Web Semantics: Science, Services and Agents on the World Wide Web 57 (2019) 100492

Contents lists available at ScienceDirect

Web Semantics: Science, Services and Agents on the World Wide Web

journal homepage: www.elsevier.com/locate/websem

Why are ontologies not reused across the same domain?

Mariano Fernández-López [a,*], María Poveda-Villalón [b,*], Mari Carmen Suárez-Figueroa [b], Asunción Gómez-Pérez [b]

[a] Escuela Politécnica Superior, Universidad San Pablo CEU, Spain
[b] Escuela Técnica Superior de Ingenieros Informáticos, Universidad Politécnica de Madrid, Madrid, Spain

ARTICLE INFO

Article history:
Received 19 December 2017
Received in revised form 15 October 2018
Accepted 18 December 2018
Available online 27 December 2018

Keywords:
Ontology reuse
Ontology reliability
Ontology heterogeneity
Percentage of reuse
Reuse patterns
Drawbacks for reuse

ABSTRACT

Even though one of the main characteristics of ontologies has always been claimed to be their reusability, throughout this paper it will be shown that ontology reuse across a given domain is not a consolidated practice. We have carried out a statistical study on ontology reuse in the ontologies collected in Linked Open Vocabularies (LOV), in addition to a particular analysis of a use case. The results of the present work show that, when building an ontology, the heterogeneity between the needed conceptualization and that of available ontologies, as well as the deficiencies in some of such ontologies (concerning documentation, licensing, etc.) are important obstacles for reusing ontologies of the same domain of the ontology under development. A possible approach to lessen these problems could be the creation of communities similar to open software ones in charge of developing and maintaining ontologies.

1. Introduction

The reuse potential of ontologies and its benefits has been one of the main claims for ontology development since the first studies in the ontological field ([1–3], etc.). This vision and argument has driven research to reach the point in which ontology registries (e.g. BioPortal[1] and LOV[2]) are available where ontologies are published regardless the applications in which they are meant to be used and the data they are supposed to annotate [4]. However, long experience in ontological engineering led the authors to suspect that ontology reuse within a given domain does not seem to be widespread. This intuition has been reinforced by the work carried out by Kamdar and colleagues [5], focused on BioPortal, which states, as one of their conclusions, that the reuse among biomedical ontologies is quite limited.

In addition, as presented in [6], ontology reuse is often not performed by means of importing the reused ontologies (hard reuse), but just by referencing the reused ontology elements URIs (soft reuse). Fig. 1 shows the difference between both ways of reuse. For example, given that *OntoC:E3* has been reused by means of soft reuse, the constraints established on it in *OntoC* are outside *OntoA*, and therefore, it is not possible to reason with them when reasoning with *OntoA*. However, in the case of hard reuse, all the triples belonging to *OntoB* as well as all the triples belonging to the ontologies imported by this ontology belong to *OntoA*, Thus, when reasoning with this ontology, such triples are also taken into account. Note that, given the transitivity property of importation, if *OntoB* imports another ontology, the triples in the latter become triples of *OntoA* as well.

Advocates of the soft reuse claim that the imported ontology may become unavailable on the web. It is also argued that, frequently, the union of the reused ontology and the reusing one is inconsistent if a hard reuse is performed.[3] Note that if the closure of *OntoA* and *OntoC* is inconsistent, while each ontology is consistent, and a soft reuse of terms of *OntoC* is carried out in *OntoA*, then no ontology could import both ontologies at the same time without provoking an inconsistency.

There are intermediate situations between hard and soft reuse. For example, a sub-tree of the reused ontology may be copied in the reusing one. In contrast to hard reuse, the update of the reused ontology is not propagated to the reusing one. Katsumi and Grüninger's [7] work is useful to understand the different situations that may arise in reuse. These authors define reuse as the application of a series of operations to the ontology to be reused with the purpose of axiomatizing the intended conceptualization. The operations identified by Katsumi and Grüninger are presented below:

* Corresponding authors.
E-mail addresses: mfernandez.eps@ceu.es (M. Fernández-López), mpoveda@fi.upm.es (M. Poveda-Villalón), mcsuarez@fi.upm.es (M.C. Suárez-Figueroa), asun@fi.upm.es (A. Gómez-Pérez).

1 http://bioportal.bioontology.org/.
2 https://lov.okfn.org/dataset/lov/.
3 https://lists.w3.org/Archives/Public/public-lod/2017Jan/0045.html.

https://doi.org/10.1016/j.websem.2018.12.010
1570-8268/

1. *as_is*: reuse of an ontology without modifying it. This operation is equivalent to hard reuse without modifying the imported ontology.
2. *extraction*: reuse by removing some original axioms. This operation is equivalent to hard reuse removing some axioms. In the extreme case, that is, if all the axioms are removed, and just some terms are taken to be reused, the operation is equivalent to soft reuse.
3. *extension*: reuse by adding new axioms to the ontology. This operation is equivalent to hard reuse adding new axioms to the imported ontology.
4. *combination*: reuse by merging several ontologies. Let there be ontologies $O_1, O_2, \ldots O_n$. The combination of them consists in creating a new ontology O_{n+1}, applying hard reuse from O_{n+1} by importing the rest of the ontologies, and, if needed, harmonize them so that they form a single ontology.

Another way of reuse is the extraction of Ontology Design Patterns (ODPs) [8]. For example, it is possible not to reuse any SSN term, but to reuse its Sensor, Observation, Sample, and Actuator (SOSA) pattern. This case of reuse is outside the scope of this paper.

Based on our experience, we also detected a set of ontologies (FOAF,[4] Dublin Core,[5] etc.) that are reused in most of the ontologies. Along this work, it was checked whether such a phenomenon takes place in LOV ontologies.

Consequently, and in order to provide an analysis of the ontology reuse, the following objectives were defined:

- **Obj 1: Measure reuse, hard reuse and soft reuse**: the questions associated with this objective are the following: what is the proportion of entities reused by another ontology (reuse)?; what is the proportion of imported entities (hard reuse)?; what is the proportion of neither imported nor defined entities (soft reuse)? Do these proportions vary depending on whether they are calculated focusing just on classes or just on properties? Is there a set of ontologies that is reused in most of LOV ontologies?
- **Obj 2: Measure reuse across the same domain**: this objective is focused on the question: what is the proportion of ontologies that reuse other ontologies belonging to the same domain ?
- **Obj 3: Identify factors that hamper reuse**: given that the essential feature of ontologies is their reuse potential, it is critical to answer the following question: what are the main obstacles when reusing ontologies across the same domain?

Objective 1 is an update on Poveda and colleague's work [6], which is mainly focused on hard and soft reuse, but it does not deepen in the rest of the objectives.

LOV [9] is a general purpose ontology registry, and, as such, it includes ontologies from very different domains. Before adding a new ontology into the registry, LOV curators carry out a verification process in order to guarantee that it contains human and machine readable information, that some minimal metadata is provided, etc. By July 2016, LOV had registered around 500 ontologies in different domains . Therefore, throughout this study LOV is considered to be a good source to acquire the data needed in order to reach the above-mentioned objectives 1 and 2, and to provide ontologies to analyze in order to reach Objective 3. The data extraction from LOV was performed by an ad-hoc developed Java code based on the OWL-API.[6] These data were afterwards aggregated with the statistic suite R.[7] To achieve Objective 3, a particular use case was defined, involving (1) the development of an ontology in the academic domain and (2) the identification of the main obstacles for the reuse of LOV ontologies of this domain. Then, taking data from all LOV ontologies, the extend to which the conclusions of this use case are generalizable was studied.

The remaining sections of this paper are organized as follows. Section 2 dwells on the research methodology we followed for the study. Section 3 provides the obtained results during the study while Section 4 offers a discussion of such results. Section 5 focuses on related work. Finally, conclusions and future trends are outlined in Section 6.

2. Research methodology

An empirical approach was followed to reach the objectives mentioned above. Since objectives 1 and 2 can be approached in a quantitative way, they were addressed in statistical terms. In contrast, Objective 3 required a qualitative approach and, therefore, was achieved basing on the authors' experience in a real ontology development use case. Once the use case has been analyzed, the generalization study was carried out computationally checking if the obstacles found in candidate ontologies to reuse in the use case were present in other LOV ontologies.

For objectives 1 and 2, the particular steps carried out are presented below:

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

The following sections (from Section 2.1 to Section 2.3) show how each one of these steps were performed. The last section, Section 2.4 shows the methodology followed to attain Objective 3.

2.1. Definition of variables

To carry out this step, authors took into account that, according to their experience, reuse is different in classes and in properties. This led to a distinction between reuse in classes, object properties and data properties. The variables defined below are evaluated for each ontology.

The following variables represent the size of the ontology:

- Number of classes (c).
- Number of object properties (op).
- Number of data properties (dp).
- Total number of entities ($total = c + op + dp$).

The following variables account for the entities that had been hard reused in the ontology:

- Number of imported classes (ic).
- Number of imported object properties (iop).
- Number of imported data properties (idp).
- Total number of imported entities ($ti = ic + iop + icp$).

These other variables account for the entities that had been soft reused in the ontology:

- Number of neither defined nor imported classes ($ndnic$). They are the classes that are in the ontology by means of soft reuse.
- Number of neither defined nor imported object properties ($ndnop$).
- Number of neither defined nor imported data properties ($ndndp$).
- Total number of neither defined nor imported entities ($tndni = ndnic + ndnop + ndndp$).

[4] http://xmlns.com/foaf/spec/.
[5] http://dublincore.org/.
[6] http://owlapi.sourceforge.net/.
[7] https://www.r-project.org/.

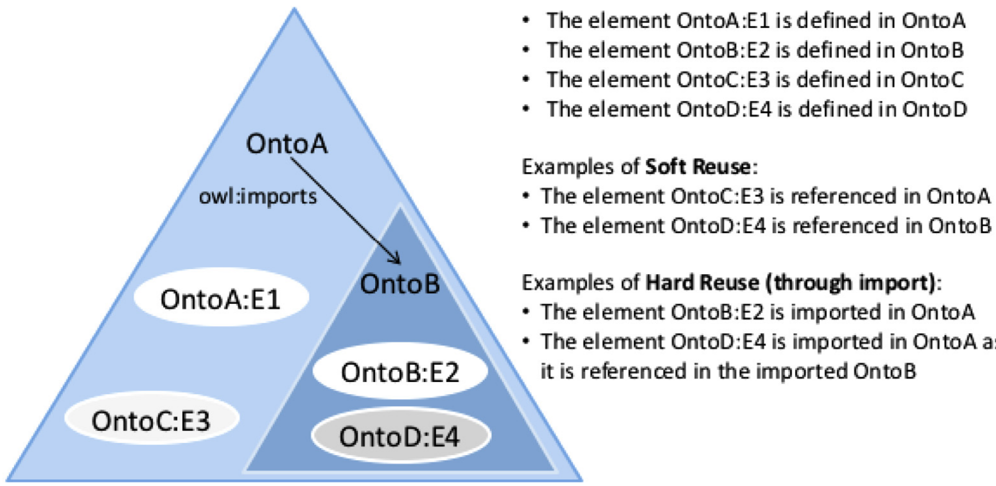

Fig. 1. Soft and Hard reuse.

For the total reuse, the following variable was defined:

- Total number the reused entities ($total\_reuse = ti + tndni$).

Out of the former variables, the following percentages of reuse are calculated:

- Percentage of reused classes ($per\_reuse\_c = (ic + ndnic) * 100/c$).
- Percentage of reused properties ($per\_reuse\_p = (iop + ndnop + iop + ndnop) * 100/(op + dp)$).
- Percentage of reused object properties ($per\_reuse\_op = (iop + ndnop) * 100/op$).
- Percentage of reused data properties ($per\_reuse\_dp = (idp + ndndp) * 100/dp$).
- Percentage of reused entities ($per\_reuse = total\_reuse * 100/total$).
- Percentage of hard reuse ($per\_i = ti/total\_reuse$)

The specific variables for Objective 3 are applied to the whole population of ontologies, and they are those shown below:

- Number of ontologies that reuse ontologies of the same domain ($total\_reuse\_same\_domain$).
- Percentage of ontologies that reuse ontologies of the same domain ($per\_reuse\_same\_domain = total\_reuse\_same\_domain * 100/total\_reuse$).

Let us note that denominators may be zero for some ontologies. In such a case, the percentages are not defined. Note also that the percentages of reuse do not necessarily show which of the reused entities are relevant for the future use of the ontology. This fact specially applies to hard reuse. That is, it is possible that *OntoA* imports *OntoB*, but most of *OntoB* entities are not present in applications based on *OntoA*. For example, *univ-ontology*,[8] a vocabulary about universities, imports FOAF. The only FOAF terms appearing in *univ-ontology* definitions are just *foaf:Person*, *foaf:firstName* and *foaf:lastName*. Hence, to calculate the percentage of reuse, an option would be to consider only these three terms as reused from FOAF. In fact, *foaf:yahooChatID*, also belonging to FOAF, but referred in no *univ-ontology* definition, could be considered as not relevant. However, all the terms of FOAF (including *foaf:yahooChatID*) really belong to *univ-ontology* and, consequently it is possible to reason with them when this ontology is queried.

8 https://github.com/ontop/ontop-examples/blob/master/university/univ-ontology.ttl (not registered in LOV).

2.2. Evaluation of the variables by means of data collection

The variables were evaluated for each ontology registered in LOV. To do so, an application to extract ontology features as well as to report errors when loading the ontologies was developed by using JAVA as implementation language.[9] Fig. 2 shows its inputs and outputs, as well as its interaction with other tools that also provide support for the analysis. The tools are depicted with rounded corners, whereas the inputs and the outputs are marked with square corners for documents and the database symbol for datasets. The different parts of the figure are presented below:

a. *Namespace-URI table (ns_uri_table)*. It is a spreadsheet that, for each LOV ontology, contains its prefix, its URI and its namespace.[10] Note that an ontology (for example, SIO[11]) may have the namespace different to the URI. The URI allows searching for the ontology in the Web.
b. *Frozen SPARQL storage*.[12] The purpose of this component is to allow for the process to be repeated independently of the ontologies and metadata that are removed from or added to LOV.
c. *File in R format*. This file[13] contains, for each ontology registered in the frozen SPARQL storage, the values of the features described above (c, op, dp, $total$, ic, iop, idp, $nindc$, $nindop$ and $nindp$). It can be processed by the R statistical tool.
d. LaTeX *file*. This file enables the generation of a PDF document with information that eases the manual analysis of the obtained results. The information provided in this file includes, for each ontology, data about metrics (annotation used, reused elements, etc.). In addition, the document includes information about exceptions raised during the ontology loading process, that is, a loading errors log.[14]

9 The java files are available at http://www.oeg-upm.net/files/mpoveda/JWS2017-Reuse/reuseOntologyAnalysis.zip.

10 The spreadsheet is available at http://www.oeg-upm.net/files/mpoveda/JWS2017-Reuse/Ns_vs_Uri_frozen.ods.

11 Its URI is http://semanticscience.org/ontology/sio.owl where its namespace is http://semanticscience.org/resource/.

12 http://ontoreuse.linkeddata.es/sparql (the dataset was frozen the 5th of July of 2016).

13 The R file is available at http://www.oeg-upm.net/files/mpoveda/JWS2017-Reuse/reuseOntologyAnalysis.zip.

14 The PDF file is available at http://www.oeg-upm.net/files/mpoveda/JWS2017-Reuse/OntologiesAnalysis.pdf.

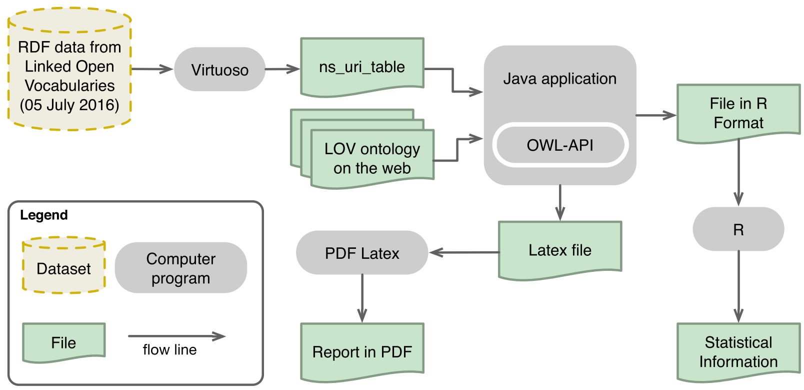

Fig. 2. The process followed to obtain the statistical information and the complementary report.

Aggregated measures of the whole population (maximum, minimum, quartiles and mean) were computed by means of the R statistical tool. The population was previously cured by removing ontologies with *c*, *op* and *dp* equals to 0. Ontologies with $per\_reuse = 100\%$ were deleted as well. There have also been cases of ontologies that had could not be properly loaded.

This research also tackles the question whether there exists a well characterized set of ontologies reused by most LOV ontologies. If the answer is *yes*, then, following the authors' experience, such ontologies are expected to be those whose URI is defined in the *http://www.w3.org/* namespace, and those identified by W3C as good ontologies.[15] Some of the ontologies belonging to the list of good ontologies are FOAF, Dublin Core, SIOC,[16] etc. Henceforth, both *W3C* and good ontologies will be called *W3C endorsed ontologies*. Thus, one of the tasks to carry out is to measure the percentage of LOV ontologies that reuse them.

To achieve Objective 2, the authors had to obtain the number of ontologies that reuse other ontologies of the same domain (*total_reuse_same_domain*) as well as its percentage with respect to the total number of ontologies that reuse (*per_reuse_same_domain*). To be able to automatize the process of determining if two ontologies belong to the same domain, the notion of LOV category has been taken as a reference. The category of each ontology is already assigned in LOV, and it has been an input data in this study. Examples of LOV categories are *Biology*, *Geography*, *Academy*, etc. There are ontologies that may be annotated with two different categories. Each LOV category was assumed to be a domain, except for the following categories: *Metadata*, *RDF* and *W3C Rec*. *Metadata* ontologies are used to annotate other ontologies. An example of an ontology belonging to this category is the Asset Description Metadata Schema (ADMS),[17] which provides terms that allow annotating code lists, taxonomies, dictionaries, vocabularies, etc. The fact that an ontology belonging to the *Metadata* category reuses another one of this category may not be due to the fact that both of them belong to the same domain. Concerning *RDF* and *W3C Rec*, they are not homogenous categories that can be considered domains. In spite of this, 27 ontologies defined in the *http://www.w3.org/* namespace that belong to other LOV categories apart from *W3C Rec* remained in the set of ontologies to study.

The steps presented below were carried out by using R:

1. The population of ontologies was reduced to those that reuse.
2. The population of ontologies was again reduced by removing those belonging to the three categories that were not considered domains.
3. The variables *total_reuse_same_domain* and *per_reuse_same_domain* were automatically evaluated by processing the resultant population of the previous step.

Modularization is an approach already followed in the development of traditional software. Considering the import of a module as a case of reuse across the same domain does not show the capability that ontologies have for being reused. Consequently, in step 3, for each pair of reusing and reused ontology, the URIs of both of them were compared to determine whether it had been a case of modularization and remove it. The number of LOV ontologies resulting from step 2 and that reused other ontologies apart from its modules was 281.

2.3. Data analysis

Given that the study has been applied to the whole collection of LOV ontologies, the data analysis was carried out by means of descriptive statistics. In some cases, the numbers were meaningful by themselves. In other cases, box plots were used. For example, to answer the question about the variation of the percentages of reuse according to the different kind of entities (Objective 1), box plots showing the percentages of reuse of classes, object and data properties were generated. These diagrams are presented in a single image in Section 3 and they allowed to appreciate how different the percentages of reuse were.

2.4. Methodology for objective 3: identify factors that hamper reuse

This part of the research had two phases. The first one consisted in the development of a use case to identify factors that may influence or hamper ontology reuse across the same domain. The second one involved the automatic analysis of the whole population of ontologies to determine to what extend the results of the use case are applicable to other cases.

To carry out the first phase, an ontology was developed solely for this purpose. The ontology is intended to specify, in a formal way, a conceptualization for the exchange of information according to the Practical Guide for the Evaluation and Monitoring of Official Degrees of the Madri+d Foundation [10]. It covers the information on official degrees that has to be public according to the Madri+d Foundation. This guide is applicable to all fourteen universities of Madrid.[18] The knowledge sources used during this task are listed below:

- The Spanish series of Royal Decrees on Higher Education [11–16].

15 https://www.w3.org/wiki/Good_Ontologies.

16 http://sioc-project.org/.

17 http://www.w3.org/ns/adms.

18 http://www.emes.es/Sistemauniversitario/UniversidadesdeMadrid/tabid/215/Default.aspx.

- The aforementioned Madri+d guide.
- The guide for accreditation of degrees of the National Agency for Quality Assessment and Accreditation of Spain (ANECA) [17].
- An experienced coordinator of a degree in Information Systems.

In addition to providing a basis for a real-case scenario of ontology reuse, the ontology, called OTN (***O**ntología de **T**ítulos u**N**iversitarios*[19]), can be used to migrate the mandatory public information of Spanish university websites to Linked Open Data, enabling therefore the integration of data coming from the different institutions of higher education. In this way, the development of a decision support system to help future students in deciding into which university course to enroll would not require particular scrapers for each university website.

The 10 ontologies registered in LOV in the category *Academy* were analyzed to check whether they could be reused in OTN.

Once the results of the first phase were known, the ontology features to be observed for the second phase were defined, evaluated and analyzed.

3. Results

This section presents the results achieved towards each objective.

3.1. Results concerning objective 1 "measure reuse, hard reuse and soft reuse" and objective 2 "measure reuse across the same domain"

Consulting the section *Frequencies of exceptions during the loading of ontologies* of the PDF report[20] generated by the Java tool, see Fig. 3, it can be observed that 386 out of the 558 ontologies available in the frozen SPARQL could be loaded. The list below summarizes the main problems that arose during ontology processing along with the number of ontologies affected by them:

1. Timeout expiration during ontology loading (after 120 s) without arising any other exception. (21 ontologies)
2. `UnparsableOntologyException`: This exception means that the ontology is not processable due to syntax errors. (46 ontologies)
3. `OWLOntologyCreationIOException`: This exception arises when the ontology is not longer available. (70 ontologies)
4. `UnloadableImportException`: This exception happens when some of the ontologies imported by the ontology could not be loaded. (34 ontologies)
5. `Exception`: This a generic exception. The precise cause has not been determined by the tool. (1 ontology)

After the curation of the population explained in Section 2, the number of ontologies was reduced to 355.

Concerning the evaluation of variables applicable to each ontology of the population, the aggregated measures are presented in Fig. 4.

Fig. 5 shows the box plots for the percentages of global reuse (*per_reuse*), reuse of classes (*per_reuse_c*), reuse of object properties (*per_reuse_op*) and reuse of data properties (*per_reuse_dp*).

With regard to the reuse of W3C endorsed ontologies, a 78.03% of the ontologies of the population reuse them.

Concerning reuse across the same domain, the population of ontologies belonging to none of the categories *Metadata*, *RDF* and *W3C Rec*, and that reused other ontologies apart from their modules, had 281 ontologies. Of these, 56 (19.92%) reused ontologies belonging to the same LOV category.

3.2. Results concerning objective 3. factors against reuse across the same domain

3.2.1. Development of the use case

The following paragraphs show the problems found when trying to reuse each one of the ontologies classified in LOV under the *Academy* category during the development of the OTN ontology.

When the option of reusing AIISO (Academic Institution Internal Structure Ontology)[21] was considered, the following drawbacks were detected:

1. *Heterogeneity between the concepts of the OTN domain and the concepts represented in AIISO*. Let us note that the fact that the candidate ontology domain is complementary to the ontology being built to axiomatize the intended conceptualization is not a problem. For example, in Spain, the notion of College does not exist, therefore it is not necessary to complement the OTN with AIISO as it is not required to represent such concept.
2. *Heterogeneity in the natural language used*. The translation from English to Spanish had required an additional time in the process of reuse. This has been a problem shared by all LOV ontologies belonging to the *Academy* category.

From the case of Courseware,[22] the following drawbacks arose:

1. *Deficiencies in the documentation*. The ontology is just available in XML. An explanatory HTML document could not be consulted.
2. *Lots of information due to an imported ontology that is not available*. When the aurhors tried to open the ontology with Protégé,[23] an error message appeared, although the ontology had been loaded. Nevertheless, all terms imported from AKTORS (the portal ontology AKT Reference Ontology) have neither formal definition nor natural language definition, since AKTORS is no longer available.
3. *Unavailable license*. We have not found any license for this ontology. Therefore, even in the case of that the ontology had contents appropriate for OTN, permission to reuse it would have been needed.

In the case of myExperiment,[24] again, just a XML file was available, although it could be loaded in Protégé without any problem. However, the ontology was focused on content management, social networking and object information, which is not in the scope of the first version of OTN. Besides, neither does it have license. An analogous situation is the one of SCI People (Scientific People Ontology),[25] that is, just a XML file is available, which is loadable in Protégé, it does not have license and, moreover, its domain does not have enough overlap with the domain of OTN to compensate the effort of reusing it.

Other ontologies whose domains do not have enough overlap with the domain of OTN are FRAPO[26] (Research Administration and

[19] The Academic Degree Ontology (***O**ntología de **T**ítulos u**N**iversitarios*) is available at https://w3id.org/def/otn.

[20] See http://www.oeg-upm.net/files/mpoveda/JWS2017-Reuse/OntologiesAnalysis.pdf.

[21] http://vocab.org/aiiso/schema.

[22] http://courseware.rkbexplorer.com/ontologies/courseware.

[23] http://protege.stanford.edu/.

[24] http://rdf.myexperiment.org/ontologies/base/.

[25] http://lod.taxonconcept.org/ontology/sci_people.owl.

[26] http://www.sparontologies.net/ontologies/frapo/source.html.

1 Frequencies of exceptions during the loading of ontologies

558 ontologies have been tried to be loaded. Below, the absolute frequency of the different exceptions loading ontologies is displayed:

1. Number of ontologies with timeout expiration during their loading (after 120 seconds): 21
2. Number of ontologies with `UnparsableOntologyException`: 46
3. Number of ontologies with `OWLOntologyCreationIOException`: 70
4. Number of ontologies with `UnloadableImportException`: 34
5. Number of ontologies with `OWLOntologyAlreadyExistsException`: 0
6. Number of ontologies with `OWLOntologyCreationException`: 0
7. Number of ontologies with just `Exception`: 1

For the 386 ontologies that have been finally loaded, metrics have been provided. Additionally, a file in R format with these metrics have been generated.

2 Metrics and failures of the particular ontologies

2.1 org

2.1.1 Annotations

```
<http://purl.org/dc/terms/modified>: "2014-02-05"^^xsd:date
rdfs:label: "Ontologia delle organizzazioni"@it
<http://purl.org/dc/terms/contributor>: _:genid30757
<http://purl.org/dc/terms/modified>: "2012-10-06"^^xsd:date
<http://purl.org/dc/terms/modified>: "2010-10-08"^^xsd:date
<http://purl.org/dc/terms/contributor>: _:genid30758
rdfs:comment: "Vocabulario para describir organizaciones, adaptable a una amplia
    variedad de ellas."@es
<http://purl.org/dc/terms/contributor>: _:genid30759
<http://purl.org/dc/terms/modified>: "2010-06-09"^^xsd:date
<http://purl.org/dc/terms/title>: "Ontolog\'ia de organizaciones"@es
<http://purl.org/dc/terms/license>: http://www.opendatacommons.org/licenses/pddl
    /1.0/
rdfs:label: "Ontologie des organisations"@fr
rdfs:comment: "Vocabulary for describing organizational structures, specializable to
     a broad variety of types of organization."@en
<http://purl.org/dc/terms/modified>: "2012-09-30"^^xsd:date
rdfs:label: "Ontolog\'ia de organizaciones"@es
rdfs:seeAlso: http://www.w3.org/TR/vocab-org/
<http://purl.org/dc/terms/created>: "2010-05-28"^^xsd:date
<http://purl.org/dc/terms/contributor>: _:genid30765
<http://purl.org/dc/terms/modified>: "2014-01-25"^^xsd:date
<http://purl.org/dc/terms/modified>: "2013-12-16"^^xsd:date
<http://purl.org/dc/terms/modified>: "2014-01-02"^^xsd:date
<http://purl.org/dc/terms/modified>: "2013-02-15"^^xsd:date
<http://purl.org/dc/terms/contributor>: _:genid30760
<http://purl.org/dc/terms/contributor>: _:genid30761
```

40

Fig. 3. Screenshot of the PDF report about the processing of LOV ontologies.

		Minimum	First Quartile	Median	Mean	Third Quartile	Maximum
Ontology Elements	Number of Classes (c)	1.00	9.00	18.00	68.60	55.50	3040.00
	Number of Object Properties (op)	0.00	6.00	17.00	52.91	52.50	2529.00
	Number of Datatype Properties (dp)	0.00	1.00	7.00	31.48	24.00	4504.00
	Number of Entities (total = c + op + dp)	2.00	23.50	53.00	153.00	133.00	8626.00
Hard Reuse	Number of Imported Classes (ic)	0.00	0.00	0.00	16.03	0.00	1768.00
	Number of Imported Object Properties (iop)	0.00	0.00	0.00	15.06	0.00	378.00
	Number of Imported Datatype Properties (idp)	0.000	0.000	0.000	5.718	0.000	148.000
Soft Reuse	Number of Classes (ndnic) that are neither defined nor imported	0.000	1.000	3.000	5.823	6.000	266.000
	Number of Object Properties (ndnop) that are neither defined nor imported	0.000	0.000	0.000	4.521	1.000	366.000
	Number of Datatype Properties (ndndp) that are neither defined nor imported	0.000	0.000	0.000	1.715	0.000	177.00
Percentages	Imported Entities (per_i) (Hard Reuse)	0.000	0.000	0.000	17.615	6.161	100.000
	Entities that are neither defined nor imported (per_ndni) (Soft Reuse)	0.000	0.9093	9.5696	18.0373	25.8333	100.000
	Reused Entities (per_reuse) (Hard and Soft Reuse)	0.00	7.69	21.43	30.40	46.78	99.57
	Reused Classes (per_reuse_c)	0.00	13.90	38.46	41.15	65.33	100.00
	Reused Properties (per_reuse_p)	0.000	0.000	5.882	23.249	39.056	100.00
	Reuse Object Properties (per_reuse_op)	0.000	0.000	4.058	23.137	35.893	100.00
	Reuse Datatype Properties (per_reuse_dp)	0.000	0.000	3.571	29.233	57.143	100.00

Fig. 4. Statistics of reuse, hard reuse and soft reuse.

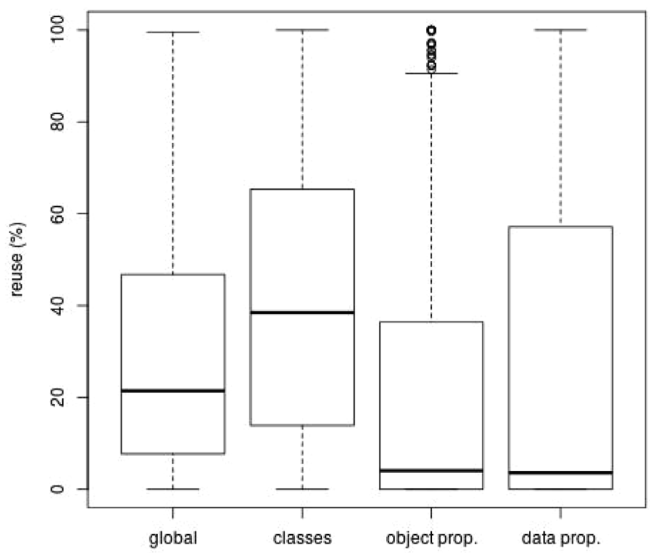

Fig. 5. Box plots for global reuse (*per_reuse*), reuse of classes (*per_reuse_c*), reuse of object properties (*per_reuse_op*) and reuse of data properties (*per_reuse_dp*).

Projects Ontology), SCORO[27] (Scholarly Contributions and Roles Ontology) and SWC[28] (Semantic Web Conference Ontology).

Teach[29] is an ontology that could be useful in a long term. However, it is more appropriate for the execution of an academic program than for its definition. Thus, for example, it contained entities like *Assignment*, *reportTemplate*, *room*, etc.

Finally, Ontoware[30] is not available online anymore.

Summarizing, due to all the previous issues, none of the above-mentioned ontologies were reused during the development of OTN.

However, domain independent ontologies have been reused. DUL[31] and DOLCE-Zero[32] were translated into Spanish for the purpose of this work[33] and were reused in the development of OTN.

3.2.2. Generalization of the use case

Below are the factors that have hampered the reuse in the use case. We also elaborate on the extent to which this particular use case can be generalized to other LOV ontologies.

1. *Heterogeneity in the natural language used*. LOV provide information above the natural language used to develop the ontology. From the result of a SPARQL query, the percentage of ontologies that use languages apart from English was calculated, which amount to (27.04%).
2. *Deficiencies in the documentation*. The *Curl* tool allowed to check each ontology for its corresponding documentation. Just 17.18% of them provide such a documentation.
3. *Lost of information due to an imported ontology that is not available*. As mentioned in Section 3.1 the complete population of LOV had to be curated due to a subset not being able to load.
4. *Unavailable license*. For each ontology, it has been checked whether it was licensed according to some of the following descriptions: *dct:license*, *dc:rights*, *dcterms:rights*, *dcterms:license*, *cc:license* or *xhv:license*. The resultant percentage is 60.28%. It should be mentioned that this figure indicates the ontologies in which a license is stated, fact that does not imply that the specified license is open enough to allow their reuse. However, that analysis is out of scope of this paper.
5. *Heterogeneity between the subdomains of the reusing and the reused ontologies*. The generalization of this factor remains for future work.

27 http://www.sparontologies.net/ontologies/scoro/source.html.
28 http://data.semanticweb.org/ns/swc/swc_2009-05-09.html.
29 http://linkedscience.org/teach/ns/.
30 http://www.ontoware.org/index.html.
31 http://www.ontologydesignpatterns.org/ont/dul/DUL.owl.
32 http://www.ontologydesignpatterns.org/ont/d0.owl.
33 https://w3id.org/def/dul-dolce-zero-en-espannol.

3.2.3. Global considerations

Given that the study was applied to a whole population of ontologies (LOV), the results have not been expressed in terms of significance. As LOV was not proved to be representative with regard to the rest of repositories, it should not be considered as a sample. Therefore, the results of this paper are strictly only valid for LOV, in the same way as Kamdar and colleagues work [18] is strictly valid for BioPortal.

4. Discussion

One of the most outstanding facts observed during the study is that 36.38% of the ontologies registered in LOV could not be appropriately loaded.

Concerning how much reuse is carried out during ontology development, the aggregated measures point out that the percentage of reused entities (*per_reuse*) have a mean of 30.40% and a median of 21.43%. Regarding reuse of W3C endorsed ontologies, 78.02% of ontologies reuse them. As it can be appreciated in Figs. 4 and 5, reuse in classes is higher than in properties. Concerning hard reuse, the median of the percentage of imported entities (*per_i*) is 0%, whereas its third quartile is 6.161%. That is, more than a half of the ontologies that reuse others do not import any entity, and more than 3/4 import, at most, a 6% of their entities. It is true that soft reuse prevents problems when the reused ontology is not loaded. Thus, the fact that a soft reused ontology is no longer available does not affect the reusing ontology. Nevertheless, this procedure of reuse loses the axioms that relate the reused entities to the rest of the entities of its source ontology. Moreover, if, in the reusing ontology, knowledge with contradictions to the reused ontologies are added, such contradictions are not automatically detected. If both ontologies, in turn, are imported by the same ontology, such contradictions will emerge. This situation has not taken place in the author's experience, but it may arise when the two imported ontologies assume different world views. For example, according to DOLCE, one of the main characteristics of an abstract entity is that it has neither spatial nor temporal qualities. Examples of such abstract entities are mathematical ones: a triangle, a circumference, etc. Nevertheless, SUMO defines *social role* as a sub class of *abstract*. Let us note that a social role has both spatial and temporal qualities, since it may take place in a particular area along a particular time interval.

Let us now suppose that ontology O_1 imports SUMO and that O_2 imports DOLCE. Let us also suppose that O_1 imports O_2. In this case, O_1 includes contradictory definitions of the term *abstract*. Given that all the details of the natural language definitions are not formalized in the ontologies, it may be case that no inconsistency appears when a computer reasons with this ontology. However, there will be a contradiction in the intended meaning of the terms of O_1.

Something similar may happen if a redefinition of classes/properties in the reusing ontology [19,20] takes place. Let us suppose, for example, the following axiom:

```
reused:property rdf:subPropertyOf reusing:property
```

If *reusing:property* were declared as inverse functional, then *reused:property* would become inverse functional as well. If soft reuse had been carried out, then, formally, the only definition would be the one of the reusing ontology. However, if a third ontology imported both the reused and the reusing ones, there would be, in the third ontology, a formal definition from the reused ontology and another formal definition in the reusing ontology.

Concerning the information that was obtained through the OTN use case, several heterogeneity problems were detected. First of all, there are ontologies, in the academic domain, focused on different aspects of such domain but out of the scope of OTN ontology, for example content management, social networking, etc.. Some ontologies represent the same domain as OTN, but from a different perspective. Others do not use Spanish, being the lack of localization [21] a key factor that hampers reuse. In fact, it has been proved that just 27.04% of the LOV ontologies use languages other than English. Other reasons to not reuse LOV ontologies belonging to the Academy category have been deficiencies in the documentation of some candidate ontologies and unreachable imported ontologies.

5. Related work

One of the first analysis on ontology reuse was done by Simperl [22]. The author presents different use cases on ontology reuse as well as methods and tools. The main conclusion is that the reuse approach is a decision-making problem in which the developer experience plays a crucial role. Another interesting study has been performed by Ghazvinian and colleagues [23] with the aim of analyzing the orthogonality in ontologies at the OBO Foundry.[34] The idea of having orthogonal ontologies implies that each term is defined in only one ontology; when such a term is needed in other ontologies, the term should be referred from the original ontology. Authors analyzed the reuse in 53 ontologies in three different time points (September 2009, March 2010, and September 2010). The results indicate that reuse among the OBO Foundry candidates increased during the study period. Let us note that the reuse is carried out inside the same repository.

Poveda and colleagues [6] analyzed soft and hard reuse in the Linked Data context. This study was performed over 73.96% of the ontologies[35] included in the LOV registry. The study shows that the proportion of reused elements was considerably high (40.53%). In addition, in that time, hard reuse was more frequent than soft reuse, although soft reuse already represented 1/4 of the total. If compared to the results of the present work, that percentage shows an evolution in such a way that the proportion of soft reuse has increased until approximately a half of the total reuse. Also in the Linked Data context, Schaible and colleagues [24] performed a survey to obtain rankings of diverse modeling examples. Participants in the survey should take into account their understanding of good reuse of vocabularies when ranking the models. This study shows that the trend has been to reuse popular terms from frequently used ontologies instead of using terms from not popular ontologies. This result is consistent with the present work.

Focusing on the BioPortal repository, Kamdar and colleagues [18] analyzed the reuse in 377 biomedical ontologies. As a general conclusion, the reuse among biomedical ontologies is quite limited (less than 5%). Let us note that this percentage is much lower than that obtained by Poveda and colleagues and that obtained by us. Consequently, this reflects that the reuse percentage in LOV is a great deal higher than that of BioPortal. Likewise, these authors have detected that the knowledge reused belongs to a small set of popular ontologies [5]. Kamdar and colleagues [5] also discovered that, in many cases, developers use incorrect representations when reusing terms from different ontologies. Thanks to this study, a set of common antipatterns when reusing ontology terms has been identified. Ochs and colleagues [25] made an analysis of 355 ontologies in BioPortal. As a result of this analysis, it can be mentioned that 55% of the analyzed ontologies includes elements (classes or properties) from other ontologies. In addition, the work showed that knowledge reuse is specially widespread among the ontologies in the OBO Foundry.

6. Conclusions and future trends

While in software development it is common to reuse APIs, experienced engineers address the reuse with caution. More precisely, usually the reuse is done by means of commercial agreements with the company providing it or the reused software is supported and maintained by a reliable community.

In the case of ontologies, scarcely ever an ontology is maintained by a company that commercializes the reuse of the ontology. An ontology being maintained and updated by a community similar to those of free software is not frequent either. In consequence, a candidate ontology to be reused may not reach minimal adoption requirements or it may not be always available online (as shown in the results). In addition, many ontologies do not include an explicit declaration of their license [26], therefore, the fact that the ontology is actually available online does not imply that it can be legally reused. Along the development of the use case, this has been found to be an obstacle to reuse ontologies.

Another important issue detected concerns the assumption that both English, as the lingua franca in certain fields as research or computer science, and the conceptualizations from countries leading the progress in such fields, are the reference when developing ontologies. In practice, however, when an official organism or a company from a concrete country needs and ontology, they usually need it in the country's official language and it has to model the specific conceptualizations of the domain given. For example, focusing in the use case, even after the Bolonia[36] harmonization process carried out by the European Union, the Spanish university system still presents particular characteristics. In order for ontologies to be adapted to the specific necessities of different groups (communities), the projects to develop them have to be supplied of the extra resources necessary to develop localized ontologies.

From the previous paragraphs the following lines of work emerge as promising. First of all, a change of mentality is needed regarding ontology development and maintenance. In this sense, processes as those used by organizations like Apache[37,38] for software development could be taken as inspiration for ontology development. In consequence, research on collaborative development [27,28] is key in order to produce reusable ontologies. Secondly, it is crucial to continue the research being carried out on ontology localization [21] in order to identify those ontology modules independent of a region or country. Finally, tools for ontology modularization (e.g. Ontofox [29] or the Protégé functionality to merge axioms) and methods (e.g. NeOn approach [30] or MIREOT [31]) deserve special attention, as in practical settings they might represent the key for users to extract only those ontology submodules needed, or to import an ontology as a whole or not to reuse ontological resources at all. In any case, reusability does not come for free, and a realistic view should be taken assuming that the reusability degree will depend on the resources dedicated to such goal.

34 http://www.obofoundry.org/.

35 196 ontologies within the 265 registered at the time of performing the analysis.

36 http://eur-lex.europa.eu/legal-content/ES/TXT/?uri=URISERV%3Ac11088.

37 https://www.apache.org/foundation/getinvolved.html.

38 https://www.apache.org/dev/new-committers-guide.html.

Acknowledgment

This work has been supported by the project Datos 4.0: retos y soluciones (TIN2016-78011-C4-4-R), funded by Ministerio de Economía, Industria y Competitividad. Authors would also like to thank Julia Bosque-Gil for her invaluable comments and revision.

References

[1] T.R. Gruber, A translation approach to portable ontology specifications, Knowl. Acquis. 5 (2) (1993) 199–220, http://dx.doi.org/10.1006/knac.1993.1008.

[2] R. Neches, R. Fikes, T. Finin, T. Gruber, R. Patil, T. Senator, W.R. Swartout, Enabling technology for knowledge sharing, AI Mag. 12 (3) (1991) 36–56.

[3] R. Studer, V.R. Benjamins, D. Fensel, Knowledge engineering: principles and methods, Data Knowl. Eng. 25 (1–2) (1998) 161–197, http://dx.doi.org/10.1016/S0169-023X(97)00056-6.

[4] M. d'Aquin, N. Noy, Where to publish and find ontologies? a survey of ontology libraries, Web Semant. Sci. Serv. Agents World Wide Web 11 (2012) 96–111, http://dx.doi.org/10.1016/j.websem.2011.08.005, http://www.sciencedirect.com/science/article/pii/S157082681100076X.

[5] M.R. Kamdar, T. Tudorache, M. Musen, A systematic analysis of term reuse and term overlap across biomedical ontologies, Semantic Web Journal (in press).

[6] M. Poveda-Villalón, M.C. Suárez-Figueroa, A. Gómez-Pérez, The landscape of ontology reuse in linked data, in: Proceedings Ontology Engineering in a Data-driven World, OEDW 2012, 2012, oeg.

[7] M. Katsumi, M. Grüninger, What is ontology reuse?, in: Formal Ontology in Information Systems - Proceedings of the 9th International Conference, FOIS 2016, Annecy, France, July 6–9, 2016, 2016, pp. 9–22, https://doi.org/10.3233/978-1-61499-660-6-9.

[8] V. Presutti, E. Blomqvist, E. Daga, A. Gangemi, Pattern-based ontology design, in: Ontology Engineering in a Networked World, Springer, 2012, pp. 35–64.

[9] P.-Y. Vandenbussche, G.A. Atemezing, M. Poveda-Villalón, B. Vatant, Linked open vocabularies (lov): a gateway to reusable semantic vocabularies on the web, Semant. Web 8 (3) (2017) 437–452.

[10] Guía práctica para la evaluación del seguimiento de títulos universitarios oficiales de la fundación madri+d. http://www.madrimasd.org/uploads/acreditacion/doc/seguimiento2013/Guia_de_Evaluacion_2014.pdf. (Accessed 20 July 2018).

[11] Real Decreto 898/1985, de 30 de abril, sobre régimen del profesorado universitario. https://www.boe.es/buscar/pdf/1985/BOE-A-1985-11578-consolidado.pdf. (Accessed 20 July 2018).

[12] Ley Orgánica 6/2001, de 21 de diciembre, de universidades. http://www.boe.es/buscar/pdf/2001/BOE-A-2001-24515-consolidado.pdf. (Accessed 20 July 2018).

[13] Real Decreto 1125/2003, de 5 de septiembre, por el que se establece el sistema europeo de créditos y el sistema de calificaciones en las titulaciones universitarias de carácter oficial y validez en todo el territorio nacional. https://www.boe.es/boe/dias/2003/09/18/pdfs/A34355-34356.pdf. (Accessed 20 July 2018).

[14] Real Decreto 1393/2007, de 29 de octubre, por el que se establece la ordenación de las enseñanzas universitarias oficiales. https://www.boe.es/boe/dias/2007/10/30/pdfs/A44037-44048.pdf. (Accessed 20 July 2018).

[15] Real Decreto 1791/2010, de 30 de diciembre, por el que se aprueba el estatuto del estudiante universitario. https://www.boe.es/boe/dias/2010/12/31/pdfs/BOE-A-2010-20147.pdf (Accessed 20 July 2018).

[16] Real Decreto 861/2010, de 2 de julio, por el que se modifica el Real Decreto 1393/2007, de 29 de octubre, por el que se establece la ordenación de las enseñanzas universitarias oficiales. https://www.boe.es/boe/dias/2010/07/03/pdfs/BOE-A-2010-10542.pdf. (Accessed 20 July 2018).

[17] Guía de apoyo para la elaboración de la memoria de verificación de títulos oficiales universitarios (grado y máster). http://www.aneca.es/content/download/12155/136031/file/verifica_gm_guia_V05.pdf. (Accessed 20 July 2018).

[18] M.R. Kamdar, T. Tudorache, M.A. Musen, Investigating term reuse and overlap in biomedical ontologies, in: Proceedings of the International Conference on Biomedical Ontology, ICBO 2015, Lisbon, Portugal, July 27–30, 2015, 2015.

[19] A. Hogan, A. Harth, A. Polleres, Scalable authoritative owl Reasoning for the Web, Int. J. Semant. Web Inf. Syst. 5 (2) (2009) 49–90, http://dx.doi.org/10.4018/jswis.2009040103.

[20] C. Lutz, D. Walther, F. Wolter, Conservative extensions in expressive description logics, in: Proceedings of the 20th International Joint Conference on Artificial Intelligence, IJCAI'07, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2007, pp. 453–458, http://dl.acm.org/citation.cfm?id=1625275.1625347.

[21] J. Gracia, E. Montiel-Ponsoda, P. Cimiano, A. Gómez-Pérez, P. Buitelaar, J. McCrae, Challenges for the multilingual web of data, Web Semant. Sci. Serv. Agents World Wide Web (ISSN: 1570-8268) 11 (2011).

[22] E. Simperl, Reusing ontologies on the semantic web: a feasibility study, Data Knowl. Eng. 68 (10) (2009) 905–925, http://dx.doi.org/10.1016/j.datak.2009.02.002.

[23] A. Ghazvinian, N.F. Noy, M.A. Musen, How orthogonal are the obo foundry ontologies? J. Biomed. Semant. (2011).

[24] J. Schaible, T. Gottron, A. Scherp, Survey on common strategies of vocabulary reuse in linked open data modeling, in: V. Presutti, C. d'Amato, F. Gandon, M. d'Aquin, S. Staab, A. Tordai (Eds.), The Semantic Web: Trends and Challenges: 11th International Conference, ESWC 2014, Anissaras, Crete, Greece, May 25-29, 2014. Proceedings, Springer International Publishing, Cham, 2014, pp. 457–472, http://dx.doi.org/10.1007/978-3-319-07443-6_31.

[25] C. Ochs, Y. Perl, J. Geller, S. Arabandi, T. Tudorache, M.A. Musen, An empirical analysis of ontology reuse in bioportal, J. Biomed. Inform. 71 (2017) 165–177, http://dx.doi.org/10.1016/j.jbi.2017.05.021.

[26] M. Poblet, A. Aryani, P. Manghi, K. Unsworth, J. Wang, B. Hausstein, S. Dallmeier-Tiessen, C. Klas, P. Casanovas, V. Rodríguez-Doncel, Assigning creative commons licenses to research metadata: issues and cases, CoRR abs/1609.05700 (2016).

[27] O. Dyachenko, Y. Zagorulko, A collaborative development of ontology-based knowledge bases, in: P. Klinov, D. Mouromtsev (Eds.), Knowledge Engineering and the Semantic Web: 5th International Conference, KESW 2014, Kazan, Russia, September 29–October 1, 2014. Proceedings, Springer International Publishing, Cham, 2014, pp. 219–228, http://dx.doi.org/10.1007/978-3-319-11716-4_19.

[28] H. Pinto, C. Tempich, S. Staab, Ontology engineering and evolution in a distributed world using diligent, in: S. Staab, R. Studer (Eds.), Handbook on Ontologies, in: International Handbooks on Information Systems, Springer Berlin Heidelberg, 2009, pp. 153–176, http://dx.doi.org/10.1007/978-3-540-92673-3_7.

[29] Z. Xiang, M. Courtot, R.R. Brinkman, A. Ruttenberg, Y. He, OntoFox: web-based support for ontology reuse, BMC Res. Not. 3 (1) (2010) 175.

[30] M. d'Aquin, Modularizing ontologies, in: Ontology Engineering in a Networked World., 2012, pp. 213–233, http://dx.doi.org/10.1007/978-3-642-24794-1_10.

[31] M. Courtot, F. Gibson, A.L. Lister, J. Malone, D. Schober, R.R. Brinkman, A. Ruttenberg, Mireot: the minimum information to reference an external ontology term, Appl. Ontol. 6 (1) (2011) 23–33.