Review #2
Submitted by Robert Meusel
Recommendation: Minor Revision
Detail Comments

The paper states nicely the current state of the Linked Open Vocabulary (LOV) initiative and the system they created to support data publishers as well as ontology engineers in their work.
The paper is well written and clearly states the different aspects of the tool. First, the authors present a comprehensive overview about LOV itself (Section 2), then focus on the architecture of the system with its different components.
Section 4 explains the different use cases where the proposed/presented system can support users with different tasks where the following section names some systems/applications/initiatives making use of LOV (or parts of
it) already.
The related work comprehensively compares the proposed system to the state-of-the-art especially giving Table 8.

# Quality, Importance, and Impact
As LOD has grown more and more in the past years, it is necessary to help users to select vocabularies and improve the reuse of existing vocabs.
To enable this support, the presented tool is clearly a step in the right direction and as shown in Section 5 is already used by other systems/applications/initiatives.
Although they use it the paper does not provide a study if the system helps users more than user systems are capable of.

# Clarity, Illustration, and Readability The system/tool is presented in a clear and straight forward way. The figures help to understand the system workflow.

In general, I would recommend to accept the paper with a minor revision:

- I see that including a vocabulary into the system is controlled and managed by the curators. So far 511 vocabs have been added. It would be great to compare this number e.g. to the number of total vocabs in the current state of the LOD cloud (Schmachtenberger et al. Adoption of the linked data best practices in different topical domains.) as we found (Meusel et al. Towards Automatic Topical Classification of LOD Datasets) over $1\,400$ different vocabularies.
An analysis about the number of datasets using the included vocabs could also help to clarify the coverage of the system. Based on the findings of the proposed paper and our findings, I am also surprised about the huge amount of different classes and properties (e.g. ~39 classes on average per vocab, which for me is really surprising). Some explaining comments might help readers to pin down the helpfulness of those 511 vocabs.
- Figure 2 and 3 either need some more explanation or should be rethought, as at the current state I am not sure about the contribution of those figures.
Does Figure 2 state the number of new vocabularies which were created within each year? Do they all follow the current W3C recommendation language? Are those filtered by the format which is recommended? Also it would be helpful instead of stating the absolute number to state the percentage of all vocabs (from the 511 included).
- Although Table 8 gives a nice comparison between the systems, I think at the current state it is in some parts misleading and some information are missing. a) Ontology discovery method: Does automatic means, the other systems cannot be pointed manually to an ontology to include it? What does manually discovery mean? In addition: Interaction with the contributor is in my eyes not really

a feature of the system but of the process how the system is used. I would also be interested, as LOV comes with a large manual effort to include a new vocab (by the curator) how the other systems handle the inclusion. Here some more explanation is needed to give an comprehensive overview about the differences of the systems.
- Figure 8 mentions "Is a good fit?" and "Does meet quality?", without a clear statement about the criteria used to answer those questions. A further explanation would improve the understandability of those decisions. What criteria need to be full-filled to pass this state? Where is the difference between those two decisions?
- Currently, I do not see the contribution of Figure 9, as the processes is nicely explained in the text itself and in my eyes needs no further figurative explanation.
- Please separate clearly Chapter 7 and 8 (where does future work belong to?) In addition, the paper would benefit from a comprehensive style, as the last two chapters are more or less a stringing together of bullet points.
- Figure 4: The 0 should be "unknown" as the information is missing/not included in the dataset or the caption must be changed respectively.
- Chapter 4 is somehow creating a topical cut between Chapter 3 and 5. I would recommend the authors to rethink the location of this part and think about an inclusion into the introduction as motivation.
- (Optional) A contribution, which would definitely enable readers to judge the helpfulness of the tool would be a user study showing the usefulness of the tool based on the described use cases. I know that the design and execution of such a study is time and resource consuming, so I mark this point as optional. I think such a study would really help and I would recommend the authors to run such a study and publish the results in the near future.

Other minor comments:

- Please state the root-URL of the system somewhere in the beginning of the paper, as I had to go through the first pages to find the webpage of the tool.
- Some words are wrongly capitalized
    - Page 3: "In total, 45 Languages ..."
    - Page 4: "At the Vocabulary level ..."
    - Page 6: "Indeed, There are ..."
- Missing bracket at the end of the last sentence before Section 3.2
- Please capitalize "Figure", "Table", "Equation" (in text)
- Section 3.3.: Please either capitalize "Term frequency inverse document frequency" or lower case all terms
- Eq 4: score(t,Q) instead of Score(t,Q)
- Page 12: "... on the Web similar o the 5 stars ..." (to?)

Review #3
Submitted by Anonymous
Recommendation: Major Revision
Detail Comments

Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web.

The presented work provides a system report that describe Linked Open Vocabularies (LOV), a high quality catalogue of reusable vocabularies for the description of data on the Web. The paper is well written and very easy to understand for the readers.

In the Introduction section the authors start giving some historical background about the project, the have provided some details about the origin of the project. However, the justification of the presented work is missing, so, I would recommend to focus on justifying why the presented work is an essential component in the selected research field more rather than providing a technological context.

In the section 2 the system architecture of LOV is presented. Then, the LOV system is split in its modules and each one is individually analysed. In particular, in the section 2.2 the authors mentioned that :"At the Vocabulary Term level, the system extracts labels that will be used..", on this case the extracted labels are coming from the dictionaries, I am wondering if the authors have just considered to use others NLP technologies to extract labels from the text. Later, the authors have mentioned that "When some metadata failed to be extracted automatically, LOV curators enhance the description available in the system", could you be more explicitly, how the curators do that? it would be nice to describe deeply how they could help to overcome the mentioned issue.

In the section 3, several two relevant activities of the system are explained: Data Publication and Ontology Engineering. In particular, in the Ontology Engineering activities commented by the authors such as: Ontology Search, Ontology Assessment, Ontology Mapping, Ontology Localization. At the beginning of this section, the authors mentioned that this system could be used to "creation and reuse of ontologies". Well, the aims of each activity is clear and from my point of view the system open an opportunity for reusing ontologies using Ontology Mapping activity. But the creation of the ontology is should be clarified because this is usually a very hard task to achieve.
So, could you clarify what mentioned activities could be used to create a new ontology?.

The Derived tools and applications section shows the variety of applications that currently are using LOV system as a vocabulary provider.

The relation work and discussion section provides a deep analysis of different existing catalogues relevant that are currently using to find vocabularies.

Finally, the conclusion and future work section describes accurately the following directions of the LOV project. However the limitations are missed.
So, it would be acceptable to provide some details about them.

In conclusion, I believe that the paper could be improved before its publication. I would really encourage the authors to keep on working on it and to improve it.

Review #4
Submitted by Johann Schaible
Recommendation: Minor Revision
Detail Comments

This manuscript was submitted as 'Tools and Systems Report' and should be reviewed along the following dimensions: (1) Quality, importance, and impact of the described tool or system (convincing evidence must be provided). (2) Clarity, illustration, and readability of the describing paper, which shall convey to the reader both the capabilities and the limitations of the tool.

==========

The paper describes the Linked Open Vocabulary (LOV) initiative in detail.
The authors provide an extensive overview on the current state of LOV, describe the system's functionality as well as the entire process from analyzing new vocabularies to enabling a LOV user to search for reusable vocabularies and vocabulary terms, and how LOV is adopted in other applications or research activities.

The paper is well structured/written, and I enjoyed reading it. The authors motivate their work very clearly, illustrate the impressive amount of harvested vocabulary indicators, and describe the different ways Linked Data practitioners can access all functionality LOV offers. The system itself is also of high quality and works just as described. Overall, the paper and the system leave a satisfying impression. However, there are still a few small uncertainties in the paper, thus, I recommend to accept the paper given that the authors revise the paper according to the following suggestions.

General remarks:
Whereas bullet-points are very helpful for structuring a document and drawing attention to some specific statements, their excessive use causes the opposite. In Section 3, you use them just fine to briefly enumerate various information types, categories or others, but in Sections 6 and 7, they are way too long to point out something very clearly. Please use bullet-point to enumerate solely the categories (e.g. "Catalogues of generic vocabularies/schemas", "Catalogues of ontologies for a specific domain"), and go into detail in the following paragraphs. However, this is my personal preference and other reviewers can of course proof me wrong.
Furthermore, let a native speaker check the paper. There are various mistakes in spelling and (especially) grammar, which at times disturb the flow of reading.

Abstract:
"LOV goes beyond existing Semantic Web vocabulary search engines and takes into consideration the value's property type, matched with a query, to …" – what is meant by "the value" and what is its property type?
Please make that clear.

Section 2:
Table 1 describes the LOV dataset content by vocabulary element type. Please clarify what you mean by "Instances" and "Datatypes", especially as they have a median value of 0. It seems that you refer to rdfs:Class, rdf:Property, and rdfs:Datatype, but the reader can get confused, as there is no rdfs:Instance according to http://www.w3.org/TR/rdf-schema/.
In Table 3, the reader can observe that searching for Agents is most prominent. But why? Traditionally, an ontology engineer filters vocabularies that represent the domain of interest directly. Therefore, please provide a brief explanation for this difference-making feature, particularly as it was introduced only in version 3 and users were not accustomed to that type of search.

About Figures 1- 5: In Figure 5, you provide a description of the axes.
Please do so for the other figures (Figure 1 – Figure 4) as well to make them more self-explanatory.

Section 3:
The Data Layer of Figure 6 is basically not explained. What are the differences of the "LOV Catalogue" and the others in the Data Layer, and what are their benefits for LOV? Please, provide more details on that.
Regarding the inter-vocabulary relationships, especially the Specialization, Generalization, and Extension: How exactly are these links established? For example, if LOV finds a class from V1 to be a subclass of V2, it automatically establishes V1 to be a specialization of V2, or is there some kind of threshold that must be exceeded? Please, make that clear.
Figure 8 seems to be obsolete. First, The general process is clear from your description. Second, it is more confusing, since the difference between "Is a good fit?" and "Does meet quality?" is not explained. Is there one anyway? It seems that a vocabulary falls in the scope of LOV, iff it meets the quality requirements 1. – 5., or is the scope something else? Please clarify.
Furthermore, the five requirements can also be checked automatically. Yet for LOV it is performed manually. The following two questions arise: a) Which guidelines do the curators follow to ensure the vocabulary meets the five requirements? b) What is the cost or effort (in average) for the curators to review/validate a submitted vocabulary? Please clarify these aspects. (This was also mentioned in the previous reviews) Section 3.3 is a bit off in its structure. Three of the four data access possibilities are described from 3.3.1 to 3.3.3, but the fourth one is described in 3.4. The readability can be increased, if 3.3.1 describes the Search Engine, 3.3.2 the additional UI facets and elements helping to navigate within the vocabularies catalogue, 3.3.3 the data dump, 3.3.4 the SPARQL Endpoint, and 3.3.5 the API, since they are all part of the data access.

Section 4 and Section 5:
The listing in Section 5 provides fairly convincing evidence for an adequate impact of LOV on the Linked Data and ontology community. Whereas, IMHO, this is sufficient for accepting the paper, there is still a need for a user-study illustrating the actual impact of LOV on the Linked Data and ontology community. Specifically, Section 4 describes the relevance of LOV in three activities to support data publication and ontology engineering. However, this relevance is never proven. A user-study could do so (Such a study is solely a nice-to-have for this submission). One simplistic example: The participants of the user-study have the task of finding equivalent classes for the FOAF classes in Table 7 with and without LOV.

Section 6:
There are no clear statements distinguishing LOV from the other catalogues.
For instance: "Catalogues of generic vocabularies/schemas similar to LOV catalogue. Example of catalogues falling in this category are vocab.org, ontologi.es, JoinUp Semantic Assets or the Open Metadata Registry". But what are the differences between LOV and these catalogues? Is LOV the only manually curated catalogue, or are there further differences? This should be made very clear in the related work section, despite the fact that it might have been already mentioned somewhere else in the paper.
In "Search Engines of ontology terms", the service vocab.cc (http://vocab.cc/) is missing. Using it, just like LOV, one is able to search for prominent vocabularies and vocabulary terms. Please clarify how LOV is different from vocab.cc and add it to Table 8.

Section 7 and Section 8:
Both comprise "Future Work" in their titles. Please concentrate on the shortcomings of LOV in the Discussion and provide a brief outlook on the future work in the conclusion of the paper.

Minor remarks about the paper:

- Generally, when referring to a section, a figure, or a table, capitalize the reference, e.g., "In Section 3, we will..." or "...as it is illustrated in Table 2.3".
- Refer to equations with the number in the brackets, i.e. instead of "Equation 1 shows..." use "Equation (1) shows..."
- In various tables: instead of describing the number of vocabularies with "Nb Vocabs", use "# of vocabs" or "#vocabs"
- Table 6: Instead of N, use |V| for specifying the number of vocabularies in the set of Vocabularies. Less variables provide a better readability.
- "...auto-completion together with http://prefix.cc for namespace..."

→ use just prefix.cc with a footnote pointing to http://prefix.cc

Minor remarks about the system:

- In the SPARQL editor it says "press CMD – Spacebar for autocomplete"

in the completionNotification HTML-div. On a Mac that this command for something else, but "CTRL – Spacebar" works

Typos:

- The last two decades has → the last two decades have
- breakdown of LOV dataset content → breakdown of the LOV dataset content
- 27.98% vocabularies → 27.98% of the vocabularies (that typo occurred more often. Please adjust all of them)
- LOV architecture is composed → The LOV architecture
- The information concerning vocabulary terms use in Linked Open Data → The information concerning the use of a vocabulary term in the Linked Open Data cloud
- In both case → in both cases
- We list below some tools... → Below we list some tools...
- Maguire et al. [17] use LOV search API → Maguire et al. [17] use the LOV search API