Style Guidelines for Naming and Labeling Ontologies in the Multilingual Web

Elena Montiel-Ponsoda
Ontology Engineering Group (UPM), Spain
emontiel@fi.upm.es

Daniel Vila-Suero
Ontology Engineering Group (UPM), Spain
dvila@fi.upm.es

Boris Villazón-Terrazas
Ontology Engineering Group (UPM), Spain
bvillazon@fi.upm.es

Gordon Dunsire
Consultant, Scotland, UK
gordon@gordondunsire.com

Elena Escolano Rodríguez
Biblioteca Nacional Española (BNE), Spain
elena.escolano@bne.es

Asunción Gómez-Pérez
Ontology Engineering Group (UPM), Spain
asun@fi.upm.es

Abstract

In the context of the Semantic Web, natural language descriptions associated with ontologies have proven to be of major importance not only to support ontology developers and adopters, but also to assist in tasks such as ontology mapping, information extraction, or natural language generation. In the state-of-the-art we find some attempts to provide guidelines for URI local names in English, and also some disagreement on the use of URIs for describing ontology elements. When trying to extrapolate these ideas to a multilingual scenario, some of these approaches fail to provide a valid solution. On the basis of some real experiences in the translation of ontologies from English into Spanish, we provide a preliminary set of guidelines for naming and labeling ontologies in a multilingual scenario.

Keywords: style guidelines; ontology naming; ontology labeling; multilingual web; linked data

1. Introduction

In the context of the Semantic Web, interoperability has become a major issue, not only because of the diversity of formats in which knowledge resources are expressed, or the differences in granularity or coverage of models, but also because of the linguistic descriptions associated with semantic representations. The labels assigned to classes and properties have proven to be of unquestionable assistance for human understanding, supporting ontology developers and adopters in checking consistency and avoiding inaccuracies. They have also been shown to be of great assistance in tasks such as ontology mapping (Svab and Svatek, 2008), information extraction (Müller et al., 2004), or natural language generation (Boncheva, 2005), to mention a few.

In this contribution we distinguish between the **local names assigned to URIs** of ontology classes and properties (e.g., BusinessEntity in the following URI of the GoodRelations ontology (Hepp, 2010): `http://purl.org/goodrelations/v1#BusinessEntity`) and **labels** given to ontology entities by means of metadata or properties created with that purpose (e.g., `rdfs:label "BusinessEntity"` in OWL ontologies, or `skos:prefLabel "BusinessEntity"` in SKOS resources). For practical reasons we will refer to OWL and use the corresponding terminology, although most of the contributions in this paper could be extrapolated to other web languages.

Local names are often believed to play the role of documenting ontologies in natural language, resulting in a misuse of URIs, as reported in recent discussions in W3C public mailing lists[1][2]. URIs are resource identifiers that do not need to contain meaningful descriptions in natural

[1] http://lists.w3.org/Archives/Public/public-lod/2011Apr/0278.html
[2] http://lists.w3.org/Archives/Public/public-lod/2011Apr/0282.html

language. We come back to this issue in section 2. On the contrary, labels are defined as annotation properties that should be used to provide a human-readable version of a resource's name (Brickley and Guha, 2004). However, as pointed out in Fliedl et al. (2010), annotation properties can hardly be found in OWL specifications, because "People do not like to make additional comments". Instead, URIs' local names are often used to document ontology classes and properties in natural language, which can result in some problems, for example if they are not considered valid anymore and need to be changed.

Apart from the annotations proposed to account for natural language descriptions of ontology classes and properties in the Dublin Core Metadata initiative (DCMI, 1998), in recent years vocabularies such as SKOS (Miles and Bechhofer, 2009) propound ways of assigning lexical descriptions to ontology elements. In particular, the SKOS extension SKOS-XL, introduces a `skosxl:Label` class that allows labels to be treated as first-order RDF resources, allowing the possibility of making assertion about these classes. For instance, relations can be explicitly made between a label in its full form and its acronym. This evolution hints at the importance of properly documenting semantic resources in natural language.

Finally, it is also important to mention the effort carried out by the community in the development of principled models to associate more complex linguistic descriptions with arbitrary ontologies and linked data (see LexInfo (Cimiano et al., 2010), LIR-Linguistic Information Repository (Montiel-Ponsoda et al., 2010), or *lemon* (McCrae et al., 2011)). Though varying in the type of linguistic descriptions accounted for in the different models, they all have been devised as external models that associate linguistic information with ontology elements, reified by building on a meta-model of OWL (Vradenic et al., 2006). For example, the LIR model focuses on capturing variants of terms such as acronyms, short forms or transliterations in several natural languages, so that translation relations can be explicitly established between term variants. The *lemon* model, on the other hand, captures a wider range of linguistic descriptions that range from labels morphological decomposition to syntactic behavior of labels with respect to ontology arguments, and accounts for translation relations among *lemon* lexicons in different languages.

Additionally, these models provide mechanisms for representing linguistic descriptions relative to an ontology in several natural languages, from the simplest option of language tagging offered by RDF literals (`rdfs:label@en`), to the ISO 639 tag (ISO 639-Codes for the representation of names of languages, 2002) associated to the `LexicalEntry` class in LIR or *lemon* to account for the language of the linguistic descriptions associated with ontology elements.

Although there has been some discussion on the actual naming of URIs, we believe that this should be revisited in the scenario of the Multilingual Web[3]. Some conventions have been proposed for English, as will be reviewed in section 2, but not for other languages. The same is true for the use of multilingual labels to document classes and properties in ontologies. The lack of guidelines in this sense has resulted in the implementation of heterogeneous solutions and the presence of inconsistencies within one and the same resource. Thus, there is a need for general guidelines to help OWL ontology designers and domain experts in the task of **naming ontology elements** (i.e., classes, object and data type properties, individuals and so on) and **formatting labels** or terms assigned to those elements in several natural languages. We are also in favor of proposing linguistically motivated guidelines, as some formats may be adequate for some languages but unusual or unnatural for others. The guidelines proposed in this paper are preliminary and mainly focused on latin scripts. Work is needed to extend it to other alphabets.

[3] Note that we are not proposing the translation of URI local names, but only the provision of some style guidelines in case non-English speaking users want to provide meaningful URIs to their ontologies in their own languages. We point the interested reader to the MultilingualWeb project for further actions in this sense http://www.multilingualweb.eu/en

2. Related Work

Some of the first recommendations for naming classes in ontologies are found in Noy and McGuiness (2001)' *Ontology Development 101: A Guide to Creating Your First Ontology*. The authors provide some guidelines that are bound to modeling decisions and the functionalities provided by ontology editors. Basically, they make users aware of the fact that some systems do not allow the same name space (local name) for classes and properties. In this sense, a difference is made by the use of lower case for naming properties and upper case for naming classes.

In the framework of the Open Biomedical Ontologies (OBO) Foundry initiative we find some conventions intended to help solve the problem of ontology integration (Shober et al., 2009). Such an issue is today even more important, if that is possible, in the light of the Linked Data initiative (Bizer et al., 2008), in which entities pointing to the same world object should be linked to bring linked data to its maximum potential. According to Shober et al. (2009), three main factors have hampered the adoption of standards or conventions in naming ontology entities in the biological and medical domains, namely, a) domain specificity, b) document inaccessibility, and c) format and implementation dependency. To help ontology editors in this domain, the authors offer a comprehensive set of naming conventions after conducting a survey with developers of OBO ontologies. We classify the whole set of naming conventions into two types: *conventions on content* and *conventions on format*. Regarding the first, the authors make these recommendations, among others:

- Avoid the use of univocal names, homonyms and conjunctions
- Use positive names or recycle strings rather than using synonyms

The purpose of these recommendations is to avoid inconsistencies within the model. However, we claim that in order to provide some help with tasks such as ontology mapping or semantic annotation, capturing term variety may be of key importance.

As for those conventions that focus on format, the following are of interest to us:

- Use explicit and concise names ("wall of esophagus" instead of "the wall of the esophagus")
- Prefer singular nominal forms
- Use space as word separators or underscores as default (Camel case should not be used as a means of word separation)
- Expand abbreviations and acronyms
- Prefer lower case beginnings for class and property names "as they would appear in normal English written text"
- Avoid character formation (use plain ASCII format, avoid accents)

The authors of these guidelines claim to have been confined by the needs of the OBO ontologies, so that some of the guidelines may not apply in other domains of knowledge. They have also only considered the English language. This becomes obvious if we look at the last recommendation in which they propose the avoidance of accents. This may not be possible for other languages such as Spanish or French in which the same word may be written with or without accent pointing to a completely different concept. In any case, we will discuss the convenience of some of these recommendations for a multilingual scenario in section 4.

More recently, Fliedl et al. (2010) propose general style guidelines for naming OWL classes and properties. Though using the term OWL labels, they are not referring to the annotation property `rdfs:label`, but to the labels given to URIs' local names. The motivation behind this work is the verbalization or rendering of ontologies in natural language taking as input the URIs' local names. The authors review the current state-of-the-art on strategies for naming concepts and individuals analyzing ontologies contained in the DAML Ontology Library[4]. They mainly concentrate on what we dubbed *conventions on format*, as can be seen from the examples below:

[4] http://www.daml.org/ontologies/

Typical format for class labels

– Wine, Wine-grape, Wine-Grape, Wine Grape, WineGrape

From the above-mentioned options they suggest the use of "an upper case working as delimiter", so-called CamelCase notation, and the use of the singular form of nouns.

Typical format for property labels

– madeFromGrape, MadeFromGrape, Made-From-Grape, Made From Grape

For the naming guidelines for property labels, they are in favor of the first option: "Property labels should always start with lower case".

One may agree that the formatting these authors propose for classes and properties is in line with the guidelines endorsed by the Semantic Web community for naming URIs. There has been a lot of discussion and work on defining guidelines for the effective use of URIs, starting with Tim Berners-Lee's article *Cool URIs don't change* (Berners-Lee, 1998). In the article, Berners-Lee introduces the concept of URI design, which has proven to be a challenging issue for the Semantic Web community, by providing some guidelines when assigning URIs to web resources and documents.

In this respect, we find some valuable guidelines presented at *Common HTTP Implementation problems*, a W3C Note in the context of the Technical Architecture Group (Théreaux, 2003), which summarizes, paraphrases and extends many of the ideas from Berners-Lee's article. Among the twelve proposed guidelines, the "Choose URIs wisely" one seems to be a perfect fit when thinking about naming ontology elements for the Web. In this guideline the authors state the following: (1) Use short URIs as much as possible, (2) Choose a case policy, (3) Avoid URIs in mixed case, and (4) As a case policy choose either "all lowercase" or "first letter uppercase".

More recently, with the commitment of some governments[5] to make a large part of their collected data available to the public, other guidelines have emerged to help public organizations to organize, identify and publish their data under the principles of Linked Data. One of the most representative examples is the effort from the UK Cabinet Office "Designing URIs for the public sector"[6]. There we find recommendations regarding the path structure of URIs (create different base URIs to establish a clear separation of different types of data, e.g schools data, roads data, etc.) and some naming conventions. Such conventions are in the same line as the W3C notes but emphasize the need of stability and scalability and make them specific to the task of publishing large amounts of data on the Web. For example, the document recommends the use of the singular form, *school,* instead of the plural, *schools,* when generating the path for schools data and the avoidance of verbose names, like the one in the example below, and to use codes instead:

Use: http://example.org/school/12345.

Instead of: http://example.org/schools/StPaulsCatholicSchool

The following table presents a summary of guidelines relevant to this paper extracted from the reviewed literature:

TABLE 1: URI design guidelines

Guideline	Source
Don't put too much semantics in the URI	(Berners-Lee, 1998)
Don´t bind URIs to some classification or topic (you may change your point of view)	(Berners-Lee, 1998)
Use short URIs as much as possible	(Théreaux, 2003)
Choose a case policy (all lowercase or first letter uppercase)	(Théreaux, 2003)
Use lower case and the singular form	(Cabinet Office, 2009)
Separate words by hyphens	(Cabinet Office, 2009)
Where the reference local name may change, or becomes overly verbose, a code may be more appropriate	(Cabinet Office, 2009)

[5] UK government http://data.gov.uk, and US government http://data.gov

[6] http://www.cabinetoffice.gov.uk/sites/default/files/resources/designing-URI-sets-uk-public-sector.pdf

In conclusion, from the works reviewed in this section we see that most proposals assume the use of meaningful URIs, which is a logical approach, as in practice most ontology have relied on URIs to document ontologies in natural language. However, the trend in most recent proposals is to recommend the use of "less semantics" in the URIs and more opaque elements, as in the Cabinet Office document. In this sense, we believe that labels and other type of linguistic annotations should be given more relevance in order to fully exploit ontologies in the Multilingual Semantic Web.

Thus, if we now try to extrapolate all of these proposals into the Multilingual Web, we encounter different obstacles. Regarding naming conventions for URIs' local names, some questions arise: If we decide to use meaningful local names for our ontology classes and properties, which format should we use for languages other than English? Will the upper case strategy work as *delimiter* for all languages? What about accents and other features of some languages? (The current encoding renders local names useless, e.g., http://geo.linkeddata.es/ontology/R%C3%ADo, which is the URI corresponding to the class *Río* (river in Spanish) from the geolinkeddata.es ontology). Although this is an encoding problem, we believe that it should be approached if meaningful URIs are to be used in languages other than English; a complete overview of these issues and the initiatives being carried out by the W3C and the IETF in that direction can be found in the article "An Introduction to Multilingual Web Addresses"[7]. Otherwise, should we be in favor of the use of opaque local names instead of meaningful ones to avoid bias towards particular languages?

As for ontology labels, there are mechanisms to account for language particularities that can be exploited according to applications' final needs. In this sense, we believe that style guidelines should be proposed by the different language communities to accommodate their requirements. This is even more urgent if ontologies are to be used by NLP applications that have to interact with documents that contain free text. Some examples of this will be shown in the next section.

3. Experiences in the Translation of Local Names and Ontology Labels

In this section, our aim is to report on the experiences we had in the manual translation of ontologies from English into Spanish. The first experience refers to the translation of the well-known FOAF (Friend of a Friend) ontology. The second refers to the conversion of some IFLA (International Federation of Library Associations and Institutions) models to RDF, in particular the so called FRBR family of models (IFLA. FRBR Review Group, 2011), as well as the ISBD (International Standard Bibliographic Description) standard (Willer et al., 2010), and their translation into Spanish (please note that some of this work is still on-going).

3.1 FOAF

For the translation of the FOAF ontology into Spanish we decided to use the ontology editor NeOn Toolkit[8]. This editor provides support for the automatic translation of ontology local names by means of the LabelTranslator plugin (Espinoza et al., 2008, 2009). Currently, the languages supported are English, Spanish, and German. Translations are then stored in the LIR model (Montiel-Ponsoda et al., 2010). However, for convenience, in this case we stored translations as additional labels in the ontology, making use of the `rdfs:label` annotation property. The translated ontology can be accessed at http://www.oeg-upm.net/files/foaf/foaf.owl.

The FOAF[9] ontology relies on meaningful local names for URIs that adopt the CamelCase format. The same literal is then specified as the label for classes and properties, with some formatting modifications. In the case of classes, upper-case letters are maintained for each word, not only in the local name, but also in the label, as for "online account" below:

[7] http://www.w3.org/International/articles/idn-and-iri/
[8] http://neon-toolkit.org/wiki/Main_Page
[9] http://xmlns.com/foaf/spec/

```
<owl:Class rdf:about="http://xmlns.com/foaf/0.1/OnlineAccount">
<rdfs:label>Online Account</rdfs:label>
```

Regarding properties, local names and labels start with lower case, as can be seen in the example for the property "is primary topic of". But, upper case in the local name is eliminated in the label. See below:

```
<owl:ObjectProperty rdf:about="http://xmlns.com/foaf/0.1/isPrimaryTopicOf">
<rdfs:label>is primary topic of</rdfs:label>
```

We also found some inconsistencies in the formatting of names and labels, as well as in the content. For example, some names and labels use underscores to delimit words within a noun phrase. See below:

```
<owl:ObjectProperty rdf:about="http://xmlns.com/foaf/0.1/topic_interest">
<rdfs:label>topic_interest</rdfs:label>
```

Most relations use only nouns (as the case of "phone", below) but really mean a verbal phrase ("has phone"):

```
<owl:ObjectProperty rdf:about="http://xmlns.com/foaf/0.1/phone">
<rdfs:label>phone</rdfs:label>
```

When translating the ontology into Spanish, we realized that looking at the local names was often not enough to understand the meaning of the concept. So, we had to resort to the labels and even to the associated comments to really understand the meaning of concepts and relations. This was the case for many object and data type properties. See the example of "current project" below. The relation means "a person works currently in a project x"

```
<owl:ObjectProperty rdf:about="http://xmlns.com/foaf/0.1/currentProject">
<rdfs:comment>A current project this person works on.</rdfs:comment>
```

This allowed us to translate the property into "trabaja en proyecto actualmente" (works currently in project), which we considered a better label for the property.

```
<rdfs:label xml:lang="es">trabaja en proyecto actualmente</rdfs:label>
```

In any case, local names and labels followed a quite compact style, eliminating articles to make them shorter. This is common for local names, but looks unnatural for labels. It may only be justified by cardinality reasons. In order to understand what we mean by cardinality reasons, let us first define cardinality of object properties in ontologies. Let us assume two ontology classes, A and B, which can be related by means of an object property X. The property X can relate (1) one element of class A to one element of class B, (2) many elements of class A to one element of class B, (3) one element of class A to many elements of class B, or (4) many elements of class A to many elements of class A. In general, the following relation combinations are possible: many-to-many, many-to-one, one-to-many, or one-to-one. Thus, we could adapt the name of the object property so as to conform to the cardinality of the relation, for example "hasLover" vs. "hasLovers", if it is assumed that the object of the relation will account for many lovers.

In this sense, in the label of the relation "trabaja en proyecto actualmente" (works currently in project) would sound more natural either to say "trabaja en un proyecto" (works in a project) or "trabaja en varios proyectos" (works in several projects), but since the object of the relation can be one or many, we have left it undetermined also in Spanish.

Finally, we decided to maintain those articles that do not interfere with modelling issues regarding cardinality constraints, as in the examples that follow. By translating "is primary topic" by "es el asunto principal", in Spanish we only leave room for one (and only one) primary topic.

```
<rdfs:label>is primary topic of</rdfs:label>
<rdfs:label xml:lang="es">es el asunto principal de</rdfs:label>
```

3.2 IFLA FRBR family of models and ISBD

IFLA has been involved in a number of projects to transform bibliographic concepts and metadata into the RDF format and other appropriate syntaxes (see Dunsire and Willer, 2011). As a result of these efforts, RDF representations of several models, namely *FRBR* (Functional Requirements for Bibliographic Records), *FRAD* (Functional Requirements for Authority Data) and *FRSAD* (Functional Requirements for Subject Authority Data), have been created with the aim of supporting resource discovery. These models are based on entity-relationship schemas that are represented as RDF classes and properties. In this section we will briefly describe the conventions adopted for the naming and labeling of these *models*. Then, we will report on the translation of the ISBD standard into Spanish.

3.2.1 FRBR family of models in RDF

Contrary to what happened in the case of the FOAF ontology, classes and properties of these models have opaque URIs. This means that developers and users have to resort to label annotation properties to understand the semantics of the models. The terminology used for class and property labels is based as closely as possible on the relevant source documentation which, for the FRBR family of models, are the published reports. Matching the terminology is intended to make it easier to relate the source documentation to the RDF elements and use it for further information about the context and background of the models. The normalization and standardization of terms or labels in library catalog organization is a practice with a long tradition. Therefore, IFLA developers inspired themselves by available guidelines in Library and Information Science, such as those mentioned in (Svenonius, 2000) and (Guidelines for Subject Authority and Reference Entries, 1993) and adapted them to the needs of the new RDF formats.

Labels for RDF classes from all of the models were based on the names in the reports but with an initial capital letter for each word, as in "Corporate Body". All RDF property labels were in lower case. A standard approach was adopted to create human-readable labels for RDF properties: the label consists of the attribute name preceded by the word "has". For example, the logical attribute "field of activity" assigned to the FRAD entity "family" is represented as an RDF property with the label "has field of activity". For the case of object properties with inverse relations "is … of" was used, as in "is a summary of", the inverse of "has a summary" which is the phrase used in the source documentation.

Some data type properties were assigned to more than one entity in the models. The name of the property was explicitly distinct in some cases in the documentation. For example, "title of the expression" is a property of the FRBR *expression* entity and "title of the manifestation" is a property of the *manifestation* entity. The corresponding labels were created in the same way and remained distinguishable: "has title of the expression" and "has title of the manifestation ".

In other cases, the documented property name was not distinct. For example, FRAD assigns the property "address" to both the person and corporate body entities. Just prefixing "has" would have resulted in identical property labels within the same namespace, which could be misleading. In these cases the entity name was added to the label to give distinguishable labels: "has address (person)" and "has address (corporate body)". This approach was used throughout the RDF representations to ensure that all property labels were unique within the namespace.

3.2.2 ISBD translation

ISBD was also transformed to RDF, and the labels for classes and properties followed the same conventions as the ones applied in the case of the FRBR models. The translation of ISBD into Spanish was partly the result of a direct translation of the English labels, and partly took into account the Spanish source documentation. As for the translation of ISBD labels and properties, the language tagging facility of RDF(S) was used, as shown in the example below:

```
<rdfs:label xml:lang="en">has content form</rdfs:label>
<rdfs:label xml:lang="es">tiene forma del contenido</rdfs:label>
```

Attempts to follow the same capitalization pattern for class labels were discarded, however, as the result was less natural in Spanish, and only the first word has an initial capital letter, as in:

```
<rdfs:label xml:lang="es">Esquema de codificación del área de edición
```

```
</rdfs:label>
```

ISBD also defines a small number of controlled vocabularies for data type properties in the aggregated statement "content form and media type area". These have been represented in RDF using SKOS (see, for example, http://metadataregistry.org/vocabulary/show/id/113.html). The Spanish translation of these vocabularies revealed a problem which was not apparent in the English terms. Some of the vocabularies qualify terms from other vocabularies in instances of the aggregated statement. For example, "cartographic" qualifies terms from the content form vocabulary. All the terms for qualifiers are adjectives which require masculine (*cartográfico*) and feminine (*cartográfica*) forms in translation into Spanish. But only one SKOS preferred label is allowed for each language. Because of this, compounds such as "cartográfico/a" were suggested. However, this is not the natural way in which this would appear in free text, for example. Therefore, ways of avoiding the use of such compounds are being currently discussed. Models such as the ones introduced in section 1 (LIR or *lemon*) would allow for the inclusion of the two forms, the masculine and the feminine, helping to solve the problem. For example, in *lemon* both adjectival forms, the masculine and the feminine, would be linked to that property in the ontology by means of a LexicalEntry with two LexicalForms (the masculine and the feminine), and the model would be able to represent that these are form variants of the same lexical entry. The adoption of such a model still needs to be explored.

4. Preliminary Version of Proposed Guidelines

Considering the examples presented in section 3, and taking into account the state-of-the-art summarized in section 2, in this section our aim is to propose some preliminary guidelines for naming and labeling ontology elements in a multilingual scenario. We make a clear distinction between URI local names and labels, as annotation properties or as part of external linguistic models related to ontology elements. We also try to motivate the decisions taken. When more than one option is considered valid according to linguistic criteria, developers are the ones to choose which option better meets their needs.

4.1. Naming

There are some reasons to favor of the use of *meaningful URI local names*, and others against, but the truth is that URIs are meant to be identifiers, so that when using natural language for defining them, they should be kept short, compact (CamelCase policy), and with "not much semantics", as summarized in Table 1. This may not be so trivial for certain domains. See for example the taxonomies of the IFRS (International Financial Reporting Standard) in the financial domain with local names of more than 10 tokens, as in the example below for the label *Minimum lease payments payable under non-cancellable operating lease, end of period later than one year and not later than five years*

```
ifrs_MinimumLeasePaymentsPayableUnderNoncancellableOperatingLeaseEndOfPeriodLat
erThanOneYearAndNotLaterThanFiveYears
```

Some reasons supporting the use of meaningful URI local names are that they help developers to quickly understand the ontology, they are easy to remember, and their use is favored by most ontology editing tools. There are also some main arguments against their use and for opting for *opaque URI local names*, namely, that they are intended for machine consumption and make ontologies stable, they should not be modified once the ontology has been published and adopted by a community of users (unless the actual meaning of concepts has changed), and that they are difficult to understand anyway even when using natural language, as was the case with some property relations in the FOAF ontology (see section 3.1).

In the specific scenario of the Multilingual Web, we may find some more reasons in favor of opaque URI local names. In the past, most ontologies have used English as default language for URI local names, and this is in fact proposed as a guideline in Flield et al. (2010). However, we are currently witnessing an increase of ontologies published in languages other than English, a trend started by initiatives such as Linked Data that encourages governments and institutions all over the world to make their data publicly available. We will have to assume that when using meaningful local names, these agents will use their own languages. For certain domains, this is also natural, since conceptualizations differ from culture to culture, and this is reflected in the language used to describe them; for some thoughts about the reuse of available conceptualizations in different cultural and language settings we refer the interested reader to (Cimiano et al., 2009).

Therefore, leaving technical problems aside, we believe that the use of opaque local names avoids any bias towards English (or any other language) and is a better option for ontologies that might support natural language descriptions in several languages. Indeed, this solution has been adopted by multilingual semantic resources such as EuroWordNet (Vossen, 1998) and the Agrovoc Thesaurus of the FAO (Liang, et al., 2008), and was also the solution provided in the transformation of the FRBR models and the ISBD standard into RDF, as reported in section 3.2.2. In this case, *labels* are to be used to document the ontology in natural languages, as also recently suggested by Tim Berners-Lee[10].

Should it be the case that designers decide to employ meaningful local names because they do not want to add further labels and comments in the ontology, then we believe that characters (accents, etc.) should be represented in local names, as being not only natural but essential for certain languages. Finally, we summarize the proposed guidelines in Table 2.

TABLE 2: Element naming guidelines

Guidelines for the Multilingual Web	Example
Use preferably opaque URI local names	`http://aims.fao.org/aos/agrovoc/c_330988`
When using meaningful URI local names, use your own language with diacritics, etc[11].	`http://geo.linkeddata.es/ontology/río`
When using meaningful URI local names, use underscore or CamelCase for word delimiter, depending on what reads more easily in your own language[12]	`http://geo.linkeddata.es/ontology/aguas_corrientes`

4.2. Labeling

The main advantage of labels is that they give us leeway in describing classes and properties of ontologies, so that they are not only made understandable for developers and consumers, but also can be reliably reused in many tasks involving natural language processing, as mentioned in the introduction of this paper. In the case of the conversion of FRBR models into RDF and the translation of the ISBD standard into Spanish, the approach followed for the selection of labels was to remain close to the terminology of the source documentation. This means providing labels that guarantee human readability and that can be easily matched in free text. Similar criteria were applied in the translation of the FOAF ontology. Some labels were not direct translations of the English labels, but paraphrases in Spanish to make the ontology more understandable to potential Spanish-speaking users.

In line with this, we revisit some of the guidelines proposed in the state-of-the-art and reformulate them to accommodate a multilingual scenario (see Table 3). We believe that labels for classes and properties should follow language conventions in any case. For classes we propose to use the singular form of nouns, since this is the way in which concepts are captured in dictionaries and encyclopedias, at least in most European languages. We would also recommend spaces as word delimiters, since it supports readability. The use of upper case should be determined from language conventions. So, for example, in Spanish only named entities would be capitalized. We would also be in favor of allowing the inclusion of as many labels as considered useful for the final applications, as synonyms. As mentioned for the case of the ISBD standard

[10] http://lists.w3.org/Archives/Public/public-lod/2011Apr/0282.html

[11] This guideline should be analyzed for non-European languages and extended or modified if needed.

[12] This should be reviewed in the case of languages such as Chinese in which spaces do not exist as word delimiters within one nominal group.

translation, this was achieved by the use of SKOS labels, but it only allows for one preferred label. In case more complex descriptions are required, we would propose the use of more complex representation of lexical and terminological descriptions (SKOS-XL, LIR, *lemon*, etc.) as suggested in section 1.

As for properties, we would recommend the use of verbal phrases and the addition of the predicate or range of the relation for disambiguation purposes. Again, this may be valid for most European languages, but should be analyzed for others. For languages such as Spanish in which some verbal forms must agree in gender and number with the subject or domain of the relation, we propose the inclusion of several labels to account for them. See examples below:

```
<rdfs:label xml:lang="es">cartográfico</rdfs:label>
<rdfs:label xml:lang="es">cartográfica</rdfs:label>
```

TABLE 3: Element labeling guidelines

Guidelines for the Multilingual Web	Example
Use the singular form for nouns	skos:prefLabel "financial asset"@en
Use space as word delimiters, or follow your own language conventions	
Use upper or lower case according to your own language conventions	rdfs:label "autor"@es rdfs:label "Miguel de Cervantes"@es
Add as many labels as needed for classes and properties	ex:FAOlabel1 rdf:type skosxl:Label; skosxl:literalForm "Food and Agriculture Organization"@en. ex:FAOlabel2 rdf:type skosxl:Label; skosxl:literalForm "FAO"@en

5. Conclusions

In this paper our aim was to review the state-of-the-art on strategies and guidelines for names and labels of ontology elements. The list of proposed guidelines is not intended to be exhaustive, and may only cover some European languages, but it aims at encouraging the community to put in some effort in this direction. We try to justify our proposal by relying on real examples extracted from our own experiences in the translation of ontologies. We believe that today the Multilingual Web is a reality, and guidelines are needed that take into account the linguistic background of developers. This will not only help them in the arduous activity of ontology design and development, but will also benefit consumers in the adoption of semantic models and technologies.

Acknowledgements

This work is supported in part by the European Union under Grant No. 248458 for the Monnet project, and by the Spanish Ministry of Science and Innovation under Grant No. TIN2010-17550 for the BabeLData project.

References

Bizer, Christian, Richard Cyganiak, and Tom Heath. (2008). How to Publish Linked Data on the Web. Retrieved April 2011, from http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/

Bontcheva, Kalina. (2005) Generating tailored textual summaries from ontologies The Semantic Web: Research and Applications, Springer (pp. 531-545).

Cimiano, Philip, Paul Buitelaar, John McCrae, and Michael Sintek. (2010). LexInfo: A Declarative Model for the Lexicon-Ontology Interface. Journal of Web Semantics, 2010.

Cimiano, Philipp, Elena Montiel-Ponsoda, Paul Buitelaar, Mauricio Espinoza, and Asunción Gómez-Pérez. (2009). A note on ontology localization. Journal of Applied Ontology, 5 (pp. 127–137).

DCMI. (1998). Dublin Core Metadata Element Set, version 1.0: Reference description. Retrieved April 11, 2011, from http://www.dublincore.org/documents/1998/09/dces/.

Brickley, Dan and Guha, R.V. (2004) RDF Vocabulary Description Language 1.0: RDF Schema. Available at: http://www.w3.org/TR/2004/REC-rdf-schema-20040210/

Dunsire, Gordon, and Mirna Willer. (2011). Standard library metadata models and structures for the Semantic Web. In Library hi tech news, Vol. 28, No. 3.

Espinoza, Mauricio; Montiel-Ponsoda, Elena and Gómez-Pérez, Asunción. (2009) Ontology Localization Proceedings of the 5th International Conference on Knowledge Capture (KCAP09), (pp. 33-40)

Espinoza, Mauricio, Gómez-Pérez, Asunción and Mena, Eduardo. (2008). LabelTranslator - A Tool to Automatically Localize an Ontology The Semantic Web: Research and Applications, Springer, (pp.792-796).

Flied, Gunther, Christian Kop and Jrgen Vhringer. *From OWL Class and Property Labels to Human Understandable Natural Language*. In Proceeding of 12th International Conference on Applications of Natural Language to Information Systems, NLDB 2007.

Guidelines for Subject Authority and Reference Entries. (1993). Working Group on "Guidelines for Subject Authority Files" of the Section on Classification and Indexing of the IFLA Division of Bibliographic Control.62 pages. Hardbound. ISBN 10 : 3-598-11180-0

Hepp, Martin. (2010).GoodRelations Language Reference. V 1.0, Release 2010-09-16. Available at: http://www.heppnetz.de/ontologies/goodrelations/v1

IAS. (2007). International Accounting Standards Board, 2007. International Financial Reporting Standards 2007 (including International Accounting Standards (IAS) and Interpretations as at 1 January 2007).

IFLA. FRBR Review Group (2011). Functional requirements: the FRBR family of models Available at: http://www.ifla.org/node/2016

ISO 639 - Codes for the representation of names of languages. (2002).

Liang, Anita., Boris. Lauser, Margherita Sini, Johannes. Keizer, Stephen. Katz. (2008). From AGROVOC to the Agricultural Ontology Service/Concept Server. An OWL model for managing ontologies in the agricultural domain. In Proceedings of the OWL: Experiences and Directions Workshop. Manchester, UK.

McCrae, John, Guadalupe Aguado de Cea, Paul Buitelaar, Philipp Cimiano, Thierry Declerck, Asunción Gómez-Pérez, Jorge Gracia, Laura Hollink, Elena Montiel-Ponsoda, Dennis Spohr, and Tobias Wunner. (2011). Interchanging Lexical Resources in the Semantic Web. Language Resources and Evaluation (in press).

Miles, Alistar, Sean Bechhofer (2009) SKOS-Simple Knowledge Organization System Reference. Retrieved April 11, 2011, from http://www.w3.org/TR/skos-reference/

Montiel-Ponsoda, Elena, Guadalupe Aguado de Cea, Asunción Gómez-Pérez, and Wim Peters. (2010). Enriching Ontologies with Multilingual Information. Journal of Natural Language Engineering, DOI:10.1017/S1351324910000082.

Müller, Hans-Michael, Eimear E. Kenny, and Paul W. Sternberg. Textpresso: An Ontology-Based Information Retrieval and Extraction System for Biological Literature PLoS Biol, 2004, 2, e309.

Sauermann, Leo, and Richard Cyganiak. (2008, December) Cool URIs for the Semantic Web. Note NOTE-cooluris-20081203. World Wide Web Consortium,

Sauermann, Leo, Cyganiak, Richard, Ayers, Danny, Max Völkel. (2008). Cool URIs for the Semantic Web. W3C Interest Group Note 03 December 2008. World Wide Web Consortium. Retrieved April, 2011, from http://www.w3.org/TR/cooluris/

Shober, Daniel, Barry Smith, Suzanna E. Lewis, Waclaw Kusnierczyk, Jane Lomax, Chris Mungall, Christ F. Taylor, Philippe Rocca-Serra, and Susanna-Assunta Sansone. (2009). Survey-based naming conventions for use in OBO Foundry ontology development. BMC Bioinformatics 2009, 10:125. DOI: 10.1186/1471-2105-10-125.

Svab-Zamazal Ondrej and Svatek Vojtech. (2008). Analysing Ontological Structures through Name Pattern Tracking. In: EKAW 2008 - 16th International Conference on Knowledge Engineering and Knowledge Management. Springer LNCS, pp. 213-228.

Svenonius, Elaine. (2000). The intellectual foundation of information organization. Cambridge, MA: MIT Press.

Théreaux, Olivier. (2003). Common http implementation problems. W3C note, World Wide Web Consortium. Retrieved April 2011, from http://www.w3.org/TR/chips/.

Vrandecic, Denny, Völker, Johanna, Haase, Peter, Duc, Thanh, and Cimiano, Philip. (2006). A metamodel for annotations of ontology elements in OWL-DL. In Proceedings of the 2nd Workshop on Ontologies and Meta-Modeling. GI Gesellschaft für Informatik.

Vossen, Piek. (1998). Introduction to EuroWordNet. In N. Ide, D. Greenstein, and P. Vossen (Eds.), Special Issue on EuroWordNet (Vol. 32 (2-3), pp. 73-89).

Willer, Mirna., Gordon Dunsire,. and Boris Bosančić. (2010) ISBD and the Semantic Web. In JLIS.it Journal of Library and Information Science. Italy, Vol. 1 No. 2, (pp. 213-236). DOI: 10.4403/jlis.it-4536