Analysis of the evolution of ontologies using OQuaRE: Application to EDAM

Manuel Quesada-Martínez, Astrid Duque-Ramos, Jesualdo Tomás Fernández-Breis*
Facultad de Informática, Universidad de Murcia, IMIB-Arrixaca, CP 30100 Murcia

ABSTRACT

In recent years, the biomedical community has developed a significant number of ontologies. The curation of biomedical ontologies is a complex task, which has the practical implication of a high number of versions of ontologies in short time, because biomedical ontologies evolve rapidly. New versions are periodically published in ontology repositories. Ontology designers need to be supported for the effective management of the evolution of biomedical ontologies given this level of activity, because the different changes may affect the engineering and quality of the ontology. This is why we think that there is a need for methods that contribute to the analysis of the effects of changes and evolution of ontologies.

In this paper we approach this issue from the ontology quality perspective. In previous works we have developed an ontology evaluation framework based on quantitative metrics, called OQuaRE. Here, OQuaRE will be used as a core component in a method that permits to analyze the different versions of biomedical ontologies using a common framework. The objective is to help ontology developers to study the evolution of ontology versions in terms of changes in the quality dimensions analyzed in OQuaRE. In this work we explain how OQuaRE can be adapted for supporting this process and report the application of the method to 16 versions of the EDAM ontology. Discussion is provided on the evolution of the quality scores of those versions according to the OQuaRE quality perspective.

1 INTRODUCTION

In recent years, the biomedical community has increased its effort in the development of good ontologies and this will continue in the future (Hoehndorf *et al.*, 2014). As a consequence, ontology developers publish their new ontologies across the Internet, and they are accessible from different sources. BioPortal (Whetzel *et al.*, 2011) contains 428 ontologies at the time of writing, and new content is published every week. BioPortal provides for automatic updates by user submissions of new versions, which are accessible via web browsers and through web services (Whetzel *et al.*, 2011).

The curation of ontologies is a complex task because of their high level of activity and rapid evolution (Malone *et al.*, 2010). For this reason, their number and versions grow rapidly. The analysis of versions was introduced by Klein and Fensel (2001), who defined ontology versioning as the ability to handle changes in ontologies by creating and managing different variants of it and pointed out the importance of highlighting differences between versions. Later, Noy *et al.* (2003) claimed that a versioning system for ontologies must compare and present structural changes rather than changes in text representation or source files. They described a version-comparison algorithm that produces a structural difference between ontologies, which were presented to users through an interface for analysing them (Noy *et al.*, 2004). Later, Malone *et al.* (2010) presented Bubastis that reports on 5 major types of ontology changes: added or removed axioms to an existing named class (NC), NCs added, NCs made obsolete and edited annotation properties. Bubastis[1] was used in (Malone *et al.*, 2010) for measuring the level of activity of bio-ontologies, and this is used in BioPortal to generate reports about changes between 2 consecutive versions. Recently, Copeland *et al.* (2013) focused on changes in asserted and inferred axioms taking into account reasoning capabilities in ontologies (Wang *et al.*, 2004).

In this work, we are interested in studying the evolution of ontologies from the perspective of ontology quality. The analysis of quality in ontologies has been addressed in different ways in the ontology evaluation community (Gangemi *et al.*, 2006; Tartir and Arpinar, 2007; Ma *et al.*, 2009; Duque-Ramos *et al.*, 2011). Gangemi *et al.* (2006) approached it as a diagnostic task based on ontology descriptions, using three categories of criteria (structural, functional and usability profiling). Similarly, Rogers *et al.* (2006) proposed an approach using four qualitative criteria (philosophical rigour, ontological commitment, content correctness, and fitness for a purpose). Quantitatively, Yao *et al.* (2005) and Tartir and Arpinar (2007) presented metrics for evaluating structural properties in the ontology. Recently, Duque-Ramos *et al.* (2011) adapted the SQuaRE standard for software quality evaluation for defining a qualitative and quantitative ontology quality model.

In this paper, we propose a method that combines ideas from the ontology evaluation and ontology versioning field by adapting the OQuaRE methods for the needs of the study of changes between versions of ontologies. In this paper, we will explain the method and we will exemplify its application to the study of 16 versions of the ontology of Bioinformatics operations, types of data, formats, and topics (EDAM)[2]. The analysis of the results will permit to detect the impact of the series of changes in the ontology on the quality measurements offered by OQuaRE, which may contribute to learn about the engineering of the EDAM ontology. Our results will be compared with the ones obtained with Bubastis to study the relations between the level of activity of ontologies and changes in the OQuaRE quality scores. We believe this kind of method may contribute to generate new insights about biomedical ontologies.

2 METHODS

2.1 OQuaRE

OQuaRE (Duque-Ramos *et al.*, 2011) is an ontology quality evaluation framework based on the software product quality SQuaRE. OQuaRE aims at defining all the elements required for ontology evaluation: evaluation support, evaluation process and

*To whom correspondence should be addressed: jfernand@um.es

[1] http://www.ebi.ac.uk/efo/bubastis/

[2] http://edamontology.org

Characteristic	Description	Associated subcharacteristics
Structural	Formal and semantic relevant ontological properties that account for: the correct use of formal properties, clarity of cognitive distinctions and appropriate use of ontology modelling primitives and principles	"formalisation", "formal relations support", "redundancy", "consistency", "tangledness", "cohesion"
Functional Adequacy	Capability of the ontologies to be deployed fulfilling functional requirements, that is, the appropriateness for its intended purpose according to state-of-theart literature Stevens *et al.* (2008)	"reference ontology", "controlled vocabulary", "schema and value reconciliation", "consistent search and query", "knowledge acquisition", "clustering and similarity", "indexing and linking", "results representation", "text analysis", "guidance and decision trees" and "knowledge reuse and inferencing"
Reliability	Capability of an ontology to maintain its level of performance under stated conditions for a given period of time	"recoverability" and "availability"
Operability	Effort needed for the ontology use. Individual assessment of such use, by a stated or implied set of users	"learnability"
Compatibility	Capability of two or more ontologies to exchange information and/or to perform their required functions while sharing a hw/sw environment	"replaceability"
Maintainability	Capability of ontologies to be modified for changes in environments, in requirements or in functional specifications	"modularity", "reusability", "analysability", "changeability", "modification stability" and "testability"
Transferability	Degree to which the ontology can be transferred from one environment (e.g., operating system) to another	"adaptability"

Table 1. OQuaRE characteristics and subcharacteristics used in our method

metrics. The main objective of OQuaRE is to provide an objective, standardized framework for ontology quality evaluation, which could be applied in a number of situations.

OQuaRE is structured in 3 levels: quality characteristics, subcharacteristics and metrics. The evaluation of an ontology comprises a score for each quality characteristic, which depends on the evaluation of the its associated subcharacteristics. Similarly, the evaluation of a particular subcharacteristic depends on its associated metrics.

Table 1 describes the OQuaRE characteristics and subcharacteristics we use in this work. OQuaRE metrics adapt successful metrics from both ontology and software engineering communities (Tartir *et al.*, 2005; Yao *et al.*, 2005), which we briefly describe in Table 2. The complete specification of the OQuaRE quality model, including the associations between subcharacteristics and metrics, can be found at http://miuras.inf.um.es/oquarewiki.

OQuaRE metrics generate quantitative values in different ranges, so they are scaled into the range 1 to 5, which is the scale used in SQuaRE based approaches. There, 1 means not acceptable, 3 is minimally acceptable, and 5 is exceeds the requirements. The scaling method is based on the recommendations and best practices of the Software Engineering community for software metrics and ontology evaluation metrics (see the scale at the OQuaRE website).

OQuaRE provides a flexible analysis framework because ontologies can be analysed at different granularity levels: metric, subcharacteristic, characteristic and globally.

2.2 The method

Fig. 1 shows different stages of our method. We propose a method focused on measuring changes between different versions of the same ontology in terms of its global quality.

DEFINITION 1. ***Versioned corpus of an ontology** (vC)**:** vC is a list of versions $[\theta_{v1}, \theta_{v1+1} ..., \theta_{vt}]$ of the same ontology θ, where a time criterion must sequentially order vC.*

Ontologies can be found in different sources and formats, so we propose to normalise vC before applying OQuaRE. In the normalisation, we check that they are consistent, remove deprecated classes and save them in a the same OWL format.

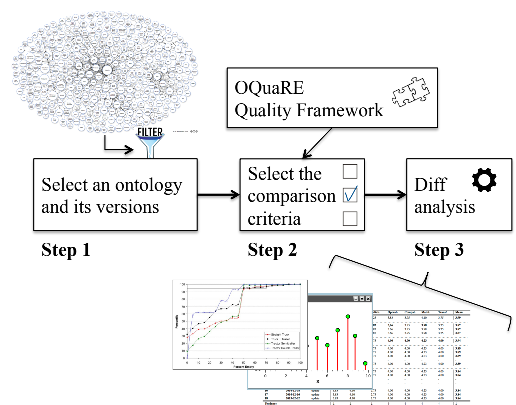

Fig. 1. Three main stages of the method for evaluating how to evolve the OQuaRE quality scores in a versioned corpus (vC) of an ontology

The second step permits to select which OQuaRE quality criteria are used to analyse the evolution of the ontologies. We define a comparison criterion as follows:

DEFINITION 2. ***Comparison criterion** ($f(\theta)$)**:** $f(\theta)$ is a quantifiable score that measures the capability of the ontology $\theta_{vi} \in vC$ to fulfill some criterion.*

When we compare different versions of the same ontology differences should be highlighted (Noy *et al.*, 2003). In our context, given an ontology θ and two consecutive versions θ_{vi}, $\theta_{v(i+1)} \in vC$:

DEFINITION 3. ***Change:** there is a change if $f(\theta_{vi}) \neq f(\theta_{v(i+1)})$, being $f(\theta)$ a comparison criterion.*

For example, if we obtain average scores of θ_{v1}=3.75 and θ_{v2}=3.87 for different versions, then the change of 0.12 indicates that the second one has a higher score than the first one. Provided

OQuaRE metric	Description
ANOnto	mean number of annotation properties per class
AROnto	number of restrictions of the ontology per classes
CBOnto	number of superclasses divided by the number of class minus the subclasses of Thing
CROnto	mean number of individuals per class
DITOnto	length of the largest path from Thing to a leaf class
INROnto	mean number of subclasses per class
NACOnto	mean number of superclasses per leaf class
NOCOnto	mean number of the direct subclasses per class minus the subclasses of Thing
NOMOnto	mean number of object and data property usages per class
LCOMOnto	mean length of all the paths from leaf classes to Thing
RFCOnto	number of usages of object and data properties and superclasses divided by the number of classes minus the subclasses of Thing
RROnto	number of usages of object and data properties divided by the number of subclassof relationships and properties
TMOnto	mean number of classes with more than 1 direct ancestor
WMCOnto	mean number of properties and relationships per class

Table 2. OQuaRE metrics and a brief description of how we calculate them

that OQuaRE metrics are scaled into the range 1-5 we need to differentiate which changes generate a variation in the scale of the metric score.

DEFINITION 4. ***Change in scale:*** *it is a sort of change where* $f(\theta_{vi})$ *and* $f(\theta_{v(i+1)})$ *are associated with different levels in the scale 1-5.*

In our example, there is no change in scale for θ_{v1} and θ_{v2}. However, changes from 3.25 to 2.87 or from 4.10 to 3.98 would make a change in the scale.

3 RESULTS AND DISCUSSION

We apply our method to the EDAM ontology. EDAM is an ontology of well established and familiar concepts that are prevalent within bioinformatics. EDAM includes types of data, data identifiers, data formats, operations and topics. We have choosen this ontology as exemplar because:

- It is well documented and its developers use a control version system[3] (CVS) so that we can trace changes.
- Its source files are accessible online. The latest version (v1.9) is published in the official project web page. Links to old versions can be found in BioPortal (18) and in the CVS (13).
- It has received 775 mean visits per month since October 2013 and 5 declared projects use EDAM, so it is a relevant ontology.
- Its number of versions and size (2 597 classes on average) makes its appropriate for this initial study.

We configured the experiment as follows. The *versioned corpus* is composed by the 18 EDAM versions in BioPortal as CVS content. We performed the *diff analysis* using OQuaRE metrics, subcharacteristics and characteristics as *comparison criteria*. We automatically processed the *versioned corpus* using a home-made software tool that implements the methods described in the previous section. This tool uses the OWL API[4] and Neo4j (http://neo4j.com) (paths metrics) for the calculation of OQuaRE metrics. 4 out of 18 versions were discarded by the tool: one could not be processed by the OWL API, and the other three were found inconsistent by Hermit (http://hermit-reasoner.com). In order to study the impact of deprecated classes in the results, we performed two studies: one with the ontologies containing the deprecated classes and one removing them. After this removal, v.13 and v.14 became consistent, so they were processed and included in the second study. Table 3 shows the results obtained in the characteristics level for the 16 versions in both studies. The whole set of results is available at [5], which includes scores and other information in the subcharacteristics and metrics levels.

3.1 Changes in Quality Scores

According to Table 3 the mean quality score ranges from 3.99 in the first version to 3.85 in the last one. The changes in this score do not generate a change in the scale. In fact, the EDAM ontology has always stayed between 3 and 4. Taking into account the OQuaRE scale, a 3-upper score reveals that good ontological principles seem to have been applied by the EDAM developers. In order to get insights about the engineering and evolution of the ontology, we continue by analysing changes in scale identified for different quality characteristics (see numbers in bold in Table 3).

3.1.1 Increase in quality scores: the Operability, Compatibility, Maintainability and Transferability characteristics increased from level 3 to 4 between v.4 and 5. Moreover, the ontology has maintained the score at this level since then. This behaviour happens for all the associated subcharacteristics. These scores are not included in the paper due to space constraints, but can be found in the result webpage. Descriptively, the highest score is found for Maintainability. The scores for its subcharacteristics "Reusability", "Analisability", "Changeability", "Modification stability" and "Testability", qualitatively make the ontology more reusable, and reduces negative side-effects due to changes in the ontology. In addition to this, these scores mean that it is easier to validate and detect flaws in EDAM. A similar reasoning can be done for the other three characteristics using the information in Table 1.

The OQuaRE metrics level reveals more information about the ontology components. In this level, the score for 9 OQuaRE metrics did not change for any version. The ones that changed are shown in Fig. 2. NOMOnto and RFCOnto are responsible for the increase from 4 to 5. The decrease in NOMOnto means that the mean number of property usage per class is lower, which is good in terms of maintainability of the ontology. RFCOnto is related to the usage of properties too.

3.1.2 Decrease in quality scores : the "Reliability" characteristic decreases from 3 to 2 between v.1 and v.2, whereas and the "Structural" characteristic does it from 4 to 3 between v.10 and v.11. The lowest score for the "Structural" characteristic is for Cohesion, which is related to the LCOMOnto metric (see Fig. 2) that uses the number of paths in the ontology in its calculation.

[3] https://github.com/edamontology/edamontology/releases

[4] http://owlapi.sourceforge.net

[5] http://miuras.inf.um.es/oquare/icbo2015

Version	Date	Status	Struct.		F. Adeq.		Reliab.		Operab.		Compat.		Maint.		Transf.		Mean	
			Org.	Nrm.	Org.	Nrm.	Org.	Nrm.	Org.	Nrm.	Org.	Nrm.	Org.	Nrm.	Org.	Nrm.	Org.	Nrm.
1	**2010-05-14**	beta	4.67	4.67	4.61	4.61	3.25	3.25	3.83	3.83	3.75	3.75	4.10	4.10	3.75	3.75	3.99	3.99
2	**2010-05-28**	beta	4.50	4.50	4.60	4.60	**2.88**	**2.88**	3,67	3.67	3.75	3.75	3.99	3.99	3.75	3.75	3.88	3.88
3	**2010-08-18**	beta	4,50	4.50	4.60	4.60	2.88	2.88	3.67	3.67	3.75	3.75	3.99	3.99	3.75	3.75	3.88	3.88
4	**2010-10-07**	beta	4,50	4.50	4.60	4.60	2.88	2.88	3.67	3.67	3.75	3.75	3.99	3.99	3.75	3.75	3.88	3.88
5	**2010-12-01**	beta	4.17	4.17	4.46	4.46	2.75	2.75	**4.00**	**4.00**	**4.00**	**4.00**	**4.23**	**4.23**	**4.00**	**4.00**	3.94	3.94
6	**2011-01-22**	beta	4.00	4.00	4.28	4.28	2.75	2.75	4.00	4.00	4.00	4.00	4.23	4.23	4.00	4.00	3.90	3.90
7	**2011-06-17**	beta	4.00	4.00	4.28	4.28	2.75	2.75	4.00	4.00	4.00	4.00	4.23	4.23	4.00	4.00	3.90	3.90
8	**2011-12-05**	beta	4.00	**3.83**	4.28	4.27	2.75	2.38	4.00	**3.83**	4.00	4.00	4.23	4.12	4.00	4.00	3.90	3.78
10	**2012-12-10**	beta	4.00	3.83	4.28	4.27	2.75	2.38	4.00	3.83	4.00	4.00	4.23	4.12	4.00	4.00	3.90	3.78
11	**2012-12-14**	release	**3.83**	3.83	4.11	4.27	2.75	2.38	4.00	3.83	4.00	4.00	4.23	4.12	4.00	4.00	3.85	3.78
12	**2014-02-18**	update	3.83	3.83	4.11	4.27	2.75	2.38	4.00	3.83	4.00	4.00	4.23	4.12	4.00	4.00	3.85	3.78
13	**2014-09-26**	update	-	3.83	-	4.27	-	2.38	-	3.83	-	4.00	-	4.12	-	4.00	-	3.78
14	**2014-11-14**	update	-	**4.00**	-	4.28	-	2.75	-	**4.00**	-	4.00	-	4.23	-	4.00	-	3.90
16	**2014-12-08**	update	3.83	4.00	4.11	4.28	2.75	2.75	4.00	4.00	4.00	4.00	4.23	4.23	4.00	4.00	3.85	3.90
17	**2014-12-16**	update	3.83	**3.83**	4.11	4.11	2.75	2.75	4.00	4.00	4.00	4.00	4.23	4.23	4.00	4.00	3.85	3.85
18	**2015-02-02**	update	3.83	3.83	4.11	4.11	2.75	2.75	4.00	4.00	4.00	4.00	4.23	4.23	4.00	4.00	3.85	3.85

Table 3. OQuaRE characteristics metric values for two versions of the EDAM ontology. These values are scaled from 1 to 5, where 1 is not acceptable and 5 exceeds the requirements.

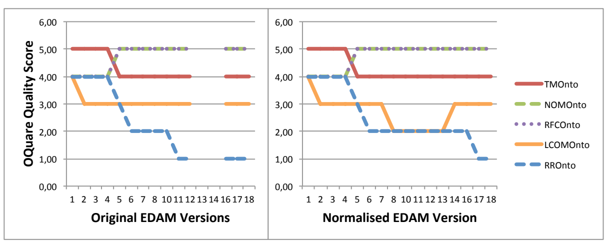

Fig. 2. Graphical representation of those OQuaRE metrics that have changes in scale between the 16 versions of the EDAM ontology

The largest decrease happens for "Formal relation support" (from 4 to 1). This fall is mainly influenced by the behaviour of the RROnto metric, which has 2 changes in scale. The first change is produced by the usage of properties, which descends 86% between v.4 and v.6. The usage of properties also decreases 8% between v.10 and v.11. This variation is smaller than the previous one but, together with an unusual increase in the number of relations (18%), it triggered the change in the RROnto scale. This increase in the number of relations is consequence of a structural change in v.11: deprecated classes were grouped as descendants of an ontology class in the first taxonomic level so the number of relations increased.

It should be noted that RROnto measures the usage of properties, not the number of them. Refactoring towards a common set of properties can often be a good sign, however the usage measures the number of times that a property is linked with an entity through an axiom. For example, while v.4 defines 16 with 6 734 usages, v.5 and v.6 define the same number of properties but with 1 979 and 937 usages respectively.

Finally, the "Structural" characteristic decrease is influenced by the "Tangledness" subcharacteristic. This is associated with TMOnto, which measures the distribution of the parents in the ontology. 10% of the classes have more than 1 direct parent in v.4, while this value grows up to 24% in v.5. This metric has a negative effect over the ontology because of the multiple inheritances, although this might reflect the biology within the ontology.

3.1.3 Influence of deprecated classes: the presence of deprecated classes grows from 3.51% (v.1) to 29.58% (v.18). Deprecated classes caused inconsistencies in v.13 and v.14. Table 3 shows that there are not significant changes at characteristic level between the ontologies with (Org) and without the deprecated classes (Nrm), but some changes happen at metric level. Fig. 2 shows the evolution of the scores for some quality metrics of the complete ontology (left) and the ontology without deprecated classes (right). The structural change previously explained for deprecated classes anticipates the drop of RROnto to v.11, whereas it happens in v.17 in the normalised version. Besides, LCOMOnto temporary descends to score level 2 between v.8 and v.13 in the normalised version. This effect on LCOMOnto cannot be appreciated in the ontologies with the deprecated classes.

3.2 Profiles of activity and its quality scores

Noy *et al.* (2003) state that the study of changes and commonalities should be a complementary process. We interpret the absence of changes as a sign of stability. However, we wonder if this stability is related to the level of activity in the ontology. For example, a difference of 0.12 between two versions might have a different interpretation depending on the number of classes added, edited and removed. We compare OQuaRE quality scores with those obtained with the five potential *profiles of activity* proposed in Malone *et al.* (2010): "initial, ad hoc", "expanding", "refining", "optimising, mature" and "dormant". They provide qualitative descriptions for setting the *profile of activity* of a set of ontologies based on the results obtained with Bubastis, which we applied to our *versioned corpus*. After that, we manually related regions with *profiles of activity* and compare them with our OQuaRE quality scores.

We can observe that some consecutive versions remain unchanged in terms of average quality, and we interpret this as a sign of stability in the quality scores. This happens in the ranges v.2-v.4, v.6-v.10 and v.11-v.18 (see Org columns in Table 3). According to the classification proposed in (Malone and Stevens, 2013), Fig. 3 shows how the ontology starts in a "refining" profile (1-4) because it "is largely refining the classes contained, rather than adding or deleting them, although some addition and deletion still occurs in lower numbers". This stage continues until v.12. From v.4 to

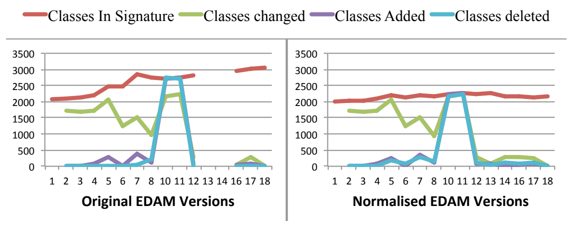

Fig. 3. Number of changes in classes obtained with Bubastis. We have applied Bubastis to every pair of consecutive ontologies θ_{vi} and $\theta_{vi+1} \in L$. We also include the number of classes in signature obtained as a primitive metric with OQuaRE.

v.7 the number of classes increases, so the ontology could be in an "expanding" profile during this stage. However, it seems that these new classes replace deprecated ones (Fig 3 right). Finally, the ontology is in "optimising, mature" stage because there is "no or very low levels of class deletions, some addition of new classes and changes to existing classes". We observe a stability during the "optimising, mature" stage related to the stability in quality scores. At characteristic level, the SD for v.11-v.18 is 0.04 on average, whereas it is 0.17 on average for v.1-v.10.

Finally, this ontology starts in a "refining" stage and evolve to the "optimising, mature". This fact might explain the high quality score and its low variability in EDAM from its first version. Although all the classes in the signature have suffered a change in the URIs between v.8 and v.10, this does not affect the quality scores.

3.2.1 Relation between the ontology status and its quality score: we labeled each version according to the status used by EDAM developers in BioPortal. From v.1 to v.10 they describe EDAM as a beta ontology. A beta version is used to describe a computer artifact that is near completion. The overlap between the stable versions (no beta) and the "optimising, mature" stage is a good indicator.

4 CONCLUSION AND FUTURE WORK

In this work, we have developed a method that combines the analysis of versions with an ontology quality evaluation framework. Its application to EDAM reveals that good ontological principles were applied in its development. The comparison between changes in quality scores and the level of the activity of the ontology justifies the low variability in the scores of the quality characteristics, as EDAM starts in a "refining" stage and evolve to the "optimising, mature" one. The analysis of changes in quality at both subcharacteristic and metric levels have shown some weaknesses and strengths of the ontology and the method. Our approach helps to identify systematically changes based on the OQuaRE metrics. However, it is out of the scope of this work to measure how the changes in quality scores relate to how the content conform to the domain represented by the ontology, which would be the main objective of complementary methods, like realism-based ones (Ceusters and Smith, 2006; Seppälä *et al.*, 2014). As future work, we propose to use the lessons learned in this experiment for improving the sensitivity of the method, in order to be more concise in the detection of changes. We cannot conclude that there is a relation between the quality and activity of classes using one ontology as exemplar, so the analysis of a wider set of ontologies is also a challenge, as it will help us to contextualise OQuaRE scores in the bio-ontology area.

ACKNOWLEDGEMENTS

This project has been possible thanks to the Spanish Ministry of Science and Innovation and the FEDER Programme through grants TIN2010-21388-C02-02, TIN2014-53749-C2-2-R, BES-2011-046192 (MQM) and EEBB-I-14-08700 (MQM), and by the Fundación Séneca (15295/PI/10).

REFERENCES

Ceusters, W. and Smith, B. (2006). A realism-based approach to the evolution of biomedical ontologies. In *AMIA Annual Symposium Proceedings*, volume 2006, page 121. American Medical Informatics Association.

Copeland, M., Gonçalves, R. S., Parsia, B., Sattler, U., and Stevens, R. (2013). Finding fault: detecting issues in a versioned ontology. In *Proceedings of the Second International Workshop on Debugging Ontologies and Ontology Mappings, Montpellier, France, May 27, 2013*, pages 9–20.

Duque-Ramos, A., Fernández-Breis, J. T., Stevens, R., and Aussenac-Gilles, N. (2011). OQuaRE: A square-based approach for evaluating the quality of ontologies. *Journal of Research and Practice in Information Technology*, **43**(2), 159–176.

Gangemi, A., Catenacci, C., Ciaramita, M., and Lehmann, J. (2006). Ontology evaluation and validation. *Proceedings of the 3rd European Semantic Web Conference (ESWC2006)*, **3**, 140–154.

Hoehndorf, R., Haendel, M., Stevens, R., and Rebholz-Schuhmann, D. (2014). Thematic series on biomedical ontologies in JBMS: challenges and new directions. *Journal of biomedical semantics*, **5**, 15.

Klein, M. and Fensel, D. (2001). Ontology versioning on the Semantic Web. *Proc of the Int Semantic Web Working Symposium SWWS*, **31**, 75–91.

Ma, Y., Ma, X., Liu, S., and Jin, B. (2009). A proposal for stable semantic metrics based on evolving ontologies. In *Artificial Intelligence, 2009. JCAI '09. International Joint Conference on*, pages 136–139.

Malone, J. and Stevens, R. (2013). Measuring the level of activity in community built bio-ontologies. *Journal of Biomedical Informatics*, **46**(1), 5–14.

Malone, J., Holloway, E., Adamusiak, T., Kapushesky, M., Zheng, J., Kolesnikov, N., Zhukova, A., Brazma, A., and Parkinson, H. (2010). Modeling sample variables with an Experimental Factor Ontology. *Bioinformatics*, **26**(8), 1112–1118.

Noy, N. F., Musen, M. a., Informatics, S. M., and Drive, C. (2003). Ontology Versioning as an Element of an Ontology-Management Framework. pages 1–17.

Noy, N. F., Kunnatur, S., Klein, M., and Musen, M. A. (2004). Tracking changes during ontology evolution. In S. McIlraith, D. Plexousakis, and F. van Harmelen, editors, *The Semantic Web ISWC 2004*, volume 3298, pages 259–273.

Rogers, J. E. *et al.* (2006). Quality assurance of medical ontologies. *Methods Inf Med*, **45**(3), 267–274.

Seppälä, S., Smith, B., and Ceusters, W. (2014). Applying the realism-based ontology-versioning method for tracking changes in the basic formal ontology. In *Formal Ontology in Information Systems - Proc. of the Eighth International Conference, FOIS 2014, September, 22-25, 2014, R. de Janeiro, Brazil*, pages 227–240.

Stevens, R., Wroe, C., Gobel, C., , and Lord, P. (2008). Applications of ontologies in bioinformatics. In S. Staab and R. Studer, editors, *Handbook on Ontologies in Information Systems*, pages 635–658. Springer.

Tartir, S. and Arpinar, I. B. (2007). Ontology evaluation and ranking using ontoqa. In *ICSC '07: Proceedings of the International Conference on Semantic Computing*, pages 185–192, Washington, DC, USA. IEEE Computer Society.

Tartir, S., Arpinar, I. B., Moore, M., Sheth, A. P., and Aleman-meza, B. (2005). Ontoqa: Metric-based ontology quality analysis. In *IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources*.

Wang, X. H., Zhang, D. Q., Gu, T., and Pung, H. K. (2004). Ontology based context modeling and reasoning using owl. In *Pervasive Computing and Communications Workshops, 2004. Proc. of the 2nd IEEE Annual Conference on*, pages 18–22. Ieee.

Whetzel, P. L., Noy, N., Shah, N., Alexander, P., Dorf, M., Fergerson, R., Storey, M. A., Smith, B., Chute, C., and Musen, M. (2011). BioPortal: Ontologies and integrated data resources at the click of a mouse. *CEUR Workshop Proceedings*, **833**, 292–293.

Yao, H., Orme, A., and Etzkorn, L. (2005). Cohesion metrics for ontology design and application. *Journal of Computer Science*, (1).