

Why are ontologies not reused across the same domain?

Mariano Fernández-López¹ María Poveda-Villalón²
Mari Carmen Suárez-Figueroa² Asunción Gómez-Pérez²

¹Escuela Politécnica Superior
Universidad San Pablo CEU

²Escuela Técnica Superior de Ingenieros Informáticos
Universidad Politécnica de Madrid

February 21, 2019

Overview

- 1 Introduction
- 2 Research methodology
- 3 Results and discussion
- 4 Conclusions
- 5 Future work
- 6 References

Introduction

Main idea

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. Moreover, some of the causes of this fact will be inferred and described.

Links

- <https://doi.org/10.1016/j.websem.2018.12.010>
- Free version
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3315457
- Twitter <https://twitter.com/PAlexop/status/1090643380034568192>



Image attributions:

- <https://en.wikipedia.org/w/index.php?curid=25688425>
- <https://en.wikipedia.org/w/index.php?curid=36179157>

Introduction

Long experience in ontological engineering led the authors to the following hypotheses:

Hypothesis 1

- It may be that most of the ontologies do not reuse ontologies of their same domain.
- Authors' belief in this hypothesis has been reinforced by the work carried out by Kamdar and colleagues [Kamdar et al., 2017], focused on BioPortal, which states, as one of their conclusions, that the reuse among biomedical ontologies is quite limited.

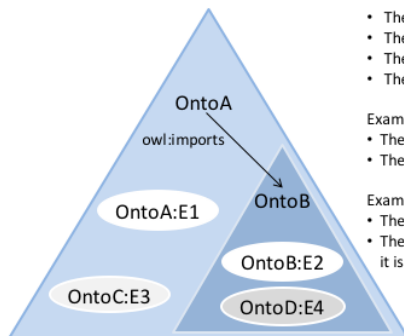


Image attribution:

- By Lusanaherandraton - Own work, Public Domain, <https://commons.wikimedia.org/w/index.php?curid=885678>

Hypothesis 2

- Ontology reuse just by referencing the reused ontology elements URIs (soft reuse) may be predominant with respect to ontology reuse by means of importing the reused ontologies (hard reuse).
- This hypothesis is reinforced by [Poveda-Villalón et al., 2012] work.



- The element OntoA:E1 is defined in OntoA
- The element OntoB:E2 is defined in OntoB
- The element OntoC:E3 is defined in OntoC
- The element OntoD:E4 is defined in OntoD

Examples of **Soft Reuse**:

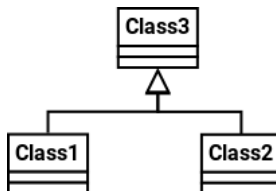
- The element OntoC:E3 is referenced in OntoA
- The element OntoD:E4 is referenced in OntoB

Examples of **Hard Reuse (through import)**:

- The element OntoB:E2 is imported in OntoA
- The element OntoD:E4 is imported in OntoA as it is referenced in the imported OntoB

Hypothesis 3

Reuse of classes may be predominant with respect to reuse of properties.



Hypothesis 4

- There may be a set of ontologies (FOAF^a, Dublin Core^b, etc.) that are being reused in most of the ontologies.

^a<http://xmlns.com/foaf/spec/>

^b<http://dublincore.org/>



Image attribution: By Hilton William Ganzo Perantunes - Own work, based on a JPG icon from the official site of the project., Public Domain, <https://commons.wikimedia.org/w/index.php?curid=1429917>

Objectives

- **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?



Objectives

- **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 its same domain?*



Introduction

Objectives

- **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?



Image attribution: De Eckhard Pecher (Arcimboldo) - Trabajo propio, CC BY 2.5,
<https://commons.wikimedia.org/w/index.php?curid=2662937>

Objectives 1 (soft and hard reuse) and 2 (reuse across the same domain) were approached in a quantitative way

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



Objective 3 (factors that hamper reuse) required a qualitative approach

It was achieved basing on the authors' experience in a real ontology development use case. Once the use case had been analysed, the generalisation study was carried out computationally by checking if the obstacles found in candidate ontologies to reuse in the use case were present in other LOV ontologies.



Definition of variables for reuse, hard reuse and soft reuse (Objective 1)

		Minimum	First Quartile	Median	Mean	Third Quartile	Maximum
Ontology Elements	Number of Classes (c)						
	Number of Object Properties (op)						
	Number of Datatype Properties (dp)						
	Number of Entities (total = c + op + dp)						
Hard Reuse	Number of Imported Classes (ic)						
	Number of Imported Object Properties (iop)						
	Number of Imported Datatype Properties (idp)						
Soft Reuse	Number of Classes (ndnic) that are neither defined nor imported						
	Number of Object Properties (ndnop) that are neither defined nor imported						
	Number of Datatype Properties (ndndp) that are neither defined nor imported						
Percentages	Imported Entities (per_i) (Hard Reuse)						
	Entities that are neither defined nor imported (per_ndni) (Soft Reuse)						
	Reused Entities (per_reuse) (Hard and Soft Reuse)						
	Reused Classes (per_reuse_c)						
	Reused Properties (per_reuse_p)						
	Reuse Object Properties (per_reuse_op)						
	Reuse Datatype Properties (per_reuse_dp)						



Definition of variables for reuse (Objective 1)

W3C endorsed ontologies

- 1 Ontologies whose URI is defined in the *<http://www.w3.org/>* namespace.
- 2 Ontologies identified by W3C as good ontologies.^a Some of the ontologies belonging to the list of good ontologies are FOAF, Dublin Core, SIOC,^b etc.

^ahttps://www.w3.org/wiki/Good_Ontologies

^b<http://sioc-project.org/>

Variable: percentage of LOV ontologies that reused W3C endorsed ontologies.

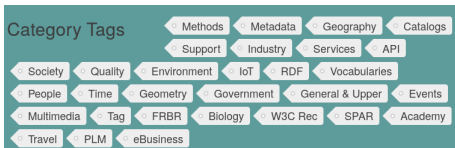


Definition of variables for reuse across the same domain (Objective 2)

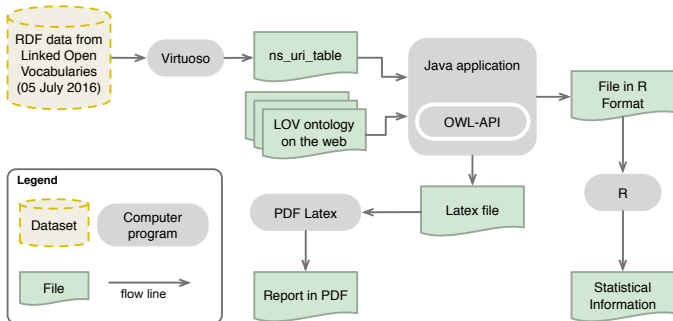
LOV domain ontologies

LOV ontologies belonging to none of the categories *Metadata*, *RDF* and *W3C Rec*.

- 1 **Variable:** percentage of LOV domain ontologies that reused other ontologies apart from their modules.
- 2 **Variable:** the percentage of these ontologies such that some of the ontologies they reused belonged to the same LOV category as themselves.



Software support to achieve objectives 1 and 2



Ontology developed to achieve objective 3

Ontología de Títulos Universitarios (OTN)

language **es**

Revisión número:

1.0

Autores:

[Mariano Fernández-López](#)

Ontologías importadas:

[skos](#)

[dul-dolce-zero-en-espanol](#)

[d0.owl](#)

[DUL.owl](#)

Descargar serialización:

Format: [RDF/XML](#) [Format: N-Triples](#) [Format: TTL](#)

Licencia:

[Licence: Creative Commons CC BY-SA](#)

[Provenance de esta página](#)

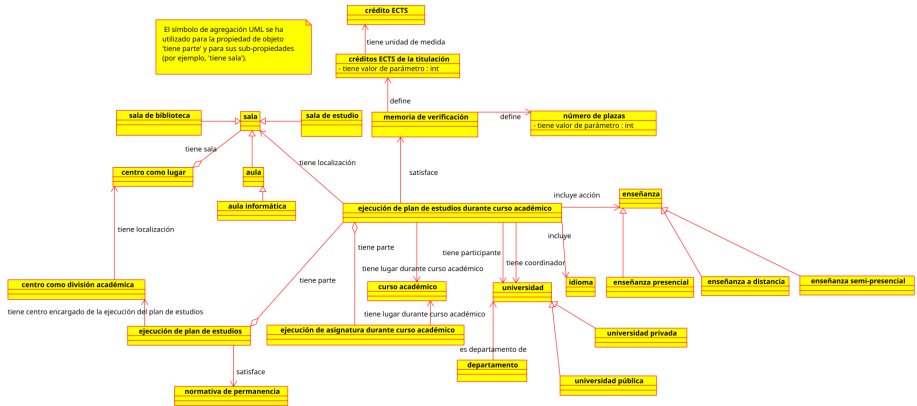
Sintesis

El propósito de la elaboración de esta ontología es especificar de manera formal una estructura conceptual para el intercambio de información de acuerdo con la Guía Práctica para la Evaluación del Seguimiento de Títulos Universitarios Oficiales de la Fundación Madrid http://www.madrimasd.org/uploads/acreditacion/doc/seguimiento2013/Guia_de_Evaluacion_2014.pdf.

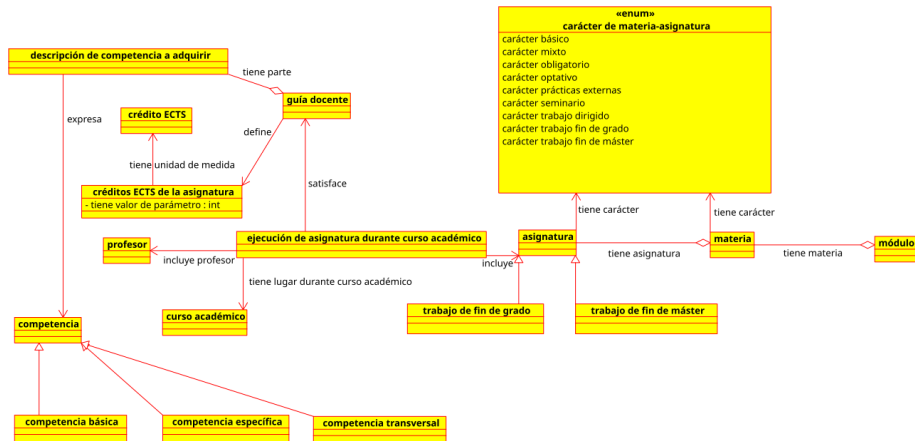
<http://ontology.linkeddata.es/publish/otn/index-es.html>



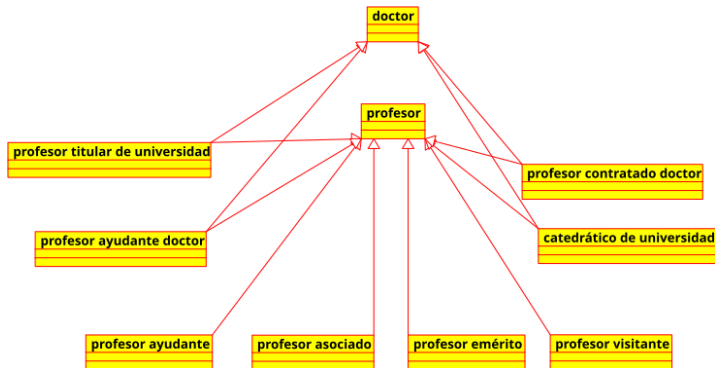
Ontology developed to achieve objective 3 (module 1)



Ontology developed to achieve objective 3 (module 2)



Ontology developed to achieve objective 3 (module 3)



Results and discussion

Results concerning Objective 1 “Measure reuse, hard reuse and soft reuse” and Objective 2 “Measure reuse across the same domain”

During the process of evaluating the variables, 386 out of the 558 ontologies available in the frozen SPARQL could be loaded

- 1 Timeout expiration during ontology loading (after 120 seconds) without arising any other exception. (21 ontologies)
- 2 `UnparsableOntologyException`: This exception means that the ontology is not processable due to syntax errors. (46 ontologies)

(It continues)



Public Domain, <https://commons.wikimedia.org/w/index.php?curid=1565425>

Results concerning Objective 1 “Measure reuse, hard reuse and soft reuse” and Objective 2 “Measure reuse across the same domain”

- 1 `OWLOntologyCreationIOException`: This exception arises when the ontology is not longer available. (70 ontologies)
- 2 `UnloadableImportException`: This exception happens when some of the ontologies imported by the ontology could not be loaded. (34 ontologies)
- 3 `Exception`: This a generic exception. The precise cause has not been determined by the tool. (1 ontology)



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.



Statistics of reuse, hard reuse and soft reuse

		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

Percentages of reuse

		Minimum	First Quartile	Median	Mean	Third Quartile	Maximum
Percentages	Imported Entities (<i>per_i</i>) (Hard Reuse)	0.000	0.000	0.000	17.615	6.161	100.000
	Entities that are neither defined nor imported (<i>per_ndni</i>) (Soft Reuse)	0.000	0.9093	9.5696	18.0373	25.8333	100.000
	Reused Entities (<i>per_reuse</i>) (Hard and Soft Reuse)	0.00	7.69	21.43	30.40	46.78	99.57
	Reused Classes (<i>per_reuse_c</i>)	0.00	13.90	38.46	41.15	65.33	100.00
	Reused Properties (<i>per_reuse_p</i>)	0.000	0.000	5.882	23.249	39.056	100.00
	Reuse Object Properties (<i>per_reuse_op</i>)	0.000	0.000	4.058	23.137	35.893	100.00
	Reuse Datatype Properties (<i>per_reuse_dp</i>)	0.000	0.000	3.571	29.233	57.143	100.00

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%.

Hard reuse and soft reuse

		Minimum	First Quartile	Median	Mean	Third Quartile	Maximum
Percentages	Imported Entities (<i>per_i</i>) (Hard Reuse)	0.000	0.000	0.000	17.615	6.161	100.000
	Entities that are neither defined nor imported (<i>per_ndni</i>) (Soft Reuse)	0.000	0.9093	9.5696	18.0373	25.8333	100.000
	Reused Entities (<i>per_reuse</i>) (Hard and Soft Reuse)	0.00	7.69	21.43	30.40	46.78	99.57
	Reused Classes (<i>per_reuse_c</i>)	0.00	13.90	38.46	41.15	65.33	100.00
	Reused Properties (<i>per_reuse_p</i>)	0.000	0.000	5.882	23.249	39.056	100.00
	Reuse Object Properties (<i>per_reuse_op</i>)	0.000	0.000	4.058	23.137	35.893	100.00
	Reuse Datatype Properties (<i>per_reuse_dp</i>)	0.000	0.000	3.571	29.233	57.143	100.00

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.

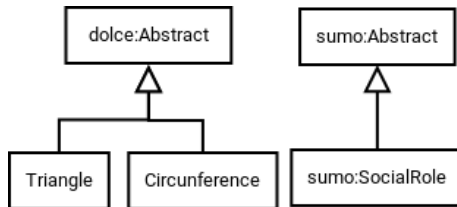
Validation of hypothesis

Hypothesis 2 is valid \implies Ontology reuse just by referencing the reused ontology elements URIs (soft reuse) is predominant with respect to ontology reuse by means of importing the reused ontologies (hard reuse).

Hypothesis 1	
Hypothesis 2	✓
Hypothesis 3	
Hypothesis 4	

Consequences of soft reuse

What does it happen if soft reuse was used in this example?



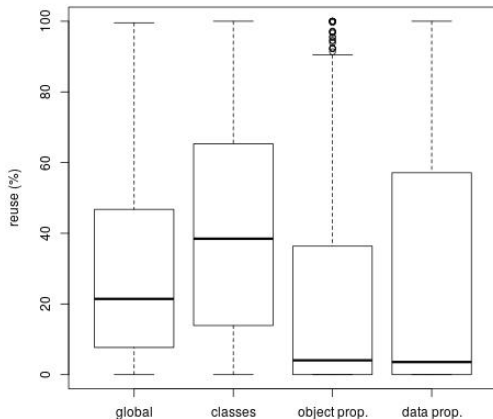
Ontologies that reuse W3C endorsed ontologies

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

Hypothesis 4 is valid \implies There are a set of ontologies, in particular, the W3C endorsed ontologies, that are being reused in most of the ontologies.

Hypothesis 1	
Hypothesis 2	✓
Hypothesis 3	
Hypothesis 4	✓

Box plots for percentages of reuse



Hypothesis 1	
Hypothesis 2	✓
Hypothesis 3	✓
Hypothesis 4	✓

Hypothesis 3 is valid \implies Reuse of classes is predominant with respect to reuse of properties.

Reuse across the same domain (Objective 2)

- 1 Percentage of LOV domain ontologies that reused other ontologies apart from their modules: 72.80%.
- 2 The percentage of these ontologies such that some of the ontologies they reused belonged to the same LOV category as themselves: 19.92%.

Hypothesis 1 is valid \implies Most of the ontologies do not reuse ontologies of their same domain.

Hypothesis 1	✓
Hypothesis 2	✓
Hypothesis 3	✓
Hypothesis 4	✓

Results concerning Objective 3. Factors against reuse across the same domain

Factors identified while building our domain ontology

- ① Predominant use of English and **conceptualizations biased towards the Anglo-Saxon world.**
- ② Deficiencies in the documentation
- ③ Loss of information due to an imported ontology that is not available.
- ④ Unavailable license.
- ⑤ Heterogeneity between the subdomains of the reusing and the reused ontologies.

Results concerning Objective 3. Factors against reuse across the same domain

Generalisation of the use case

Predominant use of English. 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%.



Image attribution: By Jolly Janner - Own work, Public Domain,
<https://commons.wikimedia.org/w/index.php?curid=7116075>

Generalisation of the use case

Deficiencies in the documentation. The *Curl* tool allowed checking each ontology for its corresponding documentation. Just 17.18% of them provided such a documentation.

```
<?xml version="1.0"?>
<rdf:RDF xmlns="http://www.arpenteur.org/ontology/Arpenteur.owl#"
  xmlns:base="http://www.arpenteur.org/ontology/Arpenteur.owl#"
  xmlns:cc="http://creativecommons.org/ns#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:gsp="http://www.opengis.net/ont/geosparql#"
  xmlns:owl2xml="http://www.w3.org/2006/12/owl2-xml#"
  xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
  xmlns:swrl="http://www.w3.org/2003/11/swrl#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:nadlShip="http://www.arpenteur.org/ontology/nadlShip-1.6.owl#"
  xmlns:swrla="http://swrl.stanford.edu/ontologies/3.3/swrla.owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:swrlabil="http://www.arpenteur.org/ontology/swrlabil.owl#"
  xmlns:xml="http://www.w3.org/XML/1998/namespace"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:time="http://www.w3.org/2006/time#"
  xmlns:arpenteur="http://www.arpenteur.org/ontology/Arpenteur.owl#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <owl:Ontology rdf:about="http://www.arpenteur.org/ontology/Arpenteur.owl">
    <owl:imports rdf:resource="http://www.opengis.net/ont/geosparql"/>
    <dc:contributor rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Mohamed BEN ELLEFI</dc:contributor>
    <rdfs:comment xml:lang="en">
      Arpenteur ontology is dedicated to photogrammetry, archeology and oceanology communities
      in order to perform tasks such as image processing, photogrammetry and modelling.
      Website: http://www.arpenteur.org
```

Results concerning Objective 3. Factors against reuse across the same domain

Generalisation of the use case

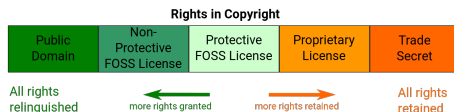
Loss of information due to an imported ontology that is not available. As mentioned before, the complete population of LOV had to be curated due to a subset not being able to load.



Results concerning Objective 3. Factors against reuse across the same domain

Generalisation of the use case

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 resulting percentage is 60.28%.



By shaddim / Mark Webbink -

https://www.redhat.com/f/summitfiles/presentation/May31/Open%20Source%20Dynamics/Troan_OpenSourceProprietyPersp.pdf, Public Domain,

<https://commons.wikimedia.org/w/index.php?curid=45955580>

Results concerning Objective 3. Factors against reuse across the same domain

Generalisation of the use case

Heterogeneity between the subdomains of the reusing and the reused ontologies. The generalisation of this factor remains for future work.

- ① Given that the study was applied to a whole population of ontologies (LOV), the results have not been expressed in terms of significance.
- ② 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 is strictly valid for BioPortal.

Conclusions

- ➊ Scarcely ever an ontology is maintained by a company that commercialises the reuse of the ontology.
- ➋ An ontology being maintained and updated by a community similar to those of free software is not frequent either (exceptions are, e.g., W3C ontologies).
- ➌ 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).
- ➍ In addition, many ontologies do not include an explicit declaration of their license [Poblet et al., 2016].
- ➎ Another important issue detected concerns the assumption of English as lingua franca.
- ➏ They take as reference conceptualizations of the Anglo-Saxon world.

- ① Processes as those used by organisations like Apache¹² for software development could be taken as inspiration for ontology development.
- ② It is crucial to continue the research being carried out on ontology localisation [Gracia et al., 2013] in order to identify those ontology modules independent of a region or country.
- ③ Finally, tools for ontology modularisation (e.g. Ontofox [Xiang et al., 2010] or the Protégé functionality to merge axioms) and methods [d'Aquin, 2012] [Courtot et al., 2011] deserve special attention, as in practical settings they might represent the key for users to extract only those ontology submodules needed.

¹<https://www.apache.org/foundation/getinvolved.html>

²<https://www.apache.org/dev/new-committers-guide.html>



Courtot M, Gibson F, Lister AL, Malone J, Schober D, Brinkman RR, Ruttenberg A

MIREOT: The minimum information to reference an external ontology term
Applied Ontology 6(1):23-33



d'Aquin M (2012)

Modularizing Ontologies

Ontology Engineering in a Networked World 213-233



Gracia J, Montiel-Ponsoda E, Cimiano P, Gómez-Pérez A, Buitelaar P, McCrae J (2013)

Multilingualism linkeddata multilingual

Web Semantics: Science, Services and Agents on the World Wide Web 11



Kamdar MR, Tudorache T, Musen M (2017)

A Systematic Analysis of Term Reuse and Term Overlap across Biomedical Ontologies

Semantic Web Journal 8(6):853-871



Poblet M, Aryani A, Manghi P, Unsworth K, Wang J, Hausstein B, Dallmeier-Tiessen S, Klas CP, Casanovas P, Rodríguez-Doncel V (2016)

Assigning Creative Commons Licenses to Research Metadata: Issues and Cases

CoRR - Computing Research Repository



Poveda-Villalón M, Suárez-Figueroa MC, Gómez-Pérez A (2012)

The Landscape of Ontology Reuse in Linked Data

Proceedings Ontology Engineering in a Data-driven World (OEDW 2012)



Xiang Z, Courtot M, Brinkman RR, Ruttenberg A, He Y (2010)

OntoFox: web-based support for ontology reuse

emphBMC research notes 3(1)

The End