Noname manuscript No.
(will be inserted by the editor)

Using LOT methodology to develop a noise pollution ontology: a Spanish use case

Paola Espinoza-Arias[1,*] · María Poveda-Villalón[1] · Oscar Corcho[1]

Received: date / Accepted: date

Abstract Local administrations generate large amounts of data due to the processes followed to attend administrative governance issues and the needs of its citizenry. Sadly, in most cases this data is not fully exploited and remains within the institutions, making their reutilization difficult. Currently, open data initiatives had gained ground worldwide and more cities are taking advantage of adopting an open data strategy, which are visible at the organizational level and at user level. In this respect, there is a need to generate guidelines that allow cities: a) to identify datasets to be shared, for example pollution, commercial premises, public services, etc. and b) to publish quality data on their portals. Data should be accurate and interoperable among cities to facilitate reuse. This work describes the development process of an ontology to represent the acoustic pollution data collected by measurement stations located in cities, providing a common model for data publication. The developed ontology reuses several well-known ontologies and includes classes, properties and instances specifically created to cover this domain. This work also includes real examples about how to instantiate the ontology.

Keywords noise pollution · open data · ontology

Paola Espinoza-Arias
* Corresponding author
E-mail: pespinoza@fi.upm.es

María Poveda-Villalón
E-mail: mpoveda@fi.upm.es

Oscar Corcho
E-mail: ocorcho@fi.upm.es

[1] Ontology Engineering Group, ETSI Informáticos, Universidad Politécnica de Madrid, 28660 Madrid, Spain.

1 Introduction

Noise pollution is one of the environmental problems affecting the health and well-being of citizens in large towns and cities. According to the World Health Organization, noise is the second largest environmental cause of health problems, just after the impact of air quality [5]. Bearing in mind this public health issue, many cities around the world have defined several regulations in order to assess and manage environmental noise. In the case of the European Union, in 2002 the Directive 2002/49/EC [10], also known as the Environmental Noise Directive (END), was passed in order to identify noise pollution levels and to trigger the necessary action both at Member State and at European Union level. The END defines the environmental noise pollution as noise caused by road, rail and airport traffic, industry, construction, as well as some other outdoor activities. In accordance with one of the key areas related with this directive, Member States are committed to reporting their strategic noise maps every five years. The purpose of these maps is presenting and assessing the calculated/measured noise levels over a geographical area in order to determine the population exposed by this pollution. In the context of national legislation and local regulations, in Spain the *Ley del Ruido 37/2003* [3] has also been defined for protecting citizens from excessive noise pollution.

In order to detect the noise level of cities, several sensors, located at strategic places (e.g. train stations, airports, wide boulevards, etc.), are used to collect these data. These sensors, in most cases, are part of an interconnected sensor network deployed over the city. Although cities are reporting their strategic noise maps, there is not an agreement about how to make available the noise level data collected to build them. Several cities are providing their noise levels following different approaches as Open Data which is defined as data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike [8].

However, the lack of use of semantics in the description of these data hinders their reusability and interoperability. In order to describe these data, ontologies, understood as formal specifications of shared conceptualizations [25], are used to avoid ambiguities, to provide semantic interoperability and, in some cases, to bring inference capabilities over the data. Developers may take advantage of the mentioned benefits of the ontologies for example to integrate noise date coming from different sensors (providing data in different formats and granularity) in order to build ambient intelligence applications for smart cities. Before undertaking any ontology development a review of the state of the art should be carried out in order to identify potential ontologies to be reused. Out of all available ontologies in the noise domain, as shown in Section 2, there is not an ontology which allows the representation of the noise pollution concepts in a general way, however, other ontologies are reused to complement the proposed model.

In this paper we describe the process and activities followed in order to develop an ontology to describe noise pollution data. The process includes defining the ontological requirements, implementing the ontology, publishing it on the web and maintaining it. Such process is detailed in Section 3. In addition, an example of use of the proposed ontology is provided in Section 4 and a discussion about the ontology development experience is provided in Section 5. Our work represents a step forward to improve semantic interoperability in the noise pollution domain. Main conclusions about the work presented and future lines of work are listed in Section 6.

2 Related work

In this section, we discuss the state of the art of noise pollution ontology-based representation. There are some available noise related ontologies, in the following we present the main features of these ontologies.

The Recommender System Context (rsctx) ontology[1] represents the context of mobile usage which may be interesting in providing recommendations to users [17]. This ontology represents several dimensions, such as noise level, traffic level, light level, etc. In the noise level domain, this ontology allows modeling the current level of noise in the environment as 1) a symbolic level (e.g very noisy, silent, etc.), and 2) a number expressed in decibels. However, the rsctx ontology does not provide a representation of the devices used to measure noise and detailed definitions about the different kinds of noise levels are missing.

M3-lite (m3lite)[2] is a taxonomy that enables testbeds to semantically annotate the IoT data produced by heterogeneous devices and store them in a federated datastore such as FIESTA-IoT[3] [1]. This ontology provides several concepts representing measurements of noise level in the environment, sensors used to detect the noise level (e.g. sound sensor, noise level sensor, microphone, etc.), and the source where the noise was originated (e.g. traffic, siren of a police car, etc.). However, the m3lite ontology represents several sources of sound which are considered as noise nuisance (e.g. animals, crown, neighbors, etc.) instead of environmental noise (e.g. rail traffic, air traffic, etc.), thus this sound source representation is out of the scope of this work. In addition, the sound sensor is classified only by two domains of interest: smart building and transportation, which are difficult its reuse for our purposes.

The ISO 37120 standard,[4] named "Sustainable development of communities - Indicators for city services and quality of life", defines city indicators for specific themes (e.g. health, education, environment, etc.) in order to measure city performance. In the particular case of our domain of study, the environment theme contains a related indicator named noise pollution, which aims to estimate the percentage of population affected by noise annoyance in a specific area. In general, the majority of the standard indicators are defined as a ratio of parameters of two populations. In the case of the environment

[1] The Recommender System Context ontology `http://softeng.polito.it/rsctx`
[2] The M3-lite ontology `http://purl.org/iot/vocab/m3-lite`
[3] The FIESTA-IoT platform `http://fiesta-iot.eu/index.php/fiesta-testbeds`
[4] The ISO 37120:2014 standard `https://www.iso.org/standard/62436.html`

theme indicators these populations include observations generated by sensors in different times and locations. In order to represent the eight environment indicators of the ISO 37120 specification, the Global City Indicator Environment (iso37120en)[5] ontology has been defined. The iso37120en ontology has built on the Global City Indicators Foundation ontology, which aims at providing ontology design patterns to model indicators and indicator metadata [12]. The iso37120en includes generic ontologies in order to represent pollutants and pollutant concentrations, taxonomic groups of animal species, and sensor devices related to the environment indicators. This ontology allows modelling noise pollution by means of a formula which includes the average sound level over a 24 hour period (Lden) and city population. However, the iso37120en is heavy, it represents only the Lden noise level and it does not model the sources which originate noise. Accordingly, as the purpose of our work is representing several noise levels (e.g. average noise level and percentile noise level) originated by several acoustic emitters, this ontology is not enough to fulfill our requirements.

3 Methodological background

The presented ontology was built following the LOT (Linked Open Terms) methodology, proposed in [21] and further developed at [13]. This methodology is based on agile techniques in which sprints and iterations represent the main workflow organization in order to align the ontology development with software development agile practices. In addition, the methodology focuses on: a) the reuse on terms (ontology classes, properties and attributes) existing in already published vocabularies or ontologies and b) on the publication of the built ontology according to Linked Data principles. In addition, it is worth mentioning that the LOT methodology builds on top of the ontological engineering activities defined in the NeOn methodology [26] when available.

The LOT methodology defines iterations over a basic workflow composed of the following activities: 1) ontological requirements specification; 2) ontology implementation; 3) ontology publication; and 4) ontology maintenance. The activities, roles involved and expected outputs are depicted in Figure 1.

The following sections will briefly present, for each of the above-mentioned activities: a) main definitions and guidelines provided by the methodology for the activity and b) how the activity was carried out during the development of the noise ontology.

3.1 Ontology requirements specification

The ontology requirements specification activity refers to the activity of collecting the requirements that the ontology should fulfill [26]. These require-

[5] The Global City Indicator Environment ontology `http://ontology.eil.utoronto.ca/GCI/ISO37120/Environment.owl`

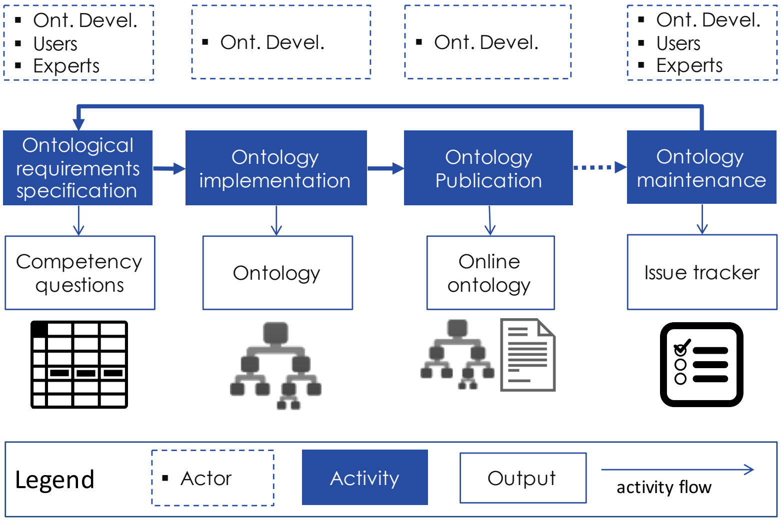


Fig. 1 LOT methodology base workflow of processes. Image taken from [13]

ments are usually related to the goal of the ontology, to the domain that the ontology should model, or to technical details of the ontology like the implementation language, among others.

In order to carry out this activity, the LOT methodology proposes the exchange of different documents between domain experts and ontology users (potentially software developers) and the ontology development team. Such documents could be manuals, API specifications, datasets, standards, formats used in the community, etc. Once the ontology engineers gather the information from documents or interviews with domain experts, they propose a set of ontological requirements. These ontological requirements can be written following the Competency Questions technique [15] or in the form of natural language sentences. Each ontological requirement may contain more information like provenance, comments, relations with other requirements, priority, the sprint in which it will be addressed, etc. Such list of ontological requirements are later validated and completed together with domain experts and ontology users in order to create the Ontology Requirements Specification Document (ORSD), which is the main output of this activity.

3.1.1 Ontology requirements specification in the noise pollution ontology

The ontology requirements specification activity was performed using different inputs. First, we used the noise pollution data available in Spanish open data portals in order to extract the most common terms available from these data sources, such as the location of noise sensor stations, measurement date, types of noise levels, etc. Second, we extracted several definitions from the Spanish document *Ley del Ruido 37/2003*, for instance, the acoustic emitters (agglomerations, major roads, major airports, etc.). Third, we had several interviews with a domain expert in order to validate the first semantic model draft and improve it including important concepts. Additional concepts were obtained from the International Standard IEC 61672-1:2013 [18] and IEC 61672-2:2013 [19], documents suggested by the expert, in order to obtain definitions for sensor classes and weighting methods respectively.

Taking the first and second aforementioned inputs, we generate a first proposal of ontological requirements written as Competency Questions. It is worth noting that even though the LOT methodology propose describing the requirements both as Competency Questions and natural language statements, for noise ontology development all the requirements were written as Competency Questions. In order to define and share them, an online spreadsheet which included the following fields was used:

- *Identifier*, unique for each requirement.
- *Competency Quest*ion, to define the requirements which the ontology should fulfill.
- *Answer*, the answer to the competency question.
- *Clarification of Competency Questions*, to include comments related with the questions.
- *References*, to include the provenance.

Once this first proposal was written, we shared the spreadsheet[6] with the domain expert in order to validate the requirements. As we previously mentioned, the expert determined which questions were correct and included new ones. Table 1 shows an excerpt from the final requirements for the noise pollution ontology.

Identifier	**Competency Question**	**Answer**
CQ1	Which is the measured noise level in a specific moment and location?	The detected noise level Ld during April in location with coordinates 40.42, - 3.69, 648 is 67.4 dB.
CQ2	Which is the location of a specific measurement station?	The location of the measurement station ”Paseo de Recoletos” is 40.42, -3,69, 648 (Longitude, Latitude, Altitude)
CQ3	In which day interval the noise level has been detected?	The noise level Ld has been detected in the time interval beteween 07:00 to 19:00.
CQ4	Which was the last calibration sensor date?	The datetime, for example, 2017-04-21T14:00:00+01:00

Table 1 Noise pollution ontology requirements excerpt

It is worth mentioning that all the files generated during the development process, including these requirements, are stored and managed in a GitHub

[6] Google spreadsheet for requirements management `https://docs.google.com/spreadsheets/d/1A-QRCsjHiLp1ROP5fwEsQr6AfyFTv7qK4YA2IyIL9GI/edit?usp=sharing`

repository[7] from the GitHub account of the Spanish thematic network on Open Data and Smart Cities.[8]

3.2 Ontology implementation

The aim of the ontology implementation activity is to build the ontology using a formal ontology implementation language, based on the ontological requirements identified by the domain experts [26]. This phase is carried out in different sprints in which a set of requirements are selected to be implemented. The ontology development team schedules and plans the implementation according to the prioritization of the requirements. After each iteration or sprint a new version of the ontology is produced. The ontology implementation is usually divided into the following sub-activities:

- **Conceptualization**: Ontology conceptualization refers to the activity of organizing and structuring the information obtained during the acquisition process, into meaningful models at the knowledge level and according to the ontology requirements specification document [26]. During this activity ontology developers may use different tools as diagrams (for example, UML based) or description logics to create the model according to the requirements previously acquired.
- **Encoding**. This activity refers to the transformation of the conceptualization into an ontology, or set of ontologies, expressed in the chosen ontology implementation language, for example OWL. In line with ontology encoding the **ontology reuse** activity could be also carried out. In this sense, when implementing the conceptualized model, ontology developers should search for existing ontologies in order to reuse them as a whole or reuse part of them. Searching in ontology registries and indexes like Linked Open Vocabularies (LOV) [28] is advisable to find potential ontologies to be reused. The ontology reuse activity is highly recommended and one of the basis of the LOT methodology as by reusing existing models, the interoperability of the resulting ontology is maximized while the resources spent in the development could be reduced. For experienced developers, the reuse activity could be carried out both during the conceptualization or the implementation one.
- **Evaluation**. This activity refers to checking the technical quality of an ontology against a frame of reference [26]. Such checking may be carried out considering different evaluation criteria, for example: domain coverage, fit for purpose or application, detection of bad practices, logical consistency checking, etc. Some recommended tools are reasoners for the logical consistency checking and ontology validators like OOPS! (OntOlogy Pitfall Scanner!) [22] for bad design practices detection. According to the

[7] Noise ontology repository `https://github.com/opencitydata/medio-ambiente/contaminacion-acustica`.

[8] Spanish network on Open Data and Smart Cities `http://www.opencitydata.es`

evaluation results, the ontology development team could go back to the conceptualization or encoding activities in order to fix bugs or improve the current version of the ontology.

3.2.1 Ontology implementation in the noise pollution ontology

Taking into account the output from the previous activity, several terms were extracted. In order to perform the ontology conceptualization activity, we used these terms to search for existing ontologies using LOV. As a result, we identified several well-known ontologies to represent these terms:

1. The W3C Semantic Sensor Network (SSN) ontology,[9] allows describing sensors in terms of capabilities, measurement processes, observations and deployments [6].
2. The OWL-Time,[10] an ontology of temporal concepts, for describing temporal properties of resources [7].
3. The WGS84 ontology,[11] for representing information about spatially-located things, such as latitude, longitude, altitude, etc. [4].
4. The Ontology of Units of Measure and Related Concepts (OM), [12] on the domain of quantities and units of measure [23]. Since this ontology has not provided a definition for the decibel unit, we selected the term definition described in the DBpedia ontology.

The first conceptualization model is shown in Figure 2. In such figure the ontologies in which each concept or relation is defined is indicated by the use of prefixes, for example the concept `ssn:Sensor` is defined in the "https://www.w3.org/2005/Incubator/ssn/ssnx/ssn#" namespace denoted by the prefix "ssn" while the local identifier is represented by "Sensor". In this figure it can be observed that the ontology reuse activity was carried out in parallel with the ontology conceptualization ontology, considering from the first draft models which ontologies could be reused.

The model depicted in Figure 2 relates the reused ontologies except for the class `om:Unit_of_measure` because there is no object property to link it with the `ssn:ObservationValue` class. Due to the lack of this link and that there are concepts which are not available in any ontology to be reused, we incrementally defined them until obtaining the final conceptualization model as shown in Figure 3. As it can be observed, the property `noise:hasUnitOfMeasure` was created to link the above-mentioned classes which is needed to model the noise ontology requirements as it can be deduced by CQ1 answer (Table 1): *The detected noise level Ld during April in location with coordinates 40.42, -3,69, 648 is 67.4 dB.* In such requirement the value of the noise level is linked to the unit of measure in which it is expressed.

[9] The W3C Semantic Sensor Network ontology `https://www.w3.org/2005/Incubator/ssn/ssnx/ssn`

[10] The OWL-Time ontology `http://www.w3.org/2006/time`

[11] The WGS84 ontology `https://www.w3.org/2003/01/geo/wgs84_pos`

[12] The Ontology of Units of Measure and Related Concepts `http://www.wurvoc.org/vocabularies/om-1.8`

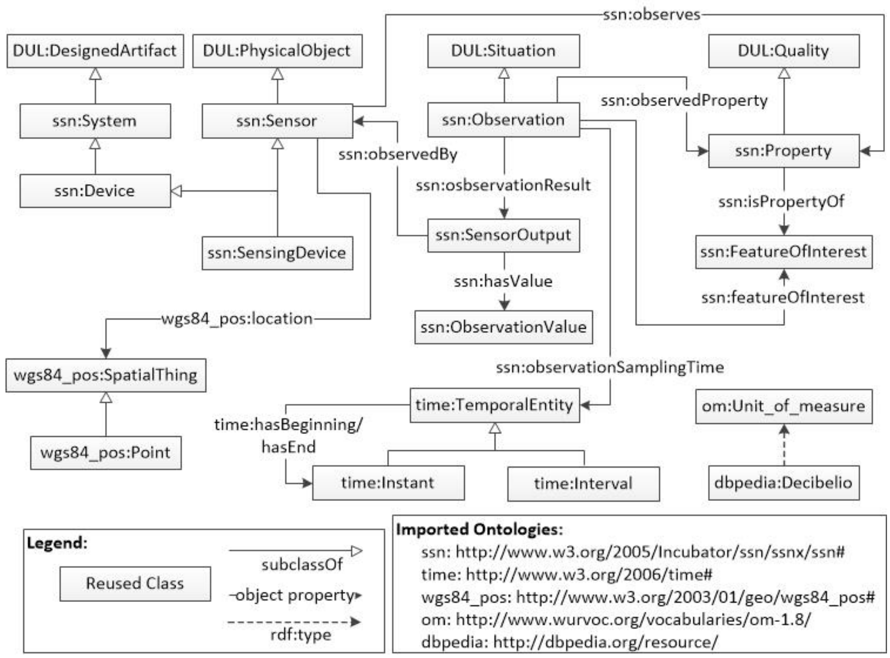


Fig. 2 Initial conceptualization of the noise pollution ontology

Once the final conceptualization model was defined, it was encoded in OWL. The ontology code is available in our GitHub repository, as explained in subsection 3.1.1. An excerpt of the ontology Turtle code is shown in Listing 1.

Even though the ontology elements contain `rdfs:label` and `rdfs:comment` annotations in English and Spanish, in the Listing 1 only the English comments are shown for space and readability issues. Furthermore, we developed a SKOS[13] thesaurus to represent the types of acoustic emitters. We adopted this modeling approach since these emitters are part of a well-structured and closed list which is defined in the *Ley del Ruido 37/2003* document and SKOS is a well-known ontology for the representation of controlled vocabularies like this case.

[13] The Simple Knowledge Organization System ontology `https://www.w3.org/TR/skos-primer`

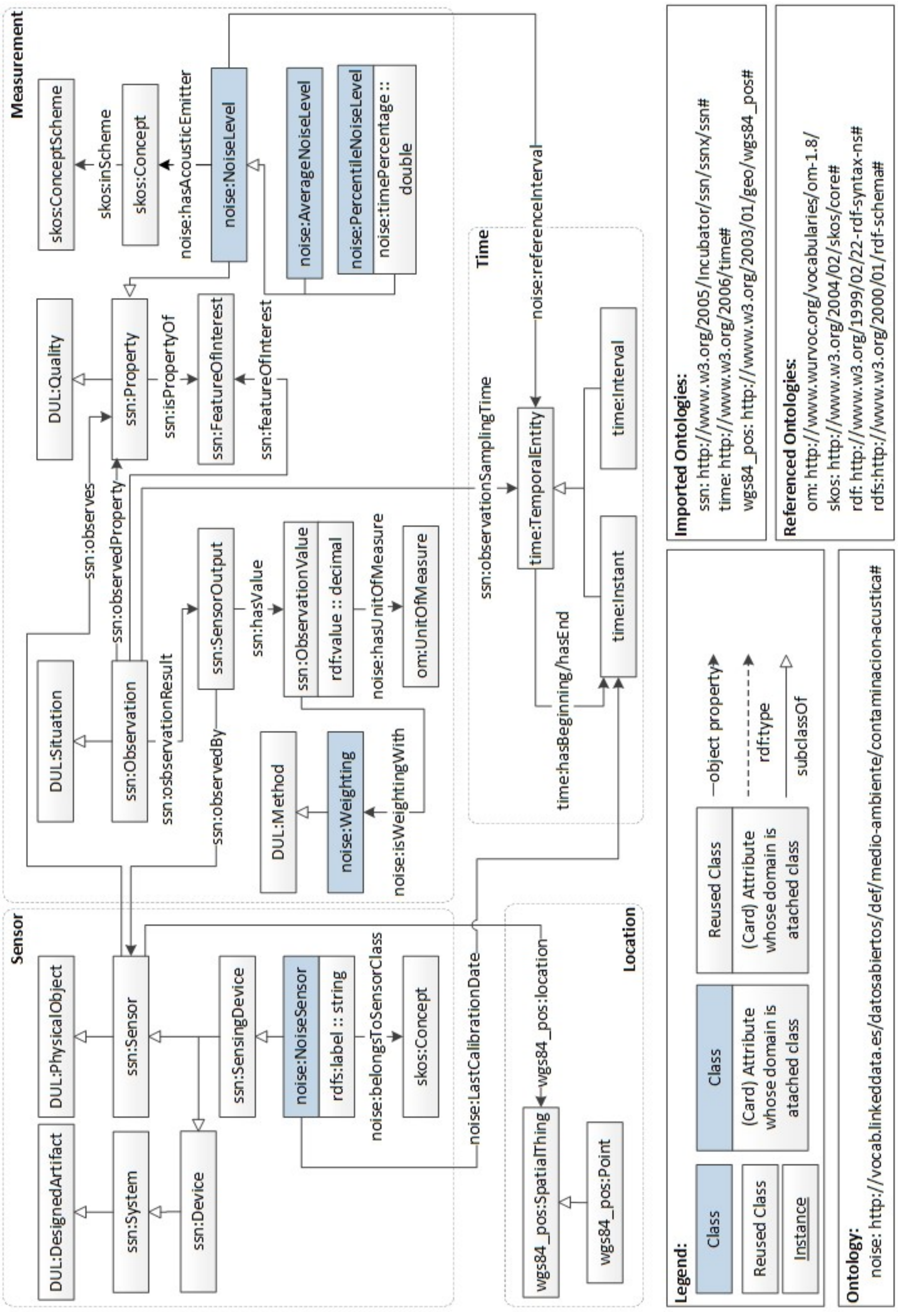
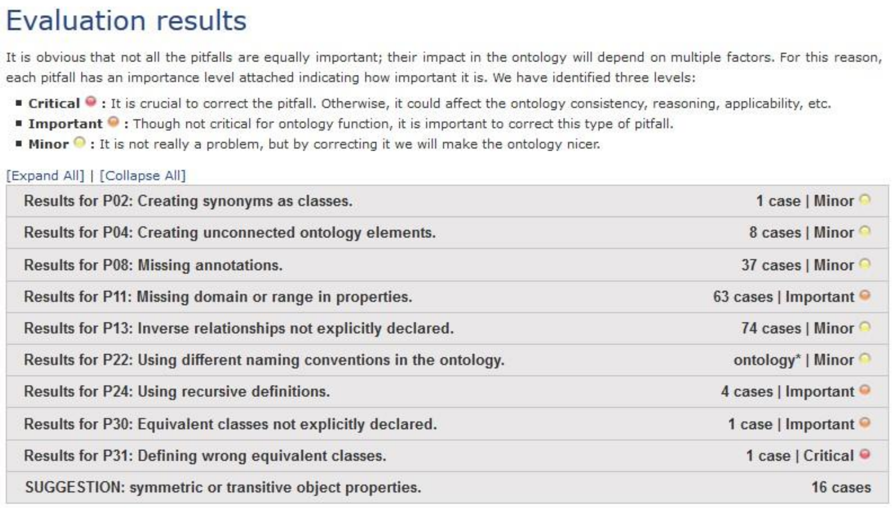


Fig. 3 Final conceptualization of the noise pollution ontology

```
### Prefixes
@prefix : <http://vocab.linkeddata.es/datosabiertos/def/medio-ambiente/contaminacion-acustica#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix ssn: <http://purl.oclc.org/NET/ssnx/ssn#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .

### Ontology metadata excerpt
<http://vocab.linkeddata.es/datosabiertos/def/medio-ambiente/contaminacion-acustica> rdf:type owl:Ontology ;
    owl:imports <http://purl.oclc.org/NET/ssnx/ssn> ,
                <http://vocab.linkeddata.es/datosabiertos/kos/medio-ambiente/contaminacion-acustica> ,
                <http://www.w3.org/2006/time> ;
    <http://creativecommons.org/ns#license> "http://creativecommons.org/licenses/by/4.0/"^^xsd:string ;
    dc:title "Acoustic Pollution"@en ,
             "Contaminacion Acustica"@es .

### Class definition examples
### http://vocab.linkeddata.es/datosabiertos/def/medio-ambiente/contaminacion-acustica#NoiseLevel
:NoiseLevel rdf:type owl:Class ;
            rdfs:subClassOf ssn:Property ;
            rdfs:comment "Sound pressure level measured by the sensor. It is a physical magnitude describing acoustic pollution and related to the produced effects. It is measured in decibels (dB)."@en ;
            rdfs:label "Nivel de Ruido"@es ,
                       "Noise Level"@en .

### http://vocab.linkeddata.es/datosabiertos/def/medio-ambiente/contaminacion-acustica#NoiseSensor
:NoiseSensor rdf:type owl:Class ;
             rdfs:subClassOf ssn:SensingDevice ;
             rdfs:comment "Device that measures the noise level. It detects changes in the air pressure, at a certain location and at a given time, and converts them into electrical pulses. These electric pulses have a frequency, representing the acoustic waves. The frequency range perceived human beings is from 20 Hz to 20 kHz."@en ;
             rdfs:label "Noise Sensor"@en ,
                        "Sensor de Ruido"@es .

### http://vocab.linkeddata.es/datosabiertos/def/medio-ambiente/contaminacion-acustica#PercentilNoiseLevel
:PercentilNoiseLevel rdf:type owl:Class ;
                     rdfs:subClassOf :NoiseLevel ;
                     rdfs:comment "Weighted frequency and time sound pressure level exceeding the N% of time of the considered interval."@en ;
                     rdfs:seeAlso "AENOR: UNE-ISO 1996-1"@es .
```

Listing 1 Noise pollution ontology code excerpt in Turtle format

Finally, as shown in Figure 4, we evaluated the ontology using OOPS! and retrieved several pitfalls. We fixed the pitfalls related to our ontology. However, we did not fix those related to the reused ontologies because we do not have authority over them. Moreover, the tool only detected one critical pitfall which corresponds to two concepts of SSN ontology but we are not reusing them therefore it did not represent a problem for our model. Last, in the case of important and minor pitfalls related to the reused ontologies, they are not crucial for ontology function thus they did not affect our ontology.



Evaluation results

It is obvious that not all the pitfalls are equally important; their impact in the ontology will depend on multiple factors. For this reason, each pitfall has an importance level attached indicating how important it is. We have identified three levels:

- **Critical** : It is crucial to correct the pitfall. Otherwise, it could affect the ontology consistency, reasoning, applicability, etc.
- **Important** : Though not critical for ontology function, it is important to correct this type of pitfall.
- **Minor** : It is not really a problem, but by correcting it we will make the ontology nicer.

[Expand All] | [Collapse All]

Results for P02: Creating synonyms as classes.	1 case | Minor
Results for P04: Creating unconnected ontology elements.	8 cases | Minor
Results for P08: Missing annotations.	37 cases | Minor
Results for P11: Missing domain or range in properties.	63 cases | Important
Results for P13: Inverse relationships not explicitly declared.	74 cases | Minor
Results for P22: Using different naming conventions in the ontology.	ontology* | Minor
Results for P24: Using recursive definitions.	4 cases | Important
Results for P30: Equivalent classes not explicitly declared.	1 case | Important
Results for P31: Defining wrong equivalent classes.	1 case | Critical
SUGGESTION: symmetric or transitive object properties.	16 cases

Fig. 4 Results of the noise pollution ontology evaluation performed in OOPS!

3.3 Ontology publication

The aim of the ontology publication activity is to make the ontology available on-line both as human-readable documentation and in a machine-readable format. Both versions of the ontology should be reachable from the ontology URI by means of content negotiation mechanisms. It should be noted that the ontology machine-readable version has been obtained during the implementation activity, however the human oriented documentation should be generated in a previous step to the publication. Such documentation is usually composed of HTML pages describing the content of the ontology and may include diagrams and examples to improve ontology readability and reusability.

3.3.1 Ontology publication of the noise pollution ontology

Although the URI strategy definition is performed during the encoding activity, we will explain it in this subsection in order to have in mind these definitions at the publication time. We defined the persistent URIs for the features of the noise pollution ontology according to the best practices described in [27] and the Spanish Technical Interoperability Standard [24]. The base URI for all elements in the ontology is *http://vocab.linkeddata.es/datosabiertos/def/medio-ambiente/contaminacion-acustica#*. We also followed an upper camel case strategy to name classes and a lower camel case strategy for object and data properties and resources, as it can be observed in Figure 3.

Once we encoded the ontology, we performed the tasks related to the publication activity. For this purpose we used OnToology [2], a web-based system that builds on top of Git-based environments and integrates a set of existing tools for documentation, evaluation and publication activities. These tools are Widoco [14] for generating the HTML documentation, AR2DTool[14] for generating diagrams, and OOPS! for evaluating the ontology. In addition, OnToology provides two alternatives for ontology publication, namely: publishing the ontology with a permanent id using the *https://w3id.org* services or downloading a bundle with all the files needed to publish the ontology in a server. For the noise ontology publication the latter option was selected.

The noise pollution ontology was published on the Web and it is available under its URI *http://vocab.linkeddata.es/datosabiertos/def/medio-ambiente/contaminacion-acustica* as a machine-readable file and a human-readable documentation, using content negotiation. This human-readable documentation in HTML, generated with Widoco, is provided in English and Spanish. Figure 5 presents a screenshot of the English version of ontology documentation in HTML format.

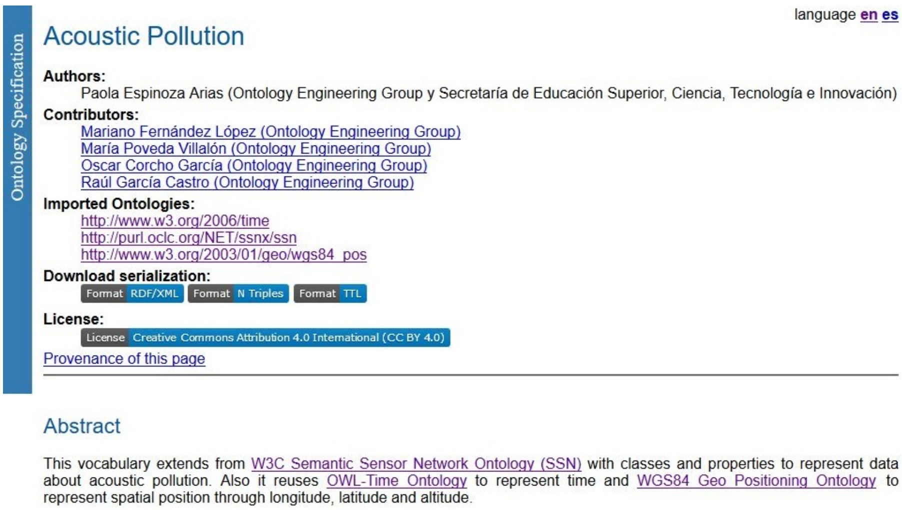


Fig. 5 English version of the noise pollution ontology HTML documentation

3.4 Ontology maintenance

The goal of this activity is to update the ontology after the last release. This may be needed due to different situations, for example: new requirements iden-

[14] Another RDF to diagram tool (AR2DTool) `https://doi.org/10.5281/zenodo.1317796`

tification, bugs detection, improvement suggestions, etc. This activity may be triggered during the ontology development process in which the new requirements implementation or bug fixing would be scheduled in one or more sprints. This activity may be also triggered after the ontology development process in which a new version or revision of the ontology should be generated.

3.4.1 Ontology maintenance in the noise pollution ontology

In order to support the maintenance of the ontology, we used the GitHub issue tracker which keep control of the list of issues. When an ontology developer, domain expert, user, etc. wants to add new requirements, changes, improvements, etc. of the ontology they should create a new issue in the GitHub repository.

All opened issues are discussed by the ontology development team and if the issue is agreed the proposal is implemented over the ontology. For example, in our GitHub repository tracker[15] an issue proposes an ontology actualization according to the new version of the SSN ontology, namely the SSN/SOSA.[16] This proposal has been agreed by the development team, and it is currently being implemented in a new ontology version.

4 Noise pollution ontology in practice

In this section, we provide an example of how users can instantiate the noise pollution ontology. This example represents the stage where a sensor collects measurements of the average noise level (Ld) captured during April 2017 at the *Paseo de Recoletos* boulevard in Madrid. This example is depicted in Figure 6. The prefix "ex" represents the namespace of this example.

As the SSN ontology is being reused, it is important to bear in mind that the observation concept (`ssn:Observation`) allows the representation of the measurement of a specific property (`ssn:Property`) of a feature of interest (`ssn:FeatureOfInterest`). Following this context, the instance observation `ex:NoiseSensorObservation_LdInterval201704`, aims at collecting the average noise level in the daytime (Ld), instantiated as `ex:LdESP`, as the observed property for the feature of interest `ex:FeatureOfInterestPaseoDeRecoletos` (*Paseo de Recoletos*).

This observation has been performed by `ex:NoiseSensor`, whose output, the result of the observation, is modeled as a sensor output instantiated by `ex:NoiseSensorSensorOutput_ObservationLdInterval201704`. In addition, the `ex:NoiseSensorObservationValue_ObservationLdInterval201704` instance represents the observation value itself, which allows specifying their decimal value, unit of measure and weighting method. Finally, the time interval where this observation has been performed, April 2017, is instantiated as

[15] Noise ontology Github tracker `https://github.com/opencitydata/medio-ambiente-contaminacion-acustica/issues`

[16] The SSN/SOSA ontology `https://www.w3.org/TR/vocab-ssn`

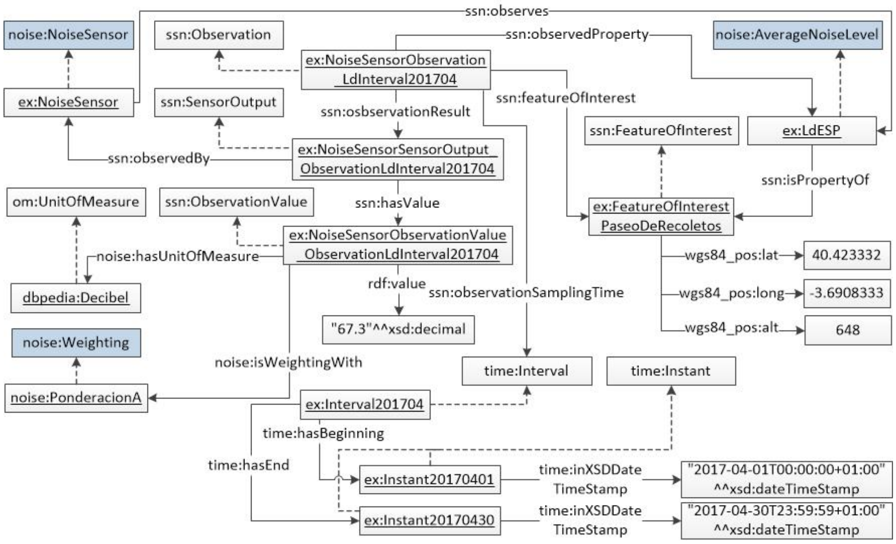


Fig. 6 Example of instantiation of an observation

`ex:Interval201704`, which is modeled according to the OWL-Time ontology specification.

The code of this example is available into the resource documentation folder[17] of our GitHub repository. Additional examples are provided in the HTML documentation of the noise pollution ontology.

5 Discussion

Methodologies for building ontologies are designed to guide developers through the whole process aiming at converting the art of building ontologies on an engineering process. However, there are still activities that require of developers know-how and aptitudes like for example the conceptualization phase. In other cases, the applicability of a methodology depends on the quantity and quality of the resources available for the project. The development of the noise ontology has not been exempt from these situations. In the rest of this section, some insights about the ontology development process and future issues are discussed.

First, during the ontology requirements specification activity the Spanish open data portals were checked in order to understand the noise datasets published by the city halls. However, most of these portals do not provide proper documentation describing the data in order to understand the dataset in a more accurate way. Thus, the collaboration with domain experts becomes

[17] Examples folder of noise GitHub repository `https://github.com/opencitydata/medio-ambiente-contaminacion-acustica/tree/master/OnToology/contaminacion-acustica.owl/documentation/resources/examples`

crucial for the correct understanding of data. While it is true that it is advisable to count with domain experts in any ontology development project, it should be noted that domains experts time is a costly resource which not all the project could afford.

Second, during the ontology implementation activity, several ontologies were reused, nevertheless, the search and selection of potential ontologies to be reused is by no means a trivial task. Although available ontology registries and indexes speed up the searching task, the analysis of retrieved ontology and their selection are still time consuming. This selection highly relies on the experience of the ontology engineer and it involves the evaluation of some criteria such as quality, trustworthy, availability, etc. of the ontologies. In this regards, the development of better tools which eases the search and selection process is an open issue in the ontology engineering field.

Third, during the ontology maintenance activity, ontology changes are allowed. Sometimes these changes may be originated by external changes related to reused ontologies. As in the noise ontology case, as was explained in subsection 3.4.1, the new SSN/SOSA ontology [16] will originate a new noise ontology version. The replacement of the SSN/SOSA ontology by its newer version would impact the presented noise ontology mostly in the following ways: a) the noise ontology would become lighter and b) most of the pitfalls detected in the noise ontology evaluation subsection 3.2.1 will not appear in the new version of the ontology. These two implications are due to the fact the new version discarded the dependency on the DOLCE Ultralite ontology.

Finally, it is also important to consider how to persist data consistency through future versions of the ontology. One aspect to bear in mind may be to ensure backward compatibility when developing the new ontology version. This compatibility implies that the semantics of the ontology changes, but the new ontology is backward compatible with the previous version [9]. Therefore, interpretation of data according the new ontology would be the same as when using the old ontology. In addition, data may include metadata including the ontology version used to represent it. In addition, it may be helpful to use a version control system to allow access to the dataset version valid at a certain time, it means, when the old ontology was operational. However, the lack of practical guidelines and best practices for ontology evolution in different scenarios hinder the reduction of the quality issues introduced by the changes while evolving an ontology [20]. Furthermore, it remains necessary to integrate the existing solutions and tools in order to achieve a platform that could help in the process of ontology evolution [29].

6 Conclusions and future work

In this work, the process followed to develop an ontology for representing the environmental noise pollution domain has been described. In addition, the noise pollution ontology itself has been presented. At the beginning, this ontology aimed at representing noise data from open data sources over the

Spanish territory, however, we modeled it in a general way in order to be also reusable in other territories and to cover important concepts related with the noise sensor devices.

The noise pollution ontology represents a step forward with regard to the state of art ontologies since it allows representing several noise levels, specific details about noise sensors (e.g. weighting methods, sensor classes, etc.) and different types of acoustic emitters. An important aspect of the noise pollution ontology is their alignment with international standards and directives, and the reuse of well-known ontologies. In this context, it is expected to achieve broader interoperability instead of using ad-hoc models to represent such data. Finally, reusing well known, adopted and tested ontologies helped us during the development of our ontology since it allows reducing time and effort instead of developing them from scratch.

In addition, this ontology has been included as an example in the open data guide [11] published by the Spanish Federation of Municipalities and Provinces (FEMP), which aims to present a work itinerary on the opening of data and its reuse for all Spanish local administrations.

Finally, as current work, we are developing the new version of the noise pollution ontology according to the SSN/SOSA ontology. We also plan to perform a revision of the ontology to determine its possible extension in order to represent new requirements or use cases, for example, to cover the representation of noise maps.

Acknowledgements This work was partially supported by a Predoctoral grant from the I+D+i program of the Universidad Politécnica de Madrid and DATOS 4.0: RETOS Y SOLUCIONES - UPM Spanish national project (TIN2016-78011-C4-4-R). We also acknowledge the contribution of Luis Gasco Sánchez and Mariano Fernández López.

Conflicts of Interest The authors declare no conflict of interest.

References

1. Agarwal, R., Gomez-Fernandez, D., Elsaleh, T., Gyrard, A., Lanza, J., Sanchez, L., Georgantas, N., Issarny, V.: Unified iot ontology to enable interoperability and federation of testbeds. In: Internet of Things (WF-IoT), 2016 IEEE 3rd World Forum on, pp. 70–75. IEEE (2016)
2. Alobaid, A., Garijo, D., Poveda-Villalón, M., Santana-Perez, I., Fernández-Izquierdo, A., Corcho, O.: Automating ontology engineering support activities with OnToology. Journal of Web Semantics (2018)
3. BOE: Ley 37/2003 de 17 de noviembre del Ruido (276), 40494–40505 (2003)
4. Brickley, D.: Basic geo (wgs84 lat/long) vocabulary. Technical report (2006)
5. COM(2017) 151 final: On the implementation of the environmental noise directive in accordance with article 11 of Directive 2002/49/EC. Tech. rep., European Comission (2017)
6. Compton, M., Barnaghi, P., Bermudez, L., García-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., et al.: The ssn ontology of the w3c semantic sensor network incubator group. Web semantics: science, services and agents on the World Wide Web **17**, 25–32 (2012)
7. Cox, S., Little, C.: Time ontology in owl. World Wide Web Consortium. Retrieved from https://www. w3. org/TR/owl-time (2017)

8. Dietrich, D., Gray, J., McNamara, T., Poikola, A., Pollock, P., Tait, J., Zijlstra, T., et al.: Open data handbook. Open Knowledge International (2009)
9. Ding, Y., Fensel, D., Klein, M., Omelayenko, B.: Ontology management: survey, requirements and directions. On-to-Knowledge Project (2001)
10. Directive EU: Directive 2002/49/EC of the European parliament and the Council of 25 June 2002 relating to the assessment and management of environmental noise. Official Journal of the European Communities, L **189**(18.07), 12–25 (2002)
11. FEMP: Datos Abiertos.- Guía estratégica para su puesta en marcha. Conjunto de datos mínimos a publicar. `http://femp.femp.es/files/3580-1617-fichero/Gu%C3%ADa%20Datos%20Abiertos.pdf` (2017)
12. Fox, M.S.: Journal of Web Semantics **48**, 48 – 65 (2018). DOI 10.1016/j.websem.2018.01.001. URL `http://www.sciencedirect.com/science/article/pii/S1570826818300015`
13. García-Castro, R., Fernández-Izquierdo, A., Heinz, C., Kostelnik, P., Poveda-Villalón, M., Serena, F.: D2.2 Detailed Specification of the Semantic Model. Tech. rep., Universidad Politécnica de Madrid (UPM) (2017). VICINITY Project. https://vicinity2020.eu
14. Garijo, D.: Widoco: a wizard for documenting ontologies. In: International Semantic Web Conference, pp. 94–102. Springer (2017)
15. Grüninger, M., Fox, M.S.: Methodology for the design and evaluation of ontologies. In: IJCAI'95, Workshop on Basic Ontological Issues in Knowledge Sharing (1995)
16. Haller, A., Janowicz, K., Cox, S.J., Lefrançois, M., Taylor, K., Le Phuoc, D., Lieberman, J., García-Castro, R., Atkinson, R., Stadler, C.: The modular SSN Ontology: A Joint W3C and OGC Standard Specifying the Semantics of Sensors, Observations, Sampling, and Actuation. Semantic Web Journal pp. 1–24 (2018). DOI 10.3233/SW-180320
17. Karpus, A., Vagliano, I., Goczyła, K., Morisio, M.: An ontology-based contextual prefiltering technique for recommender systems. In: Computer Science and Information Systems (FedCSIS), 2016 Federated Conference on, pp. 411–420. IEEE (2016)
18. level meters-Part, Electroacoustics-Sound: 1: Specifications, international standard iec 61672-1: 2013. Tech. rep., Tech. Rep (2013)
19. level meters-Part, Electroacoustics-Sound: 2: Pattern evaluation tests, international standard iec 61672-2: 2013. Tech. rep., Tech. Rep (2013)
20. Mihindukulasooriya, N., Poveda-Villalón, M., García-Castro, R., Gómez-Pérez, A.: Collaborative ontology evolution and data quality - an empirical analysis. In: M. Dragoni, M. Poveda-Villalón, E. Jimenez-Ruiz (eds.) OWL: Experiences and Directions – Reasoner Evaluation, pp. 95–114. Springer International Publishing, Cham (2017)
21. Poveda-Villalón, M.: A reuse-based lightweight method for developing linked data ontologies and vocabularies. In: E. Simperl, P. Cimiano, A. Polleres, O. Corcho, V. Presutti (eds.) The Semantic Web: Research and Applications, pp. 833–837. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
22. Poveda-Villalón, M., Gómez-Pérez, A., Suárez-Figueroa, M.C.: OOPS! (OntOlogy Pitfall Scanner!): An On-line Tool for Ontology Evaluation. International Journal on Semantic Web and Information Systems (IJSWIS) **10**(2), 7–34 (2014)
23. Rijgersberg, H., Van Assem, M., Top, J.: Ontology of units of measure and related concepts. Semantic Web **4**(1), 3–13 (2013)
24. Spain, M.o.F.P.A.: Technical Interoperability Standard for the Reuse of Information Resources. `https://administracionelectronica.gob.es/pae_Home/dam/jcr:a8d2c143-ce9a-4fc7-afe7-ef5d9ba7c4a1/ENGLISH_Interoperability_Agreement_for%20the%20Reuse%20of%20Information%20Resources.pdf` (2013)
25. Studer, R., Benjamins, V.R., Fensel, D., et al.: Knowledge engineering: principles and methods. Data and knowledge engineering **25**(1), 161–198 (1998)
26. Suárez-Figueroa, M.C., Gómez-Pérez, A., Fernandez-Lopez, M.: The NeOn Methodology framework: A scenario-based methodology for ontology development. Applied ontology **10**(2), 107–145 (2015)
27. Tandy, J., van den Brink, L., Barnaghi, P.: Spatial Data on the Web Best Practices. W3C Working Group Note (2017)
28. Vandenbussche, P.Y., Atemezing, G.A., Poveda-Villalón, M., Vatant, B.: Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the web. Semantic Web **8**(3), 437–452 (2017)
29. Zablith, F., Antoniou, G., d'Aquin, M., Flouris, G., Kondylakis, H., Motta, E., Plexousakis, D., Sabou, M.: Ontology evolution: a process-centric survey. The knowledge engineering review **30**(1), 45–75 (2015)