting the total usage of a concept or property in the LOD cloud, the usage withing one dataset, or the usage in specific triples. Additionally, other criteria that is derived from the resulting requirements can be added to optimize the recommendations. The search mechanism itself includes specific contextual information, such as the vocabularies already used in the model or the domain of the data. This context information is used to obtain better fitting results and to balance the number of vocabularies within one dataset. LOVER thereby supports the Linked Data engineer to decide whether a vocabulary is worth reusing or not. The modeling approach using LOVER is iterative, which allows for designing a first draft before refining it according to the *follow the nose* principle. This is analogical to the strategy most participants chose as their favorite approach to model Linked Open Data with respect to reusing existing vocabularies, i.e., reuse as many vocabularies as *needed* before introducing a new one. During each iteration, LOVER recommends a set of terms for every schema element, adapts, and updates the recommendations using the context information. This way, LOVER is most likely to achieve appropriate vocabulary recommendations and an optimized number of reused vocabularies, resulting in a dataset with an increased interoperability. All in all, the resulting requirement of the survey were crucial for deciding how to implement a tool like LOVER.

5 Conclusion and Future Work

We have presented a survey that helped us to investigate some factors for reusing and not reusing vocabularies, as well as illustrated the resulting requirements for future vocabulary recommendation tools. The identified factors which influence the engineer's decisions whether to reuse a vocabulary or not show that it is not easy to follow this best

[8] `http://ontoware.org/swrc/swrc/SWRCOWL/swrc_updated_v0.7.1.owl`, accessed: June 30th, 2013

practice. This especially concerns the decision-making process whether a vocabulary is worth reusing or not. Therefore, we have have derived requirements for future tools to support the engineer in this process. They express the urgent need for an appropriate search for classes and properties, extensive meta information for each vocabulary and its terms during recommendation, and for an adequate mix of different and preferably widely used vocabularies. To address the demand for such tools, we have proposed the novel LOVER approach, which supports the ontology engineer by recommending appropriate classes and properties from actively used vocabularies from the LOD cloud. The requirements from the survey were crucial for designing and implementing the LOVER approach. We illustrated the implementation of LOVER and explained how it will comply with these requirements. Currently, we are working on the formalization and the first prototype of LOVER. Regarding a first evaluation of its recommendations, we extract object properties from several triples within a well-modeled dataset, and let LOVER recommend a new property. Hereby, it will be interesting to see whether LOVER recommends the same property that has been withdrawn before. As a first prototype with an UI, we plan to incorporate LOVER within the Neologism platform[9] in 2014, and evaluate LOVER's benefits by performing a user test with LOD beginners and experts.

References

[BCH08] Chris Bizer, Richard Cyganiak, and Tom Heath. How to Publish Linked Data on the Web, July 2008.

[DFJ+04] Li Ding, Timothy W. Finin, Anupam Joshi, Rong Pan, R. Scott Cost, Yun Peng, Pavan Reddivari, Vishal Doshi, and Joel Sachs. Swoogle: a search and metadata engine for the semantic web. In *CIKM*, pages 652–659, 2004.

[HB11] Tom Heath and Christian Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Synthesis Lectures on the Semantic Web. Morgan & Claypool Publishers, 2011.

[HHP+10] Aidan Hogan, Andreas Harth, Alexandre Passant, Stefan Decker, and Axel Polleres. Weaving the Pedantic Web. 2010. peer-reviewed.

[HUH+12] Aidan Hogan, Juergen Umbrich, Andreas Harth, Richard Cyganiak, Axel Polleres, and Stefan Decker. An empirical survey of Linked Data conformance. *Web Semantics: Science, Services and Agents on the World Wide Web*, 14:14 – 44, 2012.

[KGSS12] Matthias Konrath, Thomas Gottron, Steffen Staab, and Ansgar Scherp. SchemEX - Efficient Construction of a Data Catalogue by Stream-based Indexing of Linked Data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2012.

[SGSS13] Johann Schaible, Thomas Gottron, Stefan Scheglmann, and Ansgar Scherp. LOVER: support for modeling data using linked open vocabularies. In *Proceedings of the Joint EDBT/ICDT 2013 Workshops*, EDBT '13, pages 89–92, New York, NY, USA, 2013. ACM.

[VVR11] Pierre-Yves Vandenbussche, Bernard Vatant, and Lise Rozat. Linked Open Vocabularies: an initiative for the Web of Data. In *In QetR Workshop, Chambery, France, 2011*, 2011.

[9] `http://neologism.deri.ie/`, accessed: June 30th, 2013