Semantic Web 1 (2014) 1–5
IOS Press

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

Editor(s): Michel Dumontier, Stanford Center for Biomedical Informatics Research, Stanford, CA, USA
Solicited review(s): Wei Hu, Nanjing University, China; Robert Meusel, University of Mannheim, Germany; Johann Schaible, GESIS, Germany; Robert Danitz, Fraunhofer FOKUS, Berlin, Germany; Christian Bizer, University of Mannheim, Germany; Irene Celino, CEFRIEL – Politecnico di Milano, Italy; one anonymous reviewer

Pierre-Yves Vandenbussche [a,*], Ghislain A. Atemezing [b], María Poveda-Villalón [c] and Bernard Vatant [d]
[a] *Fujitsu (Ireland) Limited, Swords, Co. Dublin, Ireland*
E-mail: pierre-yves.vandenbussche@ie.fujitsu.com
[b] *Mondeca, 35 boulevard de Strasbourg, 75010 Paris, France*
E-mail: ghislain.atemezing@mondeca.com
[c] *Ontology Engineering Group (OEG), Universidad Politécnica de Madrid, Madrid, Spain*
E-mail: mpoveda@fi.upm.es
[d] *Mondeca, 35 boulevard de Strasbourg, 75010 Paris, France*
E-mail: bernard.vatant@mondeca.com

Abstract.
One of the major barriers to the deployment of Linked Data is the difficulty that data publishers have in determining which vocabularies to use to describe the semantics of data. This systematic report describes Linked Open Vocabularies (LOV), a high-quality catalogue of reusable vocabularies for the description of data on the Web. The LOV initiative gathers and makes visible indicators such as the interconnections between vocabularies and each vocabulary's version history, along with past and current editor (individual or organization). The report details the various components of the system along with some innovations, such as the introduction of a property-level boost in the vocabulary search scoring that takes into account the property's type (e.g, `dc:comment`) associated with a matching literal value. By providing an extensive range of data access methods (full-text search, SPARQL endpoint, API, data dump or UI), the project aims at facilitating the reuse of well-documented vocabularies in the Linked Data ecosystem. The adoption of LOV by many applications and methods shows the importance of such a set of vocabularies and related features for ontology design and the publication of data on the Web.

Keywords: LOV, Linked Open Vocabularies, Ontology search, Linked Data, Vocabulary catalogue

1. Introduction

The last two decades have seen the emergence of a "Semantic Web" enabling humans and computer systems to exchange data with unambiguous, shared meaning. This vision has been supported by World Wide Web Consortium (W3C) Recommendations such as the Resource Description Framework (RDF), RDF-Schema and the Web Ontology Language (OWL). Thanks to a major effort in publishing data following Semantic Web and Linked Data principles [6], there are now tens of billions of facts spanning hundreds of linked datasets on the Web covering a wide range of topics. Access to the data is facilitated by portals (such as Datahub[1] or UK Government Data[2]) or direct publication by organisations (e.g. New York Times[3]).

*Thanks to Amélie Gyrard, Thomas Francart, Thérèze Rogez, Laurent Irles and Anthony McCauley for their help on the project.

[1] `http://datahub.io/`
[2] `http://data.gov.uk/`
[3] `http://data.nytimes.com/`

Despite the enormous volume of data now available on the Web, the Linked Data community has relatively little interest in vocabulary[4] management, focusing rather on the data itself. A vocabulary consists of classes, properties and datatypes that define the meaning of data. RDF vocabularies are themselves expressed and published following the Linked Data principles; this gives humans and machines access to the definitions of the terms used to qualify the data. Unfortunately some vocabularies are not published or are no longer not available; this breaks the semantic interoperability of the data, one of the fundamental principles of the Semantic Web [16].

The Linked Open Vocabularies (LOV) initiative[5] is an innovative observatory of the semantic vocabularies ecosystem. Started in March 2011, as part of the DataLift research project [27] and hosted by the Open Knowledge Foundation, LOV gathers and makes visible indicators not previously harvested, such as the interconnections between vocabularies, the versioning history along with past and current editor (individual or organization). The number of vocabularies indexed by LOV is constantly growing (527 as of October 2015) thanks to a community effort. It is the only catalogue, to the best of our knowledge, that accepts all types of search criteria: metadata search, ontology search, APIs, a comprehensive dump file and SPARQL endpoint access.

The purpose of LOV is to promote and facilitate the reuse of well documented vocabularies in the Linked Data ecosystem. In D'Aquin and Noy [12]'s categorisation of ontology libraries, LOV falls into the categories *"curated ontology directory"* and *"application platform"*. Specifically, LOV supports the following main activities for the design of ontologies and the publication of data on the Web [30,19,20,32]:

Ontology Search. LOV enables searching for vocabulary terms (class, property, datatype) based on domain: vocabularies (and therefore vocabulary terms) are categorised according to the domain they address.

Ontology Assessment. LOV provides a ranking (cf. Section 3.3.1 for each term retrieved by a keyword search to assist in ontology assessment.

Ontology Mapping. LOV categorizes seven different types of relationships between ontologies: metadata, import, specialization, generalization, extension and equivalence (cf. Section 3.1.1). These relationships can be useful for finding alignments between ontologies.

This report is structured as follows: in the next section, we provide statistics on the usage of LOV. In Section 3, we describe the components and features of the system. Thereafter, in Section 4, we provide an overview of some applications and research projects based on and motivated by the LOV system. In Section 5, we report on related work. The limitations and further development of LOV are discussed in Section 6. We conclude in Section 7.

2. LOV state

The LOV dataset consists of 527 vocabularies as of October 2015[6]. Figure 1 shows the evolution of the number of vocabularies inserted in the LOV dataset since March 2011. The addition of new vocabularies to LOV has been fairly constant with two exceptions: 1) the deployment of LOV version 2 [early 2012] automated most of the vocabulary analyses, resulting in the increase number of vocabularies ; and 2) the deployment of LOV version 3 [early 2015], resulting in a small decrease and plateau of the vocabularies. At that time we were considering removing offline vocabularies but finally decided to keep them with a special flag, making LOV the only source of continuity for datasets referencing unreachable vocabularies.

By observing the vocabularies contained in LOV as a whole, we can extract some information about Semantic Web adoption and dynamics. Figure 2 shows the distribution of LOV vocabularies by creation date. The distribution follows a bell curve with its peak in 2011. It is worth noting that a decrease in number of vocabulary creation does not necessarily mean a decrease in use of the technology but rather that the existing vocabularies now cover a large part of the semantic description needed. When looking at the last modified date of the same vocabularies (as illustrated in Figure 3), we see that LOV vocabularies are part of a living ecosystem in constant evolution.

Overall, the LOV dataset contains 20,000 classes and almost 30,000 properties. The median is 9 classes and 17 properties per vocabulary. Table 1 presents

[4] We use the terms "semantic vocabulary", "vocabulary" and "ontology" interchangeably.

[5] `http://lov.okfn.org/dataset/lov/`

[6] However, the figures and evaluation used in this report are based on LOV catalogue with 511 vocabularies as of June 2015.

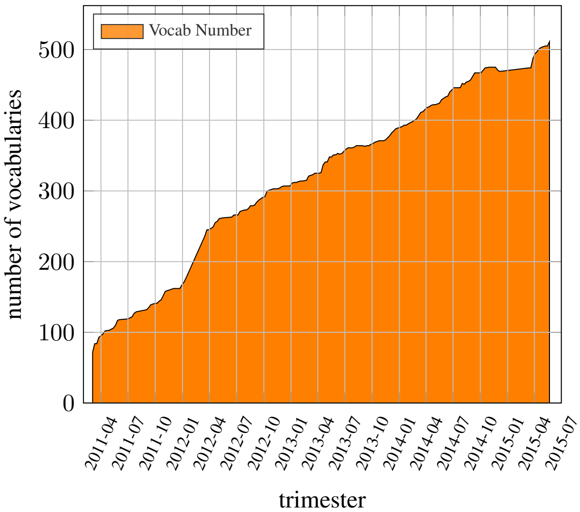

Fig. 1.: Evolution of the number of vocabularies in LOV from March 2011 to June 2015.

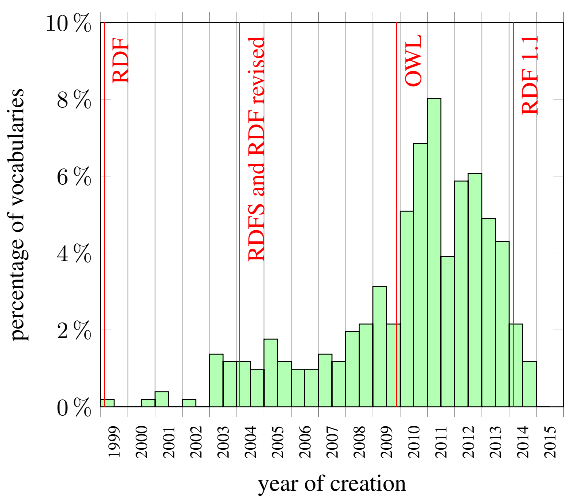

Fig. 2.: Distribution of LOV vocabularies by creation date. For indication, we use vertical red lines to mark the official release dates of the main Semantic Web languages (RDF, RDFS and OWL).

a breakdown of LOV content by vocabulary element type. In this Table, the *Classes* type refers to any instance of `rdfs:Class` or `owl:Class`; the *Properties* type refers to any instance of `rdf:Property` or by inference, any instance of subclasses of `rdf:Property` defined in the OWL language; the *Datatypes* type refers to any instance of `rdfs:Datatype`; and finally, the members of a vocabulary class are known as *instances* of the class.

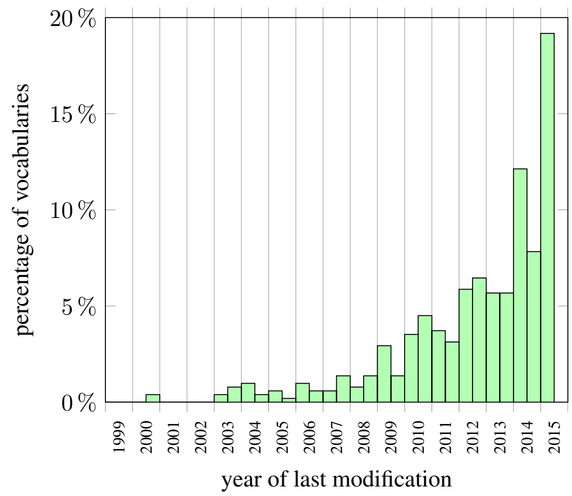

Fig. 3.: Distribution of LOV vocabularies by last modified date.

Type	**Count**	**Median per vocab**
Properties	29,925	17
Classes	20,034	9
Instances	5,232	0
Datatypes	101	0

Table 1: LOV vocabulary element types statistics.

Out of 511 vocabularies, 66.14% explicitly use the English language for labels/comments, i.e containing `@en` tag. Table 2 presents the number and percentage of vocabularies using the top five languages detected in LOV. Figure 4 shows the distribution of vocabularies per number of languages explicitly used: 27.98% of the vocabularies still do not provide any language information, and only 14.68% of the vocabularies are multilingual. In total, 45 languages are used by vocabularies in LOV. We will discuss the importance of providing multilingual vocabularies in Section 7.

Language	**# vocabs**	**% vocabs (out of 511)**
English	338	66.14%
French	37	7.24%
Spanish	25	4.89%
German	19	3.72%
Italian	18	3.52%

Table 2: Top five languages detected in the LOV catalogue, showing numbers and percentages of vocabularies using them. A vocabulary can make use of multiple languages.

From January to June 2015, more than 1.4 million searches were conducted on LOV[7]. A breakdown of searches per element type is provided in Table 3. We can see that agent search (for person or organisation) is the most prevalent; this is a new feature in LOV version 3. This might be explained by the uniqueness (when compared to other ontology catalogues) and the recent development of this feature in LOV, which now allows a user to visualise who defined or published vocabularies. Searches that include keywords (and not only filters) are mainly seek vocabulary terms. Table 4 presents the top 10 searched terms between January and June 2015. Although most of the searches are performed through the user interface, an application ecosystem using LOV APIs has surfaced, as shown in Figure 5.

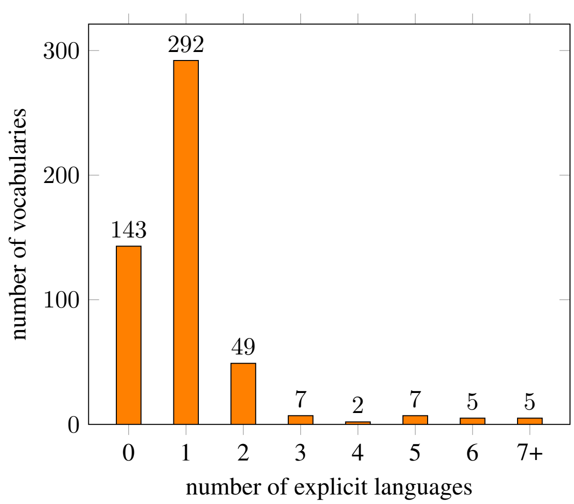

Fig. 4.: Distribution of LOV vocabularies by number of languages explicitly mentioned using language tag. “Zero" means that there is no explicit language tag declared (i.e. no literal value of the vocabulary has a language tag).

Since 2011, the Linked Open Vocabularies initiative has gathered a community of about 480 people interested in various domains, including ontology engineering and data publication. The LOV Google+ community[8] is now an important place to discuss, report and

Vocabulary Term	# searches	% searches
set	7,092	8.79%
domain	2,518	3.12%
some	2,473	3.06%
status	1,486	1.84%
iso 639	1,389	1.72%
same	1,285	1.59%
state	1,235	1.53%
supports	1,145	1.42%
start	887	1.1%
space	864	1.07%

Table 4: Top 10 terms searched from January to June 2015 by users in LOV.

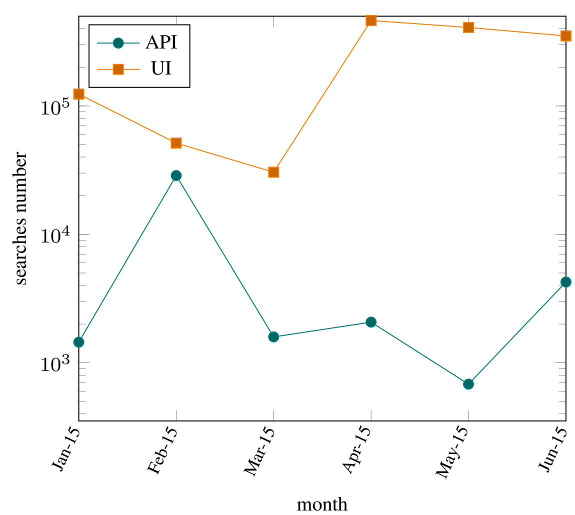

Fig. 5.: Evolution of the number of searches through UI and API methods from January to June 2015. Note that the y axis has a logarithmic scale.

announce general facts related to vocabularies on the Web. The LOV dataset itself references 389 resources of type *Person* and 111 of type *Organization* participating in vocabulary design and/or publication.

3. System Components and Features

The LOV architecture is composed of four main components (cf. Figure 6): 1) *Tracking and Analysis*. Checks for any vocabulary version update and analyses vocabularies’ specific features. 2) *Curation*. Ensures the high quality of the LOV dataset by enabling the community to suggest vocabularies or edit the catalogue. 3) *Data Access*. Provides access to the data

[7] This figure includes searches from the API and UI as well as searches with and without keywords such as “all agents that participated in vocabulary design and publication in the geo-location domain”.

[8] https://plus.google.com/communities/108509791366293651606

Element Type	# searches	% searches	# searches with keyword	% searches with keyword
Term	205,682	14.19%	80,728	92.84%
Vocabulary	178,837	12.34%	5,918	6.81%
Agent	1,064,597	73.47%	306	0.35%

Table 3: Type of elements searched from January to June 2015 by users in LOV for all searches and those with keyword.

through a large range of methods and protocols to facilitate the use of LOV dataset and 4) *Data Storage*. Offers a reliable and efficient method for storing and querying the data. Each component provides a set of features detailed in the following subsections.

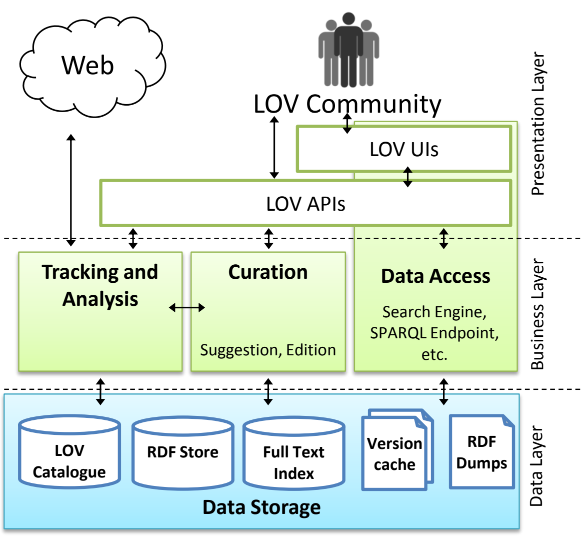

Fig. 6.: Overview of the Linked Open Vocabularies Architecture.

3.1. *Tracking and Analysis*

The *Tracking and Analysis* component dereferences[9] LOV vocabularies, stores a version locally (in Notation 3 format) and extracts relevant metadata.

3.1.1. *Vocabulary Level Analysis*

At the vocabulary level, the system extracts three types of information for each vocabulary version (Figure 7):

- The metadata associated to the vocabulary. This information is explicitly defined within the vocabulary to provide context and useful data about the vocabulary. Some high level vocabularies can be reused for that purpose, such as Dublin Core[10] to describe authors, contributors, publishers or Creative Commons[11] for the description of a license.
- Inlinks/incoming vocabularies, making explicit the links from another vocabulary based on the semantic relation of their terms.
- Outlinks/outgoing vocabularies, making explicit the links to another vocabulary based on the semantic relation of their terms.

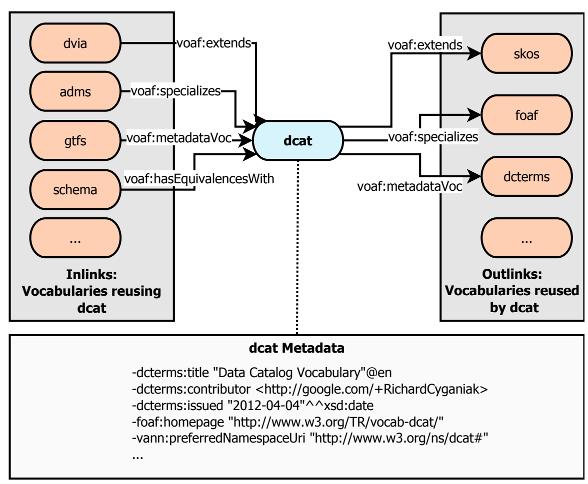

Fig. 7.: Metadata type, vocabulary inlinks and outlinks of DCAT vocabulary.

Two vocabularies can be interlinked in many different ways. Consider two vocabularies $V1$ and $V2$ such that $V1$ contains a class $c1$ and a property $p1$ and $V2$ contains a class $c2$ and a property $p2$. Relationships between these two vocabularies can be of the following types (the lines and numbers in brackets correspond to real examples presented in Listing 1):

[9]A URI is looked up over HTTP to return content in a processable format such as XML/RDF, Notation 3 or Turtle.

[10]`http://purl.org/dc/terms/`
[11]`http://creativecommons.org/ns#`

Metadata. some terms from $V2$ are reused to provide metadata about $V1$, Listing 1 lines 1-2 .

Import. some terms from $V2$ are reused with $V1$ to capture the semantic of the data (lines 3 to 4).

Specialization. $V1$ defines some subclasses or subproperties (or local restrictions) of $V2$, Listing 1 lines 5-8.

Generalization. $V1$ defines some superclasses or superproperties of $V2$, Listing 1 lines 9-11.

Extension. $V1$ extends the expressivity of $V2$, Listing 1 lines 12-15.

Equivalence. $V1$ declares some equivalent classes or properties with $V2$, Listing 1 lines 16-20.

Disjunction. $V1$ declares some disjunct classes with $V2$, Listing 1 lines 21-23.

These relationships, with the exception of *Import* which is represented by `owl:imports`, are captured by the Vocabulary of a Friend[12] (VOAF). Whenever a new vocabulary/vocabulary version is added to LOV, the system automatically detects and adds the inter-vocabulary relationships to the LOV catalogue using specific *Construct* SPARQL queries[13]. Table 5 presents a breakdown of the occurrences of each relation in LOV.

Inter-vocabulary relationship	**# relations**
voaf:metadataVoc	2,637
voaf:specializes	1,269
voaf:extends	1,031
owl:imports	373
voaf:hasEquivalencesWith	201
voaf:generalizes	57
voaf:hasDisjunctionsWith	16

Table 5: Inter-vocabulary relationship types and their number of occurrences in LOV.

3.1.2. Vocabulary Term Level Analysis

At the vocabulary term level, the system extracts two types of information:

- term type (class, property, datatype or instance defined in the namespace of the vocabulary) indexed by the system's search engine so it can be used to filter a search.
- term natural language annotations (RDF literals) with their predicate and language (e.g. `rdfs:label "Temperature"@en`). This information is provided as is for indexing by the search engine and will later be used (cf. Section 3.3.1) in the scoring algorithm.

The information concerning the usage of a vocabulary term in Linked Open Data, also called "popularity", is used in LOV search results scoring as explained in Section 3.3.1. This information is not natively present in the vocabularies and can not be inferred from the LOV dataset. We make use of the LODStats project which gathers comprehensive statistics about RDF datasets [3]. LOV regularly fetches LODStats raw data[14] described using the Vocabulary of Interlinked Datasets (VoID) [1] and the Data Cube vocabulary. We pre-process LODStats data before inserting it to LOV. Indeed, there are many duplicates in LODStats representing in fact the same vocabulary URI (e.g., `foaf` has three different records[15], and has to be mapped to a single entry in LOV)

3.2. Curation

The vocabulary collection is maintained by curators who are responsible for validating metadata information, inserting a vocabulary in the LOV ecosystem, and assigning a review on the suggested vocabulary.

3.2.1. Vocabulary Insertion

Compared to other vocabulary catalogues (cf. Section 5), LOV relies on a semi-automated process for vocabulary insertion. Whereas an automated process focuses only on volume, in our process, we focus on the quality of each vocabulary and therefore the quality of the overall LOV ecosystem. Suggestions come from the community and from inter-vocabulary reference links. Our system provides a feature to suggest[16] the insertion of a new vocabulary. This feature allows a user to check what information the LOV system can automatically detect and extract. LOV curators then check whether the vocabulary meets the following LOV quality requirements:

[12] `http://lov.okfn.org/vocommons/voaf/`

[13] The SPARQL Queries are described in the VOAF vocabulary

[14] We retrieve the statistics available at: `http://stats.lod2.eu/rdfdocs/void`. Unfortunately this file has been unavailable since June 2014 which explains some differences between the statistics we use and LODStats.

[15] `http://stats.lod2.eu/vocabularies?search=foaf`

[16] `http://lov.okfn.org/dataset/lov/suggest/`

Listing 1: Examples of Inter-vocabulary relationships.

```
# Metadata
<http://www.w3.org/2004/02/skos/core> dct:title "SKOS Vocabulary"@en
# Import - V1 imports V2
<http://purl.org/NET/c4dm/event.owl> owl:imports <http://www.w3.org/2006/time>
# Specialization - c1 is subclass of c2
<http://open.vocab.org/terms/Birth> rdfs:subClassOf <http://purl.org/NET/c4dm/event.owl#Event>
# Specialization - p1 is subproperty of p2
<http://purl.org/spar/fabio/hasEmbargoDate> rdfs:subPropertyOf <http://purl.org/dc/terms/date>
# Generalization - c1 has for narrower match c2
<http://semanticweb.cs.vu.nl/2009/11/sem/Place> skos:narrowMatch
   <http://www.w3.org/2003/01/geo/wgs84_pos#SpatialThing>
# Extension - p1 is inverse of p2
<http://vivoweb.org/ontology/core#translatorOf> owl:inverseOf <http://purl.org/ontology/bibo/translator>
# Extension - p1 has for domain c2
<http://xmlns.com/foaf/0.1/based_near> rdfs:domain <http://www.w3.org/2003/01/geo/wgs84_pos#SpatialThing>
# Equivalence - p1 is equivalent to p2
<http://lsdis.cs.uga.edu/projects/semdis/opus#journal_name> owl:equivalentProperty
   <http://purl.org/net/nknouf/ns/bibtex#hasJournal>
# Equivalence - c1 is equivalent to c2
<http://www.loc.gov/mads/rdf/v1#Language> owl:equivalentClass <http://purl.org/dc/terms/LinguisticSystem>
# Disjunction - c1 is disjoint with c2
<http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#TimeInterval>owl:disjointWith
   <http://www.ontologydesignpatterns.org/ont/dul/ontopic.owl#SubjectSpace>
```

1. a vocabulary should be written in RDF and be dereferenceable;
2. a vocabulary should be parsable without error (warnings are tolerated);
3. all vocabulary terms (classes, properties and datatypes) in a vocabulary should have an `rdfs:label`;
4. a vocabulary should refer to and reuse relevant existing ones; and
5. a vocabulary should provide some metadata about the vocabulary itself (at least a title).

If a suggested vocabulary meets these criteria it is then inserted in the LOV catalogue. During this process, LOV curators keep the authors informed and help them to improve their vocabulary quality. As a result of our experience in vocabulary publication, we developed a handbook of metadata recommendations for Linked Open Data vocabularies to help in publishing well documented vocabularies [31].

3.2.2. Vocabulary Review

When automatic extraction of metadata fails, LOV curators enhance the description available in the system and notify the vocabulary authors of the pitfalls' report. This manual task usually consists in checking for any additional information present in the HTML documentation (targeted for humans) and not reflected in the RDF description. The documentation provided by the LOV system assists users in understanding the semantics of each vocabulary term and therefore of any data using the term. For instance, information about the creator and publisher is a key indication for a vocabulary user in case help or clarification is required from the author, or to assess the stability of that artifact. About 55% of the vocabularies specify at least one creator, contributor or editor. We augment this information using manually gathered information, leading to the inclusion of data about the creator in over 85% of the vocabularies in LOV. The database stores every version of a vocabulary since its first issue. For each version, a user can access the file (particularly useful when the original online file is no longer available). A script automatically checks for vocabulary updates on a daily basis. When a new version is detected, it is stored locally, and the statistics about that vocabulary are recomputed. Similarly we ensure that curated review for each vocabulary is less than one year old by sending curators a notification when a vocabulary review is older than eleven months. In both cases, curators update the vocabulary review accordingly.

3.3. Data Access

The LOV system (code and data) is published under a Creative Commons 4.0 license[17] (CC BY 4.0). Users and applications can access the LOV data in four ways:

[17]`https://creativecommons.org/licenses/by/4.0/`

1. Query the LOV search engine to find the most relevant vocabulary terms, vocabularies or agents matching keywords and/or filters;
2. Download data dumps of the LOV catalogue in RDF Notation 3 format or the LOV catalogue and the latest version of each vocabulary in RDF N-quads format;
3. Run SPARQL queries on the LOV SPARQL Endpoint; and
4. Use the LOV API which provides a full access to LOV data for software applications.

3.3.1. Search Engine

In [9], Butt *et al.* compare eight different ranking methods grouped in two categories for querying vocabulary terms:

1. Content-based Ranking Models: tf-idf, BM25, Vector Space Model and Class Match Measure.
2. Graph-based Ranking Models: PageRank, Density Measure, Semantic Similarity Measure and Betweenness Measure.

Based on their findings, we defined a new ranking method adapting *term frequency inverse document frequency* (tf-idf) to the graph-structure of vocabularies. Compared to the other methods, tf-idf takes into account the relevance and importance of a resource to the query when assigning a weight to a particular vocabulary for a given query term. We reuse the augmented frequency variation of term frequency formula to prevent a bias towards longer vocabularies. Because of the inherent graph structure of vocabularies, tf-idf needs to be tailored so that the basic unit is not a word, but rather a vocabulary term t in a vocabulary V. Equation (1) presents the adaptation of tf-idf to vocabularies (a definition of the variables used in this paper's equations is provided in Table 6).

$$\begin{aligned} tf(t,V) &= 0.5 + \frac{0.5 * f(t,V)}{max\,\{f(t_i,V) : t_i \in V\}} \\ idf(t,\mathcal{V}) &= \log \frac{|V|}{|\,\{V \in \mathcal{V} : t \in V\}\,|} \end{aligned} \tag{1}$$

As highlighted in [9] and [26], the notion of the vocabulary term's popularity across the LOD datasets set $\mathcal{D}$ is quite important. In Equation (2) we introduce a new popularity measure, which is a function of the normalisation of the frequency $f(t,\mathcal{D})$ of a term URI t in the set of datasets $\mathcal{D}$ and the normalisation of the number of datasets in which a term URI appears $M(t) : t \in \mathcal{D}$. By using the maximum in this normalisation we emphasise the most used terms, result of a consensus within the community. This measure will give a higher score to terms that are often used in datasets and across a large number of datasets.

Variable	Description
$\mathcal{V}$	Set of Vocabularies
V	A vocabulary: $V \in \mathcal{V}$
$\|V\|$	Number of vocabularies in $\mathcal{V}$
t	A vocabulary term URI (class, property, instance or datatype): $t \in V, t \in URI$
Q	Query string
q_i	Query term i of Q
σ_V	Set of matched URIs for Q in V
$\sigma_V(q_i)$	Set of matched URIs for q_i in V : $\forall t_i \in \sigma_V$, $t_i \in V, t_i$ matches q_i
p	A term predicate: $p \in URI$
$\mathcal{D}$	Set of Datasets
D	A Dataset: $D \in \mathcal{D}$
$M(t_i)$	Number of Datasets: D in $\mathcal{D}$, $t_i \in D$

Table 6: Definition of the variables used in the equations.

$$\begin{aligned} pop(t,\mathcal{D}) = &\frac{f(t,\mathcal{D})}{max\,\{f(t_i,\mathcal{D}) : t_i \in \mathcal{D}\}} \\ &* \frac{M(t)}{max\,\{M(t_i) : t_i \in \mathcal{D}\}} \end{aligned} \tag{2}$$

RDF datasets have a consensual and stable structure, which arises from the best practices of vocabulary publication. It then becomes intuitive to assign more importance to a vocabulary term matching a query on the value of the property `rdfs:label` than `dcterms:comment`. Equation (3) extends the inner field-length norm $lengthNorm(field)$ from the Lucene-based search engine Elasticsearch, which attaches a higher weight to shorter fields, by combining it with a property-level boost $boost(p(t))$. Using this property-level boost we can assign a different score depending on which label property a query term matches. We distinguish four categories of matches:

– Local name (URI without the namespace). While a URI is not supposed to carry any meaning, it is a convention to use a compressed form of a term label to construct the local name. The local name therefore becomes an important artifact for term matching for which the highest score will be assigned. An example of local name matching the term "person" is `http://schema.org/Person`.

- Primary labels. The highest score will also be assigned for matches on the `rdfs:label`, `dce:title`, `dcterms:title`, `skos:prefLabel` properties. An example of primary label matching the term "person" is `rdfs:label "Person"@en`.
- Secondary labels. We define as secondary label the following properties: `rdfs:comment`, `dce:description`, `dcterms:description`, `skos:altLabel`. A medium score is assigned for matches on these properties. An example of secondary label matching the term "person" is `dcterms:description "Examples of a Creator include a person, an organization, or a service."@en`.
- Tertiary labels. Finally all properties not falling in the previous categories are considered as tertiary labels for which a low score is assigned. An example of tertiary label matching the term "person" is `rdarel2:name "Person"@en`.

$$norm(t, V) = lengthNorm(field) * \prod_{p \in V} boost(p(t)) \tag{3}$$

For every vocabulary in LOV, terms (classes, properties, datatypes, instances) are indexed and a full text search feature is offered[18]. Human users or agents can further narrow a search by filtering on term type (class, property, datatype, instance), language, vocabulary domain and vocabulary.

The final score of t for a query Q (Equation (4)) is a combination of the tf-idf, the importance of label properties of t on which query terms matched, and the popularity of that term in the LOD dataset. While the factorisation of the tf-idf and field normalisation factor is common for search engine ranking[19], we add a fourth parameter - the popularity - as it is fundamental in the Semantic Web. Indeed, the intention of LOV is to foster the reuse of consensual vocabularies that become *de facto* standards. The popularity metric provides an indication on how widely a term is already used (in frequency and in the number of datasets using it). We therefore add this new factor specific to the Semantic Web to the scoring equation:

$$score(t, Q) = tf(t, V) * idf(t, \mathcal{V}) * norm(t, V) * pop(t, \mathcal{D}) \quad : \forall t \{\exists q_i \in Q : t \in \sigma_V(q_i)\} \tag{4}$$

3.3.2. Data Dumps

The system provides two data dumps, one containing the LOV vocabulary catalogue only in RDF Notation 3 format[20] and another containing the LOV catalogue along with the latest version of each vocabulary and the statistics of use in LOD in RDF N-quads format[21](keeping each vocabulary in a separate named graph). As illustrated in Figure 8, the RDF model mainly reuses the Data CATalogue Vocabulary (DCAT) which allows the representation of the LOV catalogue as a `dcat:Catalog` composed of vocabulary entries (`dcat:CatalogRecord`) capturing information like the insertion date in LOV. Each entry point to the vocabulary itself is represented by a sub class of `dcat:Dataset` defined in the Vocabulary Of A Friend (VOAF). This artifact contains metadata extracted by the LOV application such as creators, first issued date, number of occurrences of the vocabulary in Linked Open Data. Each vocabulary is then linked to its various published versions represented by the `dcat:Distribution` entity on which information such as inter-vocabulary relations or languages can be found.

3.3.3. SPARQL Endpoint

The LOV SPARQL endpoint[22] offers a complementary data access method and allows clients to pose complex queries to the server and retrieve direct answers computed over the LOV dataset [8]. We use the Jena Fuseki triple store to store the N-quads file containing the LOV catalogue and the latest version of each vocabulary. We believe that this is the first service that allows users to query multiple vocabularies at the same time and to detect inter-vocabulary dependencies.

[18] `http://lov.okfn.org/dataset/lov/terms`
[19] See elasticsearch documentation: `http://bit.ly/1e37sFL`
[20] `http://lov.okfn.org/lov.n3.gz`
[21] `http://lov.okfn.org/lov.nq.gz`
[22] `http://lov.okfn.org/dataset/lov/sparql`

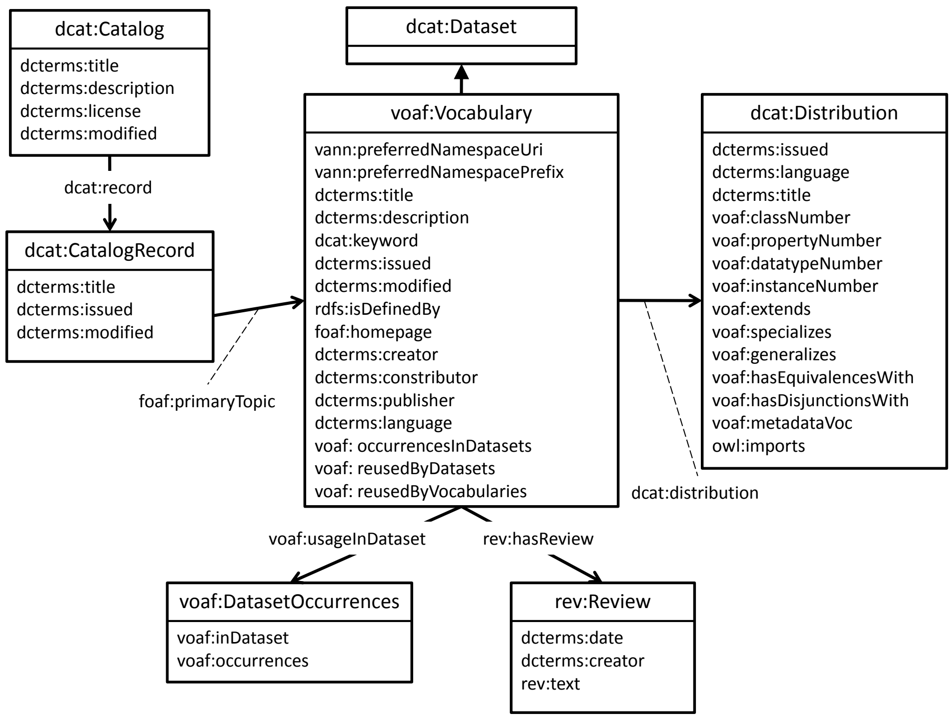

Fig. 8.: The LOV catalogue RDF schema model, in a UML class diagram representation.

3.3.4. *LOV Application Program Interfaces and User Interfaces*

LOV APIs give a remote access to the many functions of LOV through a set of RESTful services[23]. The basic design requirements for these APIs is that they should allow applications to get access to the very same information humans do via the User Interfaces. More precisely the APIs give access, through three different services (cf. Figure 9), to functions related to:

- Vocabulary terms (classes, properties, datatypes and instances). With these functions, a software application can query the LOV search engine, ask for auto-completion or a suggestion for misspelled terms.
- Vocabularies. A client can get access to the current list of vocabularies contained in the LOV catalogue; search for vocabularies, get auto-completion or obtain all details about a vocabulary.
- Agents. This provides a software agent with a list of all agent references in the LOV catalogue, a means to search for an agent, get auto-completion and details about an agent.

[23] `http://lov.okfn.org/dataset/lov/apidoc/`

LOV APIs are a convenient means to access the full functionality and data of LOV. It is particularly appropriate for dynamic Web applications using scripting languages such as JavaScript. The APIs described above have been developed for, and follow the requirements of, Ontology Design and Data Publication tools.

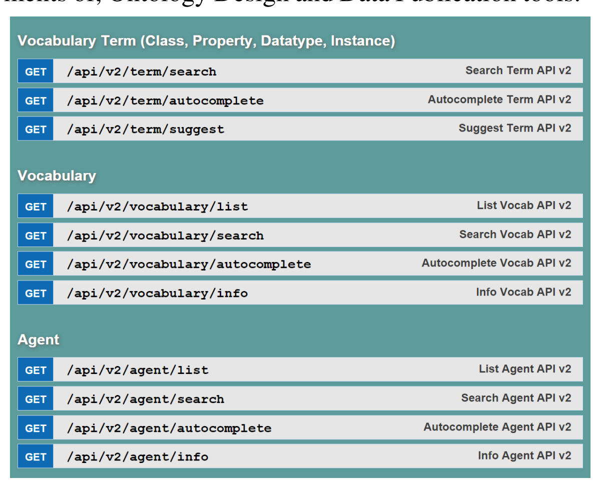

Fig. 9.: List of APIs to access LOV data.

The LOV Website offers intuitive navigation within the vocabularies catalogue. It allows users to explore vocabularies, vocabulary terms, agents and languages, and to see the connections between these entities. For instance, a user can use the agent search to look for experts in *geography* and *geometry* domains[24]. We use the d3[25] JavaScript library [7] to display charts and make the navigation more interactive; for example, we use the star graph representation to display incoming and outgoing links between vocabularies (cf. Figure 10).

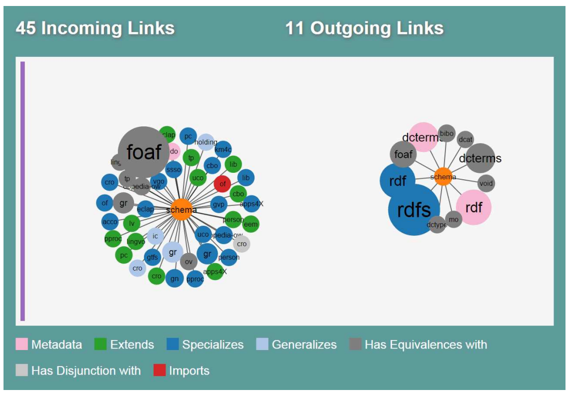

Fig. 10.: A graphical representation of the incoming and outgoing links for the Schema.org vocabulary as displayed in the UI.

3.4. Data Storage

To support the features presented above, we make use of specific storage technologies. The LOV catalogue is stored in MongoDB®, a document-based schema-less data store that scales and allows for dynamic changes in the data schema[26]. We use Jena Fuseki[27] to serve the data exported in RDF through the SPARQL protocol. The search feature is supported by Elasticsearch®, a full text index based on Lucene technology[28]. This storage solution is particularly well adapted to our User Interface technology (Node.js) as it offers RESTful APIs with output in JSON format. Finally we store each vocabulary version file and RDF dumps of LOV catalogue in the environment file system.

[24]`http://lov.okfn.org/dataset/lov/agents?&tag=Geography,Geometry`
[25]`http://d3js.org/`
[26]`https://www.mongodb.org/`
[27]`https://jena.apache.org/documentation/serving_data/`
[28]`https://lucene.apache.org/`

4. LOV Adoption

LOV, with its various data access methods, supports the emergence of a rich application ecosystem. Below we list some tools using our system as part of their service and project.

4.1. Derived tools and applications

In [18], Maguire *et al.* use the LOV search API to implement OntoMaton[29], a widget for using ontology lookup and tagging within the Google spreadsheets collaborative environment.

YASGUI (Yet Another SPARQL Query GUI)[30] is a client-side JavaScript SPARQL query editor that uses the LOV API for property and class auto-completion together with prefix.cc[31] for namespace prefix auto-completion [25]. YASGUI is itself reused by LOV for its SPARQL Endpoint User Interface.

Data2Ontology maps data objects and properties to ontology classes and predicates available in the LOV catalogue. Data2Ontology is part of the Datalift[32] platform [27], a framework for “lifting” raw data into RDF. The Data2Ontology module takes as input “raw RDF”, straightforward conversion of legacy format to RDF, with the goal of helping data publishers in selecting vocabulary terms that could be used to better represent their data.

OntoWiki[33] facilitates the visual presentation of a knowledge base as an information map, with different views on instance data [4]. It enables intuitive authoring of semantic content, with an inline editing mode for editing RDF content, similar to WYSIWIG for text documents. OntoWiki offers a vocabulary selection feature based on LOV.

Furthermore, we can mention the ProtégéLOV[34], a plug-in for the Protégé editor tool [14] that aims at improving the development of lightweight ontologies by reusing existing vocabularies at a low fine grained level. The tool searches for a term in LOV via APIs and provides three actions if the term exists : 1) to replace the selected term in the current ontology, 2) to add the `rdfs:subClassOf` axiom and 3) to add the `owl:equivalentClass`.

[29]`https://github.com/ISA-tools/OntoMaton`
[30]`http://legacy.yasgui.org/`
[31]`http://prefix.cc`
[32]`http://datalift.org/`
[33]`http://ontowiki.net/`
[34]`http://labs.mondeca.com/protolov/`

4.2. Using LOV as a Research platform

LOV has served as the object of studies in [21] where Poveda-Villalón *et al.* analysed trends in ontology reuse methods. In addition, the LOV dataset has been used to analyse the occurrence of good and bad practices in vocabularies [22].

Prefixes in the LOV dataset are regularly mapped with namespaces in the prefix.cc service. In [2], the authors perform alignments of Qnames of vocabularies in both services and provide different solutions to handle clashes and disagreements between preferred namespaces. Both LOV and prefix.cc provide associations between prefixes and namespaces but follow a different logic. The prefix.cc service supports polysemy and synonymy, and has a very loose control on its crowd-sourced information. In contrast, LOV has a much more strict policy forbidding polysemy and synonymy ensuring that each vocabulary in the LOV database is uniquely identified by a unique prefix identification allowing the usage of prefixes in various LOV publication URIs.

The LOV query log covering the period between 2012-01-06 and 2014-04-16 has been used in [9] to build a benchmark suite for ontology search and ranking. The CBRBench[35] benchmark uses eight ranking models of resources in ontologies and compares the results with ontology engineers' results. Our vocabulary term ranking method relies on and extends the outcome of this work.

In [16], the authors provide a 5 star rating for RDF vocabulary publication to boost interoperability, query federation and better interpretation of data on the Web similar to the 5 stars rating for Linked Open Data. Based on LOV's best practices criteria, all vocabularies must be 5 stars using this ranking and must provide further quality attributes imposed by LOV to facilitate vocabulary reuse.

RDFUnit[36] is a test-driven data debugging framework for the Web of Data. In [17], the authors provide an automatic test case for all available schema registered with LOV. Vocabularies are used to encode semantics to domain specific knowledge to check the quality of data.

Finally, Governatori et al. [15] analyse the current use of licenses in vocabularies on the Web based on the LOV catalogue in order to propose a framework to detect incompatibilities between datasets and vocabularies.

5. Related Work

Reusing vocabularies requires searching for terms in existing specialised vocabulary catalogues or search engines on the Web. While we refer the reader to [12] for a systematic survey of ontology repositories, below we list some existing catalogues relevant for finding vocabularies:

- *Catalogues of generic vocabularies/schemas* similar to LOV catalogue. Example of catalogues falling in this category are vocab.org[37], ontologi.es[38], JoinUp Semantic Assets or the Open Metadata Registry. Most of those repositories are not regularly updated and are created/owned by the institutions using the service.
- *Catalogues of ontologies for a specific domain* such as biomedicine with the BioPortal [33], geospatial ontologies with SOCoP+OOR[39], Marine Metadata Interoperability and the SWEET [24] ontologies[40]. The SWEET ontologies include several thousand terms, spanning a broad extent of Earth system science and related concepts (such as data characteristics), with the search tool to aid finding science data resources.
- *Catalogues of ontology Design Patterns (ODP)* focus on reusable patterns in ontology engineering [23]. The submitted patterns are small pieces of vocabularies that can further be integrated or linked with other vocabularies. ODP does not provide a search function for specific terms as is the case with some of these other catalogues.
- *Search Engines of ontology terms*. Among ontology search engines, we can cite: Swoogle [13], Watson [11], FalconS [10] and Vocab.cc [29]. These search engines crawl for data schema from RDF documents on the Web. They offer filtering based on ontology type (Class, Property) and a ranking based on the popularity. They don't look for ontology relations nor do they check if the definition of the ontology is available (usually known as dereferenciation). While in Swoogle

[35] `https://zenodo.org/record/11121`
[36] `https://github.com/AKSW/RDFUnit`

[37] `http://vocab.org/`
[38] `http://ontologi.es/`
[39] `https://ontohub.org/socop`
[40] `http://sweet.jpl.nasa.gov//`

the ranking score is displayed, Watson shows the language of the resource and the size. However, none of these services provide any relationship between the related ontologies, or any domain classification of the vocabularies. Table 7 presents a summary of key features of LOV with respect to Swoogle, Watson, Falcons and Vocab.cc.

– *Datasets and Vocabularies statistics*. In this category we can mention LODStats [3] and the vocabularies derived from the LOD Cloud. LODStats makes a bridge between datasets and vocabularies gathering up to 32 different statistical criteria based on a statement-stream-based approach for RDF datasets in Datahub[41]. LODStats maintains a comprehensive statistics on vocabularies terms (i.e. classes, properties) defined and used in a dataset. Schmachtenberg et al. [28] present a survey based on a large-scale Linked Data crawl from March 2014 to analyse the differences in best practices adoption across different application domains. Their results concerning the most used vocabularies (e.g., `foaf`, `dcterms`, `skos`, etc.) and the adoption of well-known vocabularies are inline with the findings of this paper.

While most of the related work focuses on automatic techniques to gather as many ontologies as possible, LOV focuses on maintaining a high quality collection of vocabularies that data publishers can reuse to describe their own data. To ensure the high quality of LOV data, we set up some stringent requirements for vocabularies to be inserted (cf. Section 3.2.1) such as the fact that a vocabulary URI must be dereferenceable. These kinds of requirements are not always taken into account in the aforementioned work: for instance, the authors in [28] define the notion of partly dereferenceable for vocabularies. As a consequence, anyone using a vocabulary referenced in LOV is ensured to get access to the vocabulary metadata but most importantly to its formal definition and preservation by accessing to various versions.

As part of our system evaluation we have compared the list of vocabularies in LOV with the ones in external services (LODStats and the empirical survey of Schmachtenberg et al. [28]) so as to understand the discrepancy.

LODStats contains 2,940 vocabularies extracted from datasets listed in Datahub.io. This list contains in fact a large number (2,596) of invalid vocabulary URIs and resource URIs that do not refer to a vocabulary (e.g. `http://data.kingcounty.gov/resource/d665-vvmd/` or `http://lod2.eu/view`). The domain "http://dati.opendataground.it" contains 962 Resource URIs which are instances and not vocabularies. As a result, only 344 candidate URIs in LODStats are comparable with LOV vocabularies. Out of those 344 URIs, 73 (21.22%) are covered by LOV. We randomly chose 20 vocabularies not already present in LOV for assessment. None of the randomly chosen vocabularies met LOV requirements and 8 different categories of errors were detected: 1) Failed to determine the triples content type, 2) Not found exception, 3) 403 forbidden, 4) Unknown host exception, 5) Peer not authenticated, 6) 504 gateway, 7) Bad URI and 8) Unqualified typed nodes are not allowed.

Recently, an updated comprehensive empirical survey of Linked Data conformance has been presented by Schmachtenberg et al. [28]. Their survey is based on a large-scale Linked Data crawl from March 2014 to analyse the differences of best practices adoption in different domains. Their results concerning the most used accessible vocabularies and the adoption of well-known vocabularies are inline with the findings of this paper. However, comparing the vocabularies in the LOD cloud with the LOV catalogue needs some alignments. From the 638 mentioned by Schmachtenberg et al., we removed invalid URIs such as domain names such as "umbel.org". Additionally we removed misspelled URIs and incomplete URIs. As a result, 270 candidate URIs (42.31%) can be compared with LOV vocabularies. Based on this analysis, we found that 102 vocabularies in the LOD cloud are already in the LOV catalogue, representing 38% of the 270 candidates. The general difference of our work with the one presented by Schmachtenberg et al. is that our approach applies strict criteria to include a vocabulary while their approach is dataset driven.

6. Discussion

Whilst providing access to high quality vocabularies, LOV system presents several limitations. As described in the last section, LOV system could benefit from an automatic discovery process to suggest vocabulary candidates. We could for instance extract vocabularies from the latest version of the Billion Triple Challenge or the Web Data Commons[42] dataset. Man-

[41] `http://datahub.io/`

[42] `http://webdatacommons.org/`

Feature	Swoogle	Watson	Falcons	Vocab.cc	LOV
Listing ontologies	Yes	Yes	Yes	Yes	Yes
Ontology discovery method	Automatic	Automatic	Automatic	Automatic	Automatic/Manual
Scope	SWDs	SWDs	Concepts	vocab terms	Vocabularies
Ranking	LOD metric	LOD metric	LOD metric	BTC corpus + label's property type	LOD/LOV metric
Domain filtering	No	No	No	No	Yes
Comments and review	No	Yes	No	No	Curators
Web service access	Yes	Yes	Yes	Yes	Yes
SPARQL endpoint	No	No	No	No	Yes
Read/Write	Read	Read/Write	Read	Read	Read
Ontology directory	No	No	No	Yes	Yes
Application platform	No	No	No	N/A	Yes
Storage	Cache	N/A	N/A	API	Dump/endpoint
Interaction with contributors	No	N/A	No	No	Yes
Version tracking	No	No	No	No	Yes
Inter-vocab. relationship visualization	No	No	No	No	Yes

Table 7: Comparison of LOV with respect to Swoogle, Watson, Falcons and Vocab.cc; adapted from the framework presented by d'Aquin and Noy [12]. SWD stands for Semantic Web Document.

ual curation is a critical activity to ensure the high quality of the LOV catalogue but also represents a limitation. At the moment we have been able to recruit new curators as the catalogue is growing. The version 3 of LOV system automates most of the processes and analyses but there are still some assessment and support activities that only a human can perform.

Currently, LOV's scope focuses on vocabularies for the description of RDF data and does not include any *Value Vocabularies* such as SKOS thesauri. By making the code of LOV system open source, we encourage anyone to set up an instance of the system to target such artifacts.

LOV relies on external projects such as LODStats to get the valuable information of vocabulary usage in published datasets. At the moment, the popularity information coming from LODStats does not take into account the most recent interest in publishing RDF data using markup language (e.g. `schema.org`). As a consequence, the popularity measure is incomplete and does not represent all possible use of a vocabulary. In future work we intend to extract those information from the latest datasets versions of the Billion Triple Challenge and the Web Data Commons.

From the study of LOV as a dynamic ecosystem we can draw two main lessons learned: the need for more multilingual vocabularies on the Web and the importance of long term preservation of vocabularies.

Labels are the main entry point to a vocabulary and their associated language is the key. Only 15% of LOV vocabularies make use of more than one language. Multilingualism is important at least for two reasons: 1) the most obvious one is allowing users to search, query and navigate vocabularies in their native language; and 2) translation is a process through which the quality of a vocabulary can only improve. Looking at a vocabulary through the eyes of other languages and identifying the difficulties of translation helps to better outline the initial concepts and if necessary refine or revise them. Hence multilingualism and translation should be native, built-in features of any vocabulary construction, not a marginal task.

Currently there is no solution for long-term vocabulary preservation on the Web [5]. This is a particularly important problem in a distributed and uncontrolled environment where any individual can create and publish a vocabulary. Third parties can reuse such vocabularies and therefore create a dependency on the original vocabulary availability as it retains the semantics of the data. This issue weakens the Semantic Web foundations.

7. Conclusions and Future work

In this system report we presented an overview of the Linked Open Vocabularies initiative, a high quality catalogue of reusable vocabularies for the description of data on the Web. The importance of this work is motivated by the difficulty that data publishers have in determining which vocabularies to use to describe their data. The key innovations described in this article include: 1) the availability of a high quality dataset of vocabularies available through multiple access methods 2) the curation by experts, making explicit for the first time the relationships between vocabularies and their version history; and 3) the consideration of property semantics in term search relevance scoring.

In the future, the LOV initiative could evolve in several ways. First, an area that is still largely unexplored is multi-term vocabulary search. During the ontology design process, it is common to have more than 20 concepts represented using existing vocabularies or a new one in case there is no corresponding artifact. While we are able to search for relevant terms in LOV it is still the responsibility of the ontology designer to understand the complex relationships between all these terms and come up with a coherent ontology. We could use the network of vocabularies defined in LOV to suggest not only a list of terms but graphs to represent several concepts together.

Second, we would like to provide more vocabulary based services such as vocabulary matching to help authors add more relationships to other vocabularies. Vocabulary checking is another service the community is asking for. We could integrate useful applications directly into LOV, such as Vapour[43], RDF Triple-Checker[44] and OOPS![45].

Another research direction is SPARQL query extension and rewriting based on Linked Vocabularies. Using the inter-vocabulary relationships we could transform a query to use the same semantics (same vocabulary terms) as the data source(s) being queried.

Finally, we plan to provide a user study and publish the results on the different usage of LOV by end users. In addition, we plan to include the vocabularies from LODStats and LOD Cloud that are suitable for inclusion in the LOV catalogue.

The adoption and integration of the LOV catalogue in applications for vocabulary engineering, reuse and data quality are significant. LOV has a central role in vocabulary life-cycle on the Web of data as highlighted by the W3C[46]: *"The success of LOV as a central information point about vocabularies is symptomatic of a need, for an authoritative reference point to aid the encoding and publication of data"*.

Acknowledgments

This work has been partially supported by the French National Research Agency (ANR) within the Datalift Project, under grant number ANR-10-CORD-009; the Spanish project BabelData (TIN2010-17550) and Fujitsu Laboratories Limited. The Linked Open Vocabularies initiative is graciously hosted by the Open Knowledge Foundation. We would like to thank all the members of LOV community and all the editors and publishers of vocabularies who trust in LOV catalogue. A special thank to Phil Archer, Julia Bosque Gil and Jodi Schneider for their valuable feedback and comments on this paper.

[43] http://validator.linkeddata.org/vapour
[44] http://graphite.ecs.soton.ac.uk/checker/
[45] http://oops.linkeddata.es/
[46] http://www.w3.org/2013/data/

References

[1] Keith Alexander and Michael Hausenblas. Describing linked datasets-on the design and usage of void, the vocabulary of interlinked datasets. In Christian Bizer, Tom Heath, Tim Berners-Lee, and Kingsley Idehen, editors, *In Linked Data on the Web Workshop (LDOW 09), in conjunction with 18th International World Wide Web Conference (WWW 09)*. Citeseer, 2009.

[2] Ghislain Auguste Atemezing, Bernard Vatant, Raphaël Troncy, and Pierre-Yves Vandenbussche. Harmonizing services for lod vocabularies: A case study. In Sam Coppens, Karl Hammar, Magnus Knuth, Marco Neumann, Dominique Ritze, Harald Sack, and Miel Vander Sande, editors, *WaSABi@ISWC*, volume 1106 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2013.

[3] Sören Auer, Jan Demter, Michael Martin, and Jens Lehmann. Lodstats - an extensible framework for high-performance dataset analytics. In Annette ten Teije, Johanna Völker, Siegfried Handschuh, Heiner Stuckenschmidt, Mathieu d'Acquin, Andriy Nikolov, Nathalie Aussenac-Gilles, and Nathalie Hernandez, editors, *Knowledge Engineering and Knowledge Management*, volume 7603 of *Lecture Notes in Computer Science*, pages 353–362. Springer Berlin Heidelberg, 2012. doi:10.1007/978-3-642-33876-2_31.

[4] Sören Auer, Sebastian Dietzold, and Thomas Riechert. OntoWiki – a tool for social, semantic collaboration. In *Lecture Notes in Computer Science*, pages 736–749. Springer Science and Business Media, 2006. doi:10.1007/11926078_53.

[5] Thomas Baker, Pierre-Yves Vandenbussche, and Bernard Vatant. Requirements for vocabulary preservation and governance. *Library Hi Tech*, 31(4):657–668, 2013. doi:10.1108/LHT-03-2013-0027.

[6] Tim Berners-Lee. Linked data - design issues. *W3C*, (09/20), 2006.

[7] M. Bostock, V. Ogievetsky, and J. Heer. Data-driven documents. *IEEE Trans. Visual. Comput. Graphics*, 17(12):2301–2309, dec 2011. doi:10.1109/tvcg.2011.185.

[8] Carlos Buil-Aranda, Aidan Hogan, Jürgen Umbrich, and Pierre-Yves Vandenbussche. Sparql web-querying infrastructure: Ready for action? In Harith Alani, Lalana Kagal, Achille Fokoue, Paul Groth, Chris Biemann, JosianeXavier Parreira, Lora Aroyo, Natasha Noy, Chris Welty, and Krzysztof Janowicz, editors, *The Semantic Web - ISWC 2013*, volume 8219 of *Lecture Notes in Computer Science*, pages 277–293. Springer Berlin Heidelberg, 2013. doi:10.1007/978-3-642-41338-4_18.

[9] AnilaSahar Butt, Armin Haller, and Lexing Xie. Ontology search: An empirical evaluation. In Peter Mika, Tania Tudorache, Abraham Bernstein, Chris Welty, Craig Knoblock, Denny Vrandečić, Paul Groth, Natasha Noy, Krzysztof Janowicz, and Carole Goble, editors, *The Semantic Web - ISWC 2014*, volume 8797 of *Lecture Notes in Computer Sci-*

ence, pages 130–147. Springer International Publishing, 2014. doi:10.1007/978-3-319-11915-1_9.

[10] Gong Cheng, Weiyi Ge, and Yuzhong Qu. Falcons: searching and browsing entities on the semantic web. In Jinpeng Huai, Robin Chen, Hsiao-Wuen Hon, Yunhao Liu, Wei-Ying Ma, Andrew Tomkins, and Xiaodong Zhang 0001, editors, *WWW*, pages 1101–1102. ACM, 2008. doi:10.1145/1367497.1367676.

[11] M. d'Aquin, C.Baldassare, L.Gridinoc, M. Sabou, S.Angeletou, and E.Motta. Watson: Supporting next generation semantic web applications. In *WWW/Internet conference 2007*, 2007.

[12] Mathieu d'Aquin and Natasha F. Noy. Where to publish and find ontologies? a survey of ontology libraries. *Web Semantics: Science, Services and Agents on the World Wide Web*, 11(0):96 – 111, 2012. doi:10.1016/j.websem.2011.08.005.

[13] Tim Finin, Li Ding, Rong Pan, Anupam Joshi, Pranam Kolari, Akshay Java, and Yun Peng. Swoogle: Searching for knowledge on the semantic web. In Anthony Cohn, editor, *Proceedings of the 20th National Conference on Artificial Intelligence - Volume 4*, AAAI'05, pages 1682–1683. AAAI Press, 2005.

[14] Nuria García-Santa, Ghislain Auguste Atemezing, and Boris Villazón-Terrazas. The protégélov plugin: Ontology access and reuse for everyone. In Fabien Gandon, Christophe Guéret, Serena Villata, John Breslin, Catherine Faron-Zucker, and Antoine Zimmermann, editors, *The Semantic Web: ESWC 2015 Satellite Events*, volume 9341 of *Lecture Notes in Computer Science*, pages 41–45. Springer International Publishing, 2015. doi:10.1007/978-3-319-25639-9_8.

[15] Guido Governatori, Ho-Pun Lam, Antonino Rotolo, Serena Villata, Ghislain Auguste Atemezing, and Fabien L. Gandon. Checking licenses compatibility between vocabularies and data. In Olaf Hartig, Aidan Hogan, and Juan Sequeda, editors, *COLD*, volume 1264 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.

[16] Krzysztof Janowicz, Pascal Hitzler, Benjamin Adams, Dave Kolas, and Charles Vardeman. Five stars of linked data vocabulary use. *Semantic Web*, 5(3):173–176, 2014. doi:10.3233/SW-140135.

[17] Dimitris Kontokostas, Patrick Westphal, Sören Auer, Sebastian Hellmann, Jens Lehmann, Roland Cornelissen, and Amrapali Zaveri. Test-driven evaluation of linked data quality. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 747–758, Republic and Canton of Geneva, Switzerland, 2014. International World Wide Web Conferences Steering Committee. doi:10.1145/2566486.2568002.

[18] E. Maguire, A. Gonzalez-Beltran, P. L. Whetzel, S.-A. Sansone, and P. Rocca-Serra. OntoMaton: a bioportal powered ontology widget for google spreadsheets. *Bioinformatics*, 29(4):525–527, dec 2012. doi:10.1093/bioinformatics/bts718.

[19] Sam Gyun Oh, Myongho Yi, and Wonghong Jang. Deploying linked open vocabulary (lov) to enhance library linked data. *Journal of Information Science Theory and Practice*, 2(2), Jun 2015. doi:10.1633/JISTaP.2015.3.2.1.

[20] Carlos Pedrinaci, Jorge Cardoso, and Torsten Leidig. Linked USDL: A vocabulary for web-scale service trading. In *The Semantic Web: Trends and Challenges - 11th International Conference, ESWC 2014, Anissaras, Crete, Greece, May 25-29, 2014. Proceedings*, Lecture Notes in Computer Science, pages 68–82. Springer International Publishing, 2014. doi:10.1007/978-3-319-07443-6_6.

[21] María Poveda-Villalón, Mari Carmen Suárez-Figueroa, and Asunción Gómez-Pérez. 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)*. Informatica, 2012.

[22] María Poveda-Villalón, Bernard Vatant, María Carmen Suárez-Figueroa, and Asunción Gómez-Pérez. Detecting good practices and pitfalls when publishing vocabularies on the web. In Aldo Gangemi, Michael Gruninger, Karl Hammar, Laurent Lefort, Valentina Presutti, and Ansgar Scherp, editors, *Proceedings of the 4th Workshop on Ontology and Semantic Web Patterns co-located with 12th International Semantic Web Conference (ISWC 2013), Sydney, Australia, October 21, 2013.*, volume 1188 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2013.

[23] Valentina Presutti and Aldo Gangemi. Content ontology design patterns as practical building blocks for web ontologies. In *Lecture Notes in Computer Science*, pages 128–141. Springer Science and Business Media, 2008. doi:10.1007/978-3-540-87877-3_11.

[24] Robert G. Raskin and Michael J. Pan. Knowledge representation in the semantic web for earth and environmental terminology (SWEET). *Computers & Geosciences*, 31(9):1119–1125, nov 2005. doi:10.1016/j.cageo.2004.12.004.

[25] Laurens Rietveld and Rinke Hoekstra. Yasgui: Not just another sparql client. In Philipp Cimiano, Miriam FernÃąndez, Vanessa Lopez, Stefan Schlobach, and Johanna VÃűlker, editors, *The Semantic Web: ESWC 2013 Satellite Events*, volume 7955 of *Lecture Notes in Computer Science*, pages 78–86. Springer Berlin Heidelberg, 2013. doi:10.1007/978-3-642-41242-4_7.

[26] Johann Schaible, Thomas Gottron, Stefan Scheglmann, and Ansgar Scherp. LOVER. In *Proceedings of the Joint EDBT/ICDT 2013 Workshops on - EDBT '13*. Association for Computing Machinery (ACM), 2013. doi:10.1145/2457317.2457332.

[27] François Scharffe, Ghislain Atemezing, Raphaël Troncy, Fabien Gandon, Serena Villata, Bénédicte Bucher, Fayçal Hamdi, Laurent Bihanic, Gabriel Képéklian, Franck Cotton, Jérôme Euzenat, Zhengjie Fan, Pierre-Yves Vandenbussche, and Bernard Vatant. Enabling linked-data publication with the datalift platform. In *26th Conference on Artificial Intelligence (AAAI-12)*, 2012.

[28] Max Schmachtenberg, Christian Bizer, and Heiko Paulheim. Adoption of the linked data best practices in different topical domains. In Peter Mika, Tania Tudorache, Abraham Bernstein, Chris Welty, Craig Knoblock, Denny Vrandečić, Paul Groth, Natasha Noy, Krzysztof Janowicz, and Carole Goble, editors, *The Semantic Web - ISWC 2014*, volume 8796 of *Lecture Notes in Computer Science*, pages 245–260. Springer International Publishing, 2014. doi:10.1007/978-3-319-11964-9_16.

[29] Steffen Stadtmüller, Andreas Harth, and Marko Grobelnik. Accessing information about linked data vocabularies with vocab.cc. In Juanzi Li, Guilin Qi, Dongyan Zhao, Wolfgang Nejdl, and Hai-Tao Zheng, editors, *Semantic Web and Web Science*, Springer Proceedings in Complexity, pages 391–396. Springer New York, 2013. doi:10.1007/978-1-4614-6880-6_34.

[30] Maria del Carmen Suárez-Figueroa. *NeOn Methodology for Building Ontology Networks: Specification, Scheduling and Reuse*. PhD thesis, Universidad Politecnica de Madrid, Spain, June 2010. http://oa.upm.es/3879/.

[31] Pierre-Yves Vandenbussche and Bernard Vatant. Metadata recommendations for linked open data vocabularies. Technical report, 2012.

[32] Serena Villata and Fabien Gandon. Licenses compatibility and composition in the web of data. In Juan F. Sequeda, Andreas Harth, and Olaf Hartig, editors, *Proceedings of the Third International Workshop on Consuming Linked Data, COLD 2012, Boston, MA, USA, November 12, 2012*, number 905 in CEUR Workshop Proceedings, Aachen, 2012.

[33] P. L. Whetzel, N. F. Noy, N. H. Shah, P. R. Alexander, C. Nyulas, T. Tudorache, and M. A. Musen. BioPortal: enhanced functionality via new web services from the national center for biomedical ontology to access and use ontologies in software applications. *Nucleic Acids Research*, 39(suppl):W541–W545, jun 2011. doi:10.1093/nar/gkr469.