Lights and shadows in creating a glossary about ontology engineering

Mari Carmen Suárez-Figueroa, Guadalupe Aguado de Cea and Asunción Gómez-Pérez

This paper addresses the lack of an explicitly agreed and defined terminology in the ontology engineering field, and particularly, the need for a glossary, which consists of terms and definitions for actions when developing ontologies. The novelty of this paper lies in the precise description of a methodology for building a glossary with the processes and activities involved in ontology development as well as the relations between them (such as subtype, composition and synonym). The methodology proposed in this paper takes its inspiration from ideas taken from earlier research on methodological processes for creating multilingual terminological products and for defining a glossary in a particular domain as well as on domain knowledge organization. The description of our methodology includes the approach followed and the steps carried out, as well as the key issues that arise when the glossary was being created. So far as we are aware, this is the first attempt to normalize the terminology (denominations and definitions) of process and activities in ontology building.

Keywords: ontology development, process, activity, glossary creation.

1. Introduction

During the last two decades, increasing attention has been focused on ontologies and their development. The term *Ontology* (Oxford Dictionary 1987; Munn and Smith 2008), which originally comes from philosophy, is a neologism coined in early modern times from Greek roots: *ontos*, which means "being", and *logos*, which stands for "to talk about or to give an account of something". Accordingly, "ontology" is the field that has being as its subject matter. However, in computational fields, the term ontology has been adopted to describe the parcels of knowledge that can be (computationally) represented in a program (Studer et al. 1998).

Terminology 19:2 (2013), **202–236**. DOI 10.1075/term.19.2.03sua
ISSN 0929–9971 / E-ISSN 1569–9994

Thus, in these fields the term ontology is used with a different meaning to refer to "a formal specification of the conceptualization of a part of reality", because the main motivation behind ontologies is that they allow the sharing and reuse of knowledge bodies in computational form.

The development of ontologies has become an engineering discipline, the so-called *Ontology Engineering* (Gómez-Pérez et al. 2003), which refers to the set of processes and activities that concern the ontology development and the ontology life cycle, the methods and methodologies for building ontologies, and the tool suites and languages that support them. However, it has been noticed that ontology practitioners use different terminology to name the processes and activities involved in ontology building. This is due to the meteoric development of this new field that has favoured the Tower of Babel (Gómez-Pérez and Juristo-Juzgado 1993) on the definitions for ontology engineering processes and activities. Nevertheless, a common terminology is completely necessary for the ontology engineering field since, as established in Bloom's taxonomy (Bloom 1956), the lowest level of knowledge in any field is terminology, and moreover no proper knowledge transmission is possible without terminology. Therefore, no effort should be discarded in order to normalize the terminology in Ontology Engineering. One way to overcome this confusion and the negative effects produced by any misunderstanding in the terminology used when developing an ontology is to create a glossary in the ontology engineering field, by means of consensus among experts, since no similar glossary exists.

Thus, our aim in this research work was to create a glossary that: (a) contains and unifies the terminology in the ontology engineering field, particularly the terminology related to the ontology development process as well as the explicit relations between the different terms; and (b) is built by means of consensus among epistemological communities. The decision of creating this kind of glossary was taken within the NeOn[1] project consortium, due to: (a) the misunderstandings that prevailed during different working meetings held among ontology researchers, technology developers, and users at the beginning of that project; and (b) the need of having unambiguous understanding of the methodological components, i.e. the actions involved in the ontology network development process, to be included in the NeOn Methodology (Suárez-Figueroa 2010; Suárez-Figueroa et al. 2012).

When such a decision was taken, there existed a very preliminary glossary,[2] proposed in 2001, that focused only on the terminology related to the methods and techniques used to define, share, and merge ontologies. However, this small glossary was not dedicated to the ontology development process and it did not include definitions of actions, as needed in the abovementioned project.

All in all, the creation of this glossary was affected by the following factors:

a. First, the very nature of the ontology engineering field. This dynamic field, as has just been mentioned, is in constant evolution, and new situations in ontology development are continuously emerging, such as the creation of ontology networks and the collaborative dimension of their development, and the new initiative of having data available in the Web, known as Linked Data. Thus, new precise terms are needed to name the novel situations.
b. Second, in this new field people from different communities participate, with diverse knowledge background (domain experts, software engineers, knowledge engineers, ontology engineers, etc.). In this situation, reaching consensus on the terms and their meanings is tantamount to winning an obstacle race. To this end, consensus on terminology, — defining consensus as a state of mutual agreement among members of a group where all opinions have been heard and addressed to the satisfaction of the group (Saint and Lawson 1994) — is needed.
c. Third, due to the great variety in backgrounds of the people involved, the resulting terminology was a clear example of cross-fertilization of other related disciplines, mainly Software Engineering and Knowledge Engineering, from which ontology practitioners have borrowed many terms (Devedzić 2002).[3] Thus, even existing terms had to be re-defined to describe more precisely the new situations.
d. Finally, an additional factor, that seemed a considerable hindrance to the creation of the new terminology and was derived from the wide interdisciplinary origin, was the lack of initial agreement on two fundamental "units of understanding" (Temmerman 2000) in ontology development: "process" and "activity". This fact was extremely important because many of the steps followed in ontology building had to be categorised and consequently defined as either a process or an activity.

Bearing in mind all the above mentioned factors and considering that, to the best of our knowledge, there was no explicit methodological process describing the "on-the-fly" creation of a glossary agreed by participants with different backgrounds (domain experts, software engineers, ontology engineers, philosophers, terminologists, etc.), we designed an **iterative and collaborative methodology** to create a glossary in the ontology engineering field, particularly a glossary on the ontology development process.

In our case, we faced the need to develop this glossary within certain time constraints in order to find common grounds for providing unambiguous understanding of the methodological components, i.e. the actions involved in the ontology network development process to be included in the NeOn Methodology. We also aimed at providing global coherence to the documents to be produced throughout the NeOn project. Although corpus-based approaches in terminology engineering have greatly developed as described in Ibekwe SanJuan et al. (2005), creating an 'on-the-fly' glossary as a communication tool within a project in which, up to a point, the corpus was built at the same time presented several conundrums that required some specific approach as explained below.

Therefore, we created our methodology taking as basis some ideas from the research done by Cabré and Tebé (2005) and Nuopponen (2007) for creating multilingual terminological products, and by Velardi et al. (2006) for defining a glossary in the area of interoperability of enterprise applications and software, as will be explained in Section 3. To support our methodology, new technologies, in particular wikis, that could allow people geographically distributed to contribute the process, were used. As a result of applying the iterative and collaborative methodology mentioned above, the **NeOn Glossary of Processes and Activities** was obtained. This glossary includes the definitions of all the processes and activities potentially involved when developing ontology networks.

Currently this glossary is being used by numerous ontology practitioners, although it cannot be thought of as completely extended to the whole ontology community. However, we consider this effort as the first step for solving the lack of a standard glossary in the ontology engineering field, in contrast with the software engineering field that boasts the IEEE Standard Glossary of Software Engineering Terminology (IEEE 610.12:1990).

The structure of this paper is the following. Section 2 presents the factors that affect the creation of a glossary of ontology development terminology. Section 3 describes the terminological and methodological approaches analysed. Section 4 deals with the methodology proposed for the creation of the glossary in the ontology engineering field, while Section 5 shows a summary of the results obtained after applying the methodology. The complete glossary can be found in Annex I. Finally, Section 6 provides some concluding remarks.

2. Factors affecting the creation of the glossary

As already mentioned in Section 1, a set of different factors have affected the creation of the glossary of the terminology on ontology development. Some of these factors result from the indeterminacy inherent to borrowing and merging terms from different fields. But as Antia (2007) suggested glossing Beaugrande's (1997) modelling of cognition and communication, both indeterminacy and determinacy can be seen as two positive poles that favour that knowledge and communication systems can remain dynamic. In this section we explain in detail such factors.

2.1 Ontology Engineering: an evolving field

Ontology Engineering (OE) is in constant evolution, as new situations and needs in ontology development emerge, such as the creation of ontology networks and the collaborative dimension of their development, or the new initiative of having data available in the Web, known as Linked Data. Due to this constant evolution, the process of creating an "on-the-fly" glossary faces several terminological problems:

- Known terms are assigned new meanings that have to be delimited. This is the case, for example, of ontology conceptualization, ontology documentation, and ontology selection.
- Concepts are named differently in different moments (such as ontology modification (Stojanovic 2004) and ontology update (Stojanovic et al. 2002), whose difference is not clear enough). Specialists in the field sometimes prioritize certain features, thus they name the concepts focusing on those features, whereas some other times, though keeping most of the essential features, they highlight new ones, thus changing or creating terms and producing certain terminological instability that accordingly show the ceaseless evolution of the field.
- Some terms are defined in different ways at different stages as the field develops (such as ontology merging (Fernández-López et al. 1997; Kalfoglou and Schorlemmer 2003; Kotis and Vouros 2004)).
- New processes and activities (e.g. ontology customization, ontology modularization, or non-ontological resource reuse) have emerged when large ontologies are built in complex settings, and thus new terms have been created.

This terminological chaos in Ontology Engineering favours the use of unstable and inconsistent terminology to name the processes and activities involved in ontology development. This situation is the result of a lack of standardization in the ontology engineering terminology; that is, practitioners in the field have not yet reached an explicit consensus on the definitions for all the processes and activities involved in ontology engineering. To cope with this evolving feature during the creation of the glossary, we decided to focus only on those terms, related to the ontology development process, that were found in the literature and were mentioned in any of the project meetings as relevant terms for this field.

At that stage, our objective was to devise some steps towards the 'standardization' process by focusing not only on the terminological interest, aimed at establishing provisions for common and repeated use, but also on the sense that a set of technical specifications established in collaboration with and with the approval of the parties concerned in the process can be considered as a standardization process.[4]

2.2 Ontology Engineering: A multidisciplinary field

Ontology Engineering is a discipline that attracts people with different backgrounds of knowledge (domain experts, software engineers, knowledge engineers, ontology engineers, philosophers, linguists, etc.) coming from many varied disciplines. This entire group of people was involved in the creation of the glossary on ontology development. So, agreement on the terms selected and the meaning of those terms, as well as the relations between them (such as subtype, composition, and synonym), had to be decided by consensus of all different parties. In this respect, we came across a situation described by Kerremans et al. (2008) in which human subjects will never share exactly the same thoughts about given referents because of the different ways in which these referents can be experienced. So, the "situatedness" dimension presented by these authors emphasizing the situational context in which understanding takes place was also considered when creating the glossary.

However, the active participation of these people with different skills, who were at that moment geographically dispersed, was not always easy. To deal with this difficulty technologically, we proposed the use of a multi-user collaborative framework that could (a) facilitate and accelerate the feedback and revision process, and (b) reduce the time needed to complete the glossary creation. Since the terminology to be included in the glossary should be collaboratively built and agreed upon by all the participants involved in the process, we decided to use the wiki technology (Leuf and Cunningham 2001). In addition to this collaborative platform, we also proposed to have physical meetings and use mailing lists to achieve consensus at the final stages of the glossary creation.

2.3 Cross-fertilization terminology

Terminology corresponding to ontology development does not only dwell on the specific vocabulary related to ontologies: concepts, relations, attributes, instances, but also achieves cross-fertilization by absorbing terms from other engineering fields that provide new visions of the possible actions to be performed in this new field. By cross-fertilization, we understand the interaction between two or more fields of study that are mutually productive. In this respect, it is clear that new scientific or technological fields absorb terms and concepts from other disciplines by adapting them to the new circumstances. Take, for example, the terminology corresponding to viral medicine and its projection and use in computer terminology, which has benefitted with a great number of metaphorical terms from medicine, or more simply, the very name ontology, as mentioned in the introduction.

So, in the same way as artificial intelligence practitioners borrow and reformulate terminology from epistemology, and the neural networks community receives new terms from neurobiology (Ahmad 2000), the ontology engineering community has borrowed terminology mainly from two other related disciplines: Software Engineering (SE) and Knowledge Engineering (KE). This cross-fertilization has been possible respectively because: (a) ontologies are part of software products, and sometimes ontologies are considered as a kind of software (Suárez-Figueroa 2010); and (b) ontologies are a way to represent domain and general knowledge.

At this stage, it is worth mentioning that SE and KE are complementary disciplines. The major difference between KE, a field that covers the development of knowledge-based systems (KBSs), and SE, a field that deals with the application of engineering to software, is the requirement for knowledge engineers to capture, represent, analyse and exploit knowledge in order to produce a successful knowledge-based system (Kingston 1994). However, both disciplines have similar aims, that is, to turn the process of developing systems (classical and knowledge-based, respectively) from an art into an engineering discipline (Studer et al. 1998). As interrelated disciplines, the convergence between them has slowly gestated at the level of methods, techniques, tools, and procedures (Acuña et al. 1999). Given that we strictly focus on the development process, it is important to note that the description of the software construction process is a subject that has been studied in SE for many years now.

The situation in KE, however, is quite different. The issue of the technical activities to be performed to build a KBS was debated in the 80s, but KE has never taken an interest in fully defining all the activities to be performed when building a KBS (Acuña et al. 1999). In brief, the software process model to be used in both disciplines, SE and KE, seeks to define a series of activities to be performed to produce software (whether conventional or knowledge-based).

Moving now to the ontology engineering field, many similarities and analogies arise if we put this field in the context of the above mentioned disciplines. On the one hand, desirable qualities for ontologies, such as being decomposable, extensible, maintainable, modular, translatable, are also desirable for software components in SE (Devedzić 2002). For this reason, even though practitioners from these two disciplines may use different terminology, at some point they refer to similar processes; or it could also happen that practitioners use the same terminology with analogous meaning. Let us take some examples for both situations. For the first case, we can mention 'modular decomposition' in SE and 'modularization' in OE that refer to the same action, which can be defined as the process of decomposing a product into modules to facilitate its design and development (based on IEEE 610.12:1990). For the second case, the term 'implementation' is the same in both fields and the meaning is also similar, that is, the process of translating a design or a model into hardware components, software components or both (based on IEEE 610.12:1990).

With respect to KE, we have realized that this field has fertilized ontology engineering terminology by providing some terms that have been adopted without variation on their definitions. This means that KE terms and definitions are directly used in OE. Examples of this situation are terms such as 'conceptualization', 'formalization', and those related to the evaluation process ('evaluation', 'verification', 'validation' and 'assessment') (Gómez-Pérez et al. 2003; Gómez-Pérez et al. 1995) that have the same name designations and definitions both in KE and in OE.

Thus, Ontology Engineering is creating its own terminology by absorbing terms from more permanent fields such as SE and KE, among others.

2.4 Terminological confusion: Process or activity?

As already mentioned in Section 1, two key terms in engineering development are 'process' and 'activity'. However, these terms are not clearly delimited within the ontology engineering community when deciding whether a particular action (to be done in ontology development) is a process or an activity. This confusion affects the organizational map of the steps performed in ontology building. So, these terms should be represented in terminological resources by considering the situational contexts over and above the lexical ones, as suggested by Kerremans et al. (2008).

Indeed, this lack of clarity does not only occur in the ontology engineering community but also in related fields such as SE. In other more general situations, this distinction has also become important, as Nuopponen (2007) claims. She

Table 1. Summary of process and activity definitions

Definitions of Process	Definitions of Activity
General Linguistic Resources	
a series of actions or operations conducing to an end (Merriam-Webster[5])	an organizational unit for performing a specific function (Merriam-Webster[5])
a particular course of action intended to achieve a result (WordNet On Line[6]). It is important to note that process is hypernym of activity	any specific behaviour (WordNet On Line[6]). It is worth mentioning that activity is hypernym of action
Specialized Resources	
a sequence of steps performed for a given purpose. A process is composed of activities (IEEE 610.12:1990)	a defined body of work that is to be performed, including its required input and output information (IEEE 1074-1997)
a function that must be performed in the software life cycle. A process is composed of activities (IEEE 1074.1-1995)	a constituent task of a process (IEEE 1074.1-1995)

argues that although some terminologists could state that in terminology the process description may be uninteresting, for various other information specialists (e.g. technical writers, information system designers) detailed process mapping is necessary. In this latter group we can also include ontology engineers as they really require differentiating process from activity at the conceptual and ontological level.

Although these two terms (process and activity) can be said to belong to the general vocabulary, we have realized that they have different meanings both in specialized resources (such as IEEE software glossaries) and in general linguistic resources (such as the Merriam-Webster online and WordNet online).

Let us analyze the nuances provided by both general linguistic resources and specialized resources in their definitions of process and activity. Table 1 summarizes the most usual definitions for these two terms.

The definitions presented in Table 1 are too general and they do not focus on engineering development issues, or if they do, they seem to be incomplete for ontology engineering experts. For these reasons and with the aim of better delimiting the actions corresponding to process and activity, we created a new definition (thus, a semantic neologism[7]) for these two key words. We also considered development issues and conceptual relations such as the order in the actions performed, and the information required and obtained after the actions had been carried out:

- ***Activity*** is an action to be performed, including its required input and output information.
- ***Process*** is a group or set of ordered activities.

3. Approaches to build a glossary

It is well known that in terminology work there are two different approaches, namely the traditional onomasiological approach proposed by Wüster (1968), and the semasiological one, widely accepted by translation-oriented approaches (Sager 1993) and corpus-based approaches (Condamines 2005). These two approaches, the onomasiological and semasiological one, have been traditionally linked to the researchers' background and the different purposes they have when facing terminology issues.

On the one hand, the onomasiological approach puts the emphasis on the systematic ordering of concepts and their relations as well as on the univocity of terms. Wüster and his followers (Felber 1979) intended to eliminate the ambiguity from technical domains. Thus, they aimed to achieve a better communication among technical users and to convince them of the benefits of standardized terminology. However, this pragmatic approach was later on complemented with an integrated, interdisciplinary view of terminology, as Budin (2010) explains. He proposes an integrating approach that combines the socio-terminology dimension with a socio-computational dimension based on corpus. Nevertheless, this author warns about relying totally on the results extracted from corpora, in that data may not give relevant information about what domain experts really think about the terms concerned, unless the text explicitly reflects the terms in discussion and there are other indicators that explain the socio-cognitive factors relevant to them.

In the last fifteen years, with the ontologies becoming mainstream technologies, the onomasiological approach has gained new vigour, as these knowledge representations have to be understood both by computers that require concepts to be unambiguous as well as by humans. In that sense, both terminologies and ontologies aim to reach a common understanding of a domain, by sharing knowledge, although there are also differences both in formalization, purpose, and explicitation, to mention just a few. In the computational field, one of the challenges at this stage is to represent all the terminological as well as the multilingual and linguistic information, which can be in other lexicographic resources, and/or in ontologies (Huang et al. 2010, Montiel-Ponsoda et al. 2011).

On the other hand, the semasiological approach generally germinated in linguistic environments, both among translators and sociolinguists. Experts such as linguists, terminologists and translators took texts as point of departure to understand terms, extract them, capture lexical relations and build language resources, with different purposes. In this sense, corpus-based approaches have proved essential not only for the linguistic analysis of specialized languages (Bowker and Pearson 2002), but also for terminology extraction (Condamines 1995, 2005, Temmerman and Kerremans 2003, L'Homme 2004), definitional patterns extraction (Sierra et al. 2008), taxonomy extraction (Nazar et al. 2012), identifying conceptual relations (Condamines and Rebeyrolle 2001), ontology building (Jacquemin 2001; Malaisé et al. 2005), enriching terminological databases (Schumann 2012), among many others.

However, in the case depicted in this paper, the starting point is the specialized community of ontology engineering practitioners that, for functional and pragmatic reasons, decide to agree on certain terminological aspects, i.e. concept extension, conceptual relations and terms to be used at a certain time and for specific purposes. So, in this sense, as Rey (2005: 324) puts it the needs of a language for naming, designation and lexical materialisation are a matter of onomasiology. That is to say, ontology practitioners face the situation of naming a new concept (a particular action in ontology development), and agree to do it by convention. Moreover, the focus of this glossary is specifically on processes and activities needed when building ontology networks, so the volume of the terms was not expected to be very high.

As mentioned in Sections 2.2 and 2.3, the ontology engineering field is sprinkled with experts with different backgrounds, thus some terms are extracted from contexts in different content fields, such as SE and KE and agreed in a sort of intra-language translation. Moreover, new terms are sometimes a clear product of metaphorical changes, in which the resemblance of old processes is updated and adapted to the new field. In this case, ontology practitioners depart from a name (a lexical unit) and search for its meaning and the most appropriate definition. But also new terms can show a trend line of climbing up the mainstream research topics. In fact, this is quite common in dynamic, cutting edge fields.

In all these circumstances, some existing methodological work can be taken as the inspiration for the creation of the ontology engineering glossary. It includes Cabré and Tebé (2005) on creating multilingual terminological products, Velardi et al. (2006) on defining a glossary in the area of interoperability of enterprise applications and software, and other more specific experiments on enhancing a typology of process-related relations (e.g. Nuopponen (2007)). Moreover, a translation-oriented method that combines both approaches (onomasiological and semasiological) is the Termontography method (Temmerman and Kerremans 2003), as explained below.

In the former work, Cabré and Tebé (2005) present a proposal for a methodology for creating multilingual terminological work. One of the most important aspects in this proposal is the participation of terminologists as well as translators and domain experts with the aim of achieving corpus-based linguistic enrichment in a terminological resource. The methodology proposed, which is based on a semasiological approach, follows four phases:

- Phase I (work preparation), which includes: (a) analysis of the starting nomenclature; (b) establishment of the work team; (c) establishment of the textual and lexicographic corpus, and (d) establishment of work criteria.
- Phase II (data preparation), whose activities are: (a) conversion of the original format; (b) importing data into the selected data base, and (c) data base management.
- Phase III (terminological work), in which the following activities are proposed: (a) search for equivalents; (b) checking those equivalents; and (c) updating work criteria.
- Phase IV (work review), which includes the following activities: (a) methodological revision of the work; (b) content revision by experts; and (c) data export to the required format.

The second analysis (Velardi et al. 2006) proposed a specific methodology composed also by four stages, which can be summarized as follows:

– First stage to define the purpose of the glossary.
– Second stage to build an initial list of terms and definitions using a semiautomatic glossary acquisition.
– Third stage to set up the collaborative glossary online module to support the sharing and extension of the glossary.
– Fourth stage, to complete the glossary by means of manual inputs and reviews, that is, the extension of the glossary.

At another level, Nuopponen's work (2007) is very interesting, although it cannot be taken as a methodology to build a glossary. Her detailed description of a specific process, the Japanese tea ceremony, provides a thorough conceptual map of all the concept relations participating in that process (temporal, agent, object, instrumental, teleological, resultive and locational).

Finally, it is worth mentioning the Termontography method (Temmerman and Kerremans 2003), which is meant to capture and represent knowledge acquired from texts in a platform to be used by translators (Kerremans et al. 2008). This method follows a middle-out approach and combines theories and methods for multilingual terminological analysis of the sociocognitive approach with methods and guidelines for ontological analysis. Firstly, an initial framework of categories and inter-categorial relationships is developed top-down in close collaboration with field specialists of the domain. Such a framework serves as a template for the manual and semi-automatic extraction of knowledge from a corpus. This framework is evolved and enriched in a bottom-up fashion as the knowledge elicited via textual material is then confronted with the categorical frame. The result of this method is a termontological database, which can be exported to an ontology and a terminological dictionary.

4. Methodology to create a glossary in Ontology Engineering

Despite the discrepancies found in the literature, ontology experts have reached a certain implicit consensus on which processes and activities should be undertaken in ontology development and on their possible definitions. With the purpose of unifying and making this terminology explicit, we decided to build a glossary that includes key terms involved in ontology developments. So, it was also necessary to organize the conceptual map of processes and activities to better delimit the relationships among the different actions carried out in ontology development. This decision was taken within the NeOn project consortium, due to the misunderstandings that prevailed during the meetings held among ontology researchers, technology developers, linguists, and users at the beginning of the project.

As previously mentioned, this initiative of creating a precise glossary of ontology engineering terminology is the first attempt to normalize the terminology (denominations and definitions) of processes and activities in ontology building that also reflects the relationships among all of them.

Thus, we have designed an **iterative and collaborative methodology** to create a glossary in the ontology engineering field, particularly a glossary on the ontology development process. During the creation of the glossary we have followed a middle-out approach, inspired on the approaches and methodologies summarized in Section 3. In such a middle-out approach ontology practitioners have to think firstly of the names for particular actions in ontology development; then, they decide on explicit definitions for those names and for the sequence of relations among the concepts described. Secondly, they take terms they know and introduce them in the new field, after delimiting the meaning intension and deciding on the most suitable definitions. In addition, our methodology is also grounded in a consensus reaching process, which is defined as a dynamic and iterative process composed of several rounds, where the experts express and discuss their opinions in order to reach the maximum agreement about a set of alternatives before taking a decision (Herrera-Viedma et al. 2005).

This section sketches the overall methodological process followed to create the glossary for the processes and activities required when developing ontology networks by means of consensus. A high level view of this process can be summarized as follows:

1. To select those actions existing in the ontology engineering field for which there was a formal or a non-formal definition (Trimble 1985; Pearson 1998) in documents. In these cases, the goal in the glossary creation process was to review and improve (if needed) such existing definitions.
2. To compile all the new actions for which there were no definition at all. In this case, the goal was to propose a definition and an accepted term.
3. To classify all the actions included in the glossary as processes and activities. In this case, the goal is to differentiate process from activity at the conceptual and ontological level.

The general process for identifying and defining the processes and activities involved in the development of ontology networks is composed of the following phases, which are explained in detail below: (1) Phase I. Work preparation, (2) Phase II. Initial glossary preparation, (3) Phase III. Terminological work, (4) Phase IV. Glossary revision, and (5) Phase V. Glossary publication. We have created this general process based on the phases proposed by Cabré and Tebé (2005) and we have included some specific tasks keeping in mind the steps proposed by Velardi et al. (2006). We have also considered domain knowledge organization, as shown in Figure 2, and it is also suggested by the Termontography method (Temmerman and Kerremans 2003). A schema of our iterative and collaborative process is

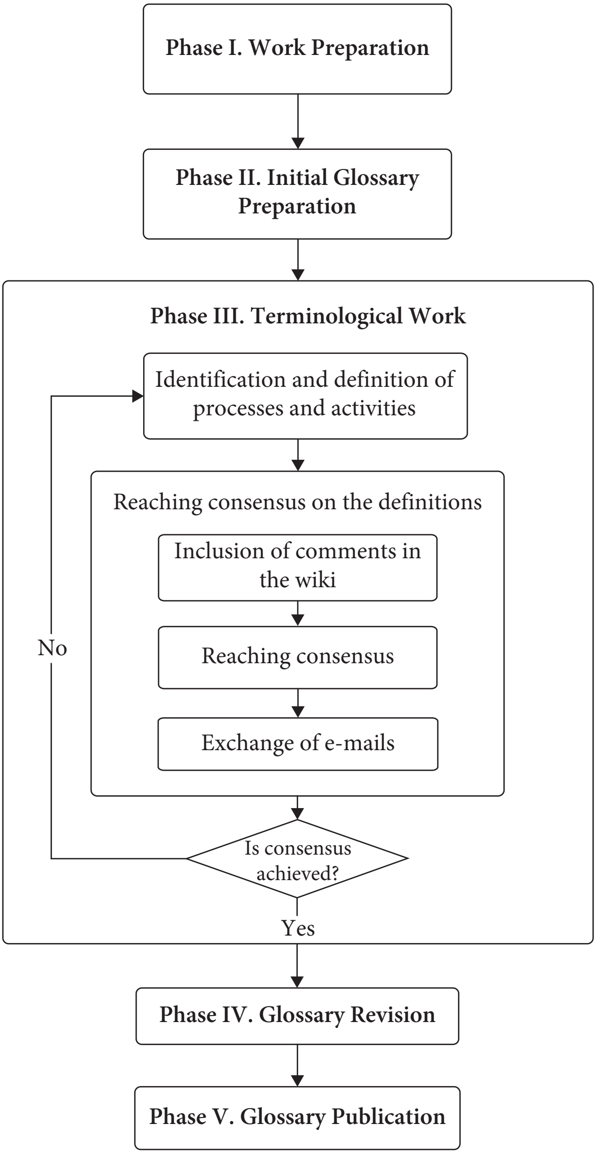

Figure 1. General process for creating the glossary

shown in Figure 1. It is worth mentioning that the final glossary, which is included in Annex I, was obtained after the third round of the global process.

4.1 Phase I. Work preparation

In this phase the following activities were performed:

- Determining the purpose of the glossary. The main objective of the ontology engineering glossary was twofold: (a) to consolidate and make explicit the terminology on the ontology development processes and activities used by people involved in the NeOn project, and (b) to avoid ambiguity when designing ontology networks in this project. The most important benefit of creating this kind of glossary is the conceptual representation of the terminology used regarding ontology building actions. Thus, the final glossary represents a shared view of relevant concepts in ontology development together with their definitions and terms. In this respect, we conform to ISO 1087-1:2000, in which a **glossary** is defined as *a terminological dictionary that contains designations and definitions from one or more specific subject fields*.
- Setting deadlines. People involved in the glossary creation had to agree on a limit period of time to build the glossary. In this case, a targeted time period of one year was agreed for both reaching a consensus and creating the glossary.
- Establishing the game rules. It is important to understand the main rules involved in the creation of the glossary. The rules are related to the consensus-based approach to be followed during the glossary creation. Before starting, it was necessary to delimit some concepts; thus, the meanings of 'consensus' and 'consensus reaching process' were explained to the people concerned.
  - Consensus is defined as a state of mutual agreement among members of a group where all opinions have been heard and addressed to the satisfaction of the group (Saint and Lawson 1994).
  - Consensus reaching process is defined as a dynamic and iterative process composed of several rounds, where the experts express and discuss their opinions in order to reach the maximum agreement about a set of alternatives before taking a decision (Herrera-Viedma et al. 2005).
- Establishing work criteria. During the creation of the glossary the following linguistic and terminological criteria were followed:
  - Basic principles for defining a term, mainly based on the traditional aristotelic genus-species definition, in which a definition is a descriptive representation of a concept which serves to differentiate it from related concepts. (Landau 1984) According to this, the selection of an appropriate superordinate is crucial for the intelligibility of the defining statement. In

  Pearson's words (Pearson 1998, 86), the superordinate or closest generic concept should preferably be one step up in the hierarchy from the term being defined. Moreover, the same superordinate should be used for all terms that belong to the same class. This has been relevant for the conceptual organization of processes and activities.
  - In this regard, for practical reasons, we decided to use a general structure for each definition that consists of (a) an entry term, the "definiendum" (designating the concept being defined) followed by (b) a full stop and (c) the definition, the "definiens". Such a definition should be intensional and begin with the structure 'it refers to' followed by the broader generic (superordinate) concept associated with the concept being defined, together with delimiters indicating the characteristics that delimit the concept being defined from coordinate concepts.
  - Other basic principles on definitions (Cabré 1993), widely accepted in lexicographic practice, were also taken into account.
    - Based on content: avoid circularity, define every word in a definition, and make sure that every word's definition says what the word means.
    - Based on style: (a) conciseness, i.e. every definition should say the most in the least number of words; (b) clarity, or avoidance of ambiguity, i.e. words should be used unambiguously; (c) appropriateness, i.e. the definition should be appropriate to the target reader; and (d) priority of essential traits, i.e. a definition should highlight the essential features of meaning.
- Setting up the work team. A varied number of skilled people (known as the Glossary team), geographically dispersed, collaborated on the consensus reaching process. This team had a well-balanced and representative participation of the different people involved in the process: ontology engineers, ontology editors, linguists and users within the NeOn project. The team was composed of 25 people belonging to 9 institutions. Its different members had the following specific roles (or functions): coordinator and contributor.
  - Coordinator: the first author of this paper, who belongs to one of the institutions participating in the glossary creation, led the whole process of building the glossary and acted also as contributor.
  - Contributor: the other participants introduced process and activity definitions and commented on them.
- Selecting the working platform. Coming from so many institutions one of the vital decisions was to decide the platform and technological support on which the glossary would be developed. The following steps were taken:
  - To select the technological support for creating the glossary: we decided to use a multi-user collaborative framework, which facilitates and accelerates

the feedback and revision process and reduces the time needed to complete this process. Specifically, we decided to use the wiki technology (Leuf and Cunningham 2001) because it supports a higher level of consensus building by community members; thus, if a user disagrees with a statement, he can very easily modify it, delete it, comment it, etc. (Viégas et al. 2004). Currently, there are hundreds of wiki services available (West and West 2008) such as Wikimatrix, Wikiversity, Google Docs, MediaWiki, and TWiki. We decided to use MediaWiki, for compatibility reasons, because the project wiki had also been established using this free software. Face-to-face meetings and mailing lists were also used to reach consensus at the final stages.

– To create a wiki page dedicated to the consensus reaching process: the glossary coordinator created a non public space within the NeOn project wiki, where all the people involved in the glossary creation could discuss the ontology engineering terminology as well as express and exchange different opinions in order to reach a final agreement.
– To create a template for gathering information about processes and activities. In order to facilitate participation and save time to the rest of the participants, the glossary coordinator created, in the dedicated wiki, a template for gathering information about the processes and activities for

Table 2. Template for processes and activities in the NeOn Glossary

Template Slot	Description
Process or Activity Name	Name of the process or activity
Definition	One or several natural language (NL) definitions (with the corresponding references, if needed)
Type of Process/ Activity according to IEEE	In the Software Engineering field, activities are grouped administratively into five main activity groups (IEEE 1074-2006). Based on that, the following groups are proposed: a. Ontology Management b. Ontology Pre-Development c. Ontology Development d. Ontology Post-Development e. Ontology Support Activity The type for a particular process or activity should be unique
Input	A list of the information required to be input of the process or activity
Output	A list of the information that is required to be output of the process or activity
References	References for the NL definitions (if needed)
Comments	Other comments about the process or activity

building ontologies. The template, shown in Table 2, gathers not only possible process and activity definitions, but also other general and semantic information useful for the glossary creation (e.g. the input and output of the process or activity and certain relations among them).

4.2 Phase II. Initial glossary preparation

During this phase, the work team had to create an initial list of actions that had been already defined (formally or informally) in any document of the ontology engineering field. The glossary coordinator manually extracted an initial corpus of ontology development actions (denominations and definitions) from the following documents:

- Ontology development methodologies (METHONTOLOGY (Gómez-Pérez et al. 2003), On-To-Knowledge (Staab et al. 2001), DILIGENT (Pinto et al. 2004)); and
- NeOn project use cases (Iglesias et al. 2006; Gómez-Pérez et al. 2006).

This initial corpus was also extended with inputs from different ontology engineering papers and ontology experts. This initial glossary contained 39 ontology engineering actions with their definitions and relations. This list was made available in the dedicated wiki following the template presented in Table 2, and thus, it was introduced as the seed set for debate.

The general idea is that any action or work to be done in ontology development that needs an input and generates an output is considered to be included in the glossary. This restriction is relevant for the final glossary. For such an action or work, a term (lexical unit) and its definition (description unit) should be created. After analyzing the relations between the different actions, the decision of whether the action is an activity or a process should be taken. At this stage, the conceptual analysis was also carried out.

4.3 Phase III. Terminological work

Within this phase, the work team performed two main tasks: (1) to identify and define collaboratively processes and activities in the wiki, according to the initial list, and (2) to reach a consensus on process and activity definitions.

With respect to the first aforementioned task, the work team was given total freedom to incorporate more processes and activities and/or definitions in the initial glossary, and also to include more general information about the process or activity (such as input and output or the classification following the groups based on IEEE 1074-2006). Regarding the task of reaching consensus on definitions, the team used the wiki, held physical ad-hoc meetings, and interchanged e-mails. The general approach during this task was as follows:

a. Inclusion of comments in the wiki. Each participant introduced comments for each process and activity available in the wiki. These comments could refer to, for example, the particular definition a participant would prefer if there were more than one. Participants could also add a proposal for a new definition for each process and activity in the wiki.
b. Reaching consensus. Three rounds to review and agree on the definitions of processes and activities were held. Two of them were based on physical ad-hoc meetings. The third and last round was based on mailing lists. During these rounds, the team took into account some informal rules for accepting or rejecting a particular process and activity definition; these rules are shown in Table 3.
c. Exchange of e-mails. Finally, several e-mails were sent to the NeOn project consortium to encourage other people, not participating in the glossary creation, to put forward their ideas about ontology development terminology.

Besides this internal agreement, the work team decided to obtain support from external users with the aim of achieving the widest consensus. The

Table 3. Informal rules used in the consensus reaching process

IF	THEN
The team's comments were generally positive and no major objections were raised	The definition was considered as final
The team's comments were generally positive, but someone had a major objection to the definition	The definition was modified until no major objection was encountered
The team's comments were generally negative	The definition was ruled out
The team's comments were mixed	There were three possibilities: – discussions continued until positive or negative results were achieved; – discussions were postponed until the next meeting; – the issue was postponed until more information was available in the wiki
Discussions seemed to be going on forever without any possibility of reaching an agreement	The team could – decide to drop the definition, or the process or activity; – send for approval by voting the definition. The selected voting procedure was based on absolute majority

comments received were mainly about differentiating or making equal some process and activity definitions, including missing aspects in some activity definitions, and deleting some processes and activities. All the comments received were considered during the second and third round of the process.

In sum, fifty one processes and activities were identified and defined in this phase. A summary of the results is shown in Table 4.

4.4 Phase IV. Revision

In this phase, two types of revisions were performed over the glossary already obtained. On the one hand, a content revision with respect to completeness; and, on the other hand, a revision regarding terminological principles.

As for the first one, the glossary coordinator realized that some processes and activities and their corresponding definitions (e.g. Non-Ontological Resource Reengineering and Non-Ontological Resource Reuse) were not included in the glossary. Thus, after analysing the different scenarios for building ontology networks (Suárez-Figueroa 2010; Suárez-Figueroa et al. 2012), such terms and their definitions were included in the glossary. Specifically 8 new terms and definitions were included in the glossary.

As for the latter revision, the definitions as well as their corresponding comments included in the updated glossary were checked and reviewed from a linguistic point of view by an expert terminologist. This expert terminologist took into account the linguistic and terminological criteria established in the work preparation phase.

4.5 Phase V. Glossary publication

Once the final glossary was accepted by all participants, it was published in a public website.[8]

5. The NeOn Glossary of Processes and Activities

As a result of the process explained in Section 4, the glossary with the main processes and activities to be performed in the ontology network development process has been obtained. This glossary collects the 59 main processes and activities that are involved in the specific field of ontology engineering and provides definitions and explanations agreed by consensus. The glossary is ordered alphabetically by type of term (process or activity) and each term is related to other terms, if necessary. The vocabulary included in the glossary is monolingual (English).

Table 4. Summary of the results obtained in Phase III

		Number of processes and activities identified	**Number of consensual definitions obtained and relations established**				**Number of processes and activities deleted**		
1st Round	*Type: Physical ad-hoc meeting*	46	25	O. Modularization O. Module O. Partitioning O. Assessment O. Evaluation O. Validation O. Verification O. Diagnosis O. Repair	Knowledge Acquisition O. Elicitation O. Learning O. Population O. Conceptualization O. Formalization O. Implementation O. Specification	O. Pruning O. Enrichment O. Extension O. Specialization O. Documentation O. Summarization O. Update O. Upgrade	0		
				Subtype relations between processes and activities					
2nd Round	*Type: Physical ad-hoc meeting*	61	4	O. Aligning O. Configuration Management	O. Localization O. Mapping		5	O. Articulation O. Combining O. Coordination	O. Mediation O. Reconciliation
				Synonym relations between processes and activities					
3rd Round	*Type: Mailing list*	56	22	O. Environment Study O. Feasibility Study O. Control O. Quality Assurance Scheduling O. Annotation O. Customization O. Reuse	O. Comparison O. Search O. Selection O. Evolution O. Modification O. Versioning O. Reengineering	O. Reverse Engineering O. Restructuring O. Forward Engineering O. Integration O. Matching O. Merging O. Translating	5	O. Adaptation O. Personalization O. Generation	O. Transforming O. Valuation

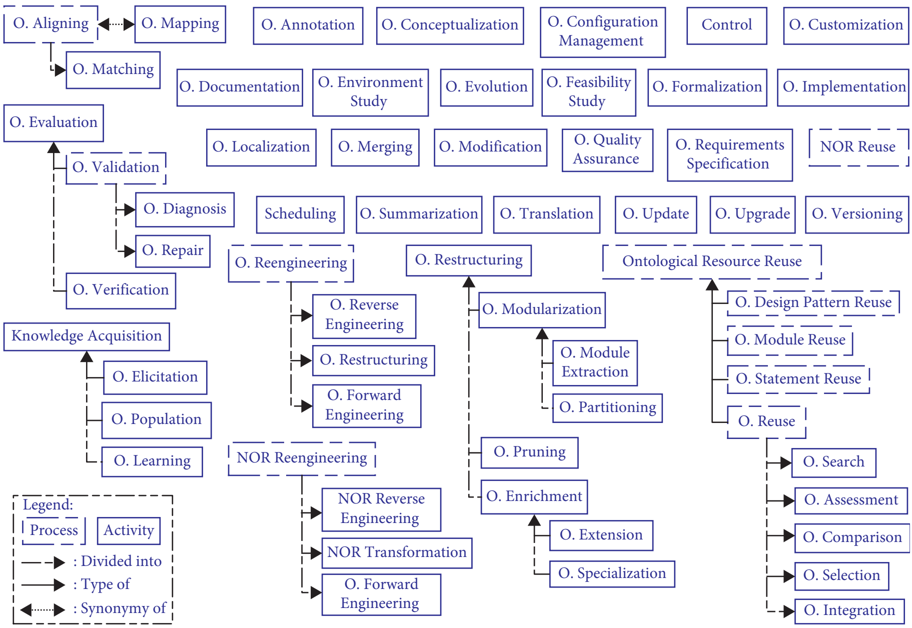

Figure 2. NeOn Glossary of Processes and Activities

Figure 2 shows the whole NeOn Glossary of Processes and Activities, where the different types of arrows have the following meanings: (1) arrows with dashed lines mean "a process is divided into activities"`; (2) arrows with solid lines mean "type of"; and (3) arrows with dotted lines mean "synonymy".

It is worth mentioning that within the NeOn Glossary, the following three situations can be distinguished:

a. Some of the processes and activities have maintained their definition taken from the literature. This is the case, for example, of "ontology configuration management", "ontology evolution", and "ontology formalization". This situation is directly related to the cross-fertilization factor presented in Section 2.3.
b. For most of the processes and activities, the definition has been changed and adapted according to past definitions found in the literature and based on participants' comments and discussions. We can mention, for example, "ontology conceptualization", "ontology documentation", "ontology matching", and "ontology selection". This situation is a clear result of the evolutive nature of the ontology engineering field as mentioned in Section 2.1.
c. New processes and activities and their definitions have been created during the glossary creation, either because the process or activity did not previously exist in the literature, or because they had not been defined, as it happens in the following cases: "ontology annotation", "ontology comparison", "ontology

customization", "ontology diagnosis", "ontology elicitation", "ontology enrichment", "ontology extension", "ontology learning", "ontology localization", "ontology module extraction", "ontology partitioning", "non-ontological resource reengineering", "ontology repair", "non-ontological resource reuse", "ontology search", and "ontology upgrade". This situation shows the dynamicity and evolving features of the ontology engineering field as well as its multidisciplinary dimension as we presented in Sections 2.1 and 2.2, respectively.

In addition, the NeOn Glossary includes the following relations among the defined terms: (a) subtype (a particular general action can be divided hierarchically speaking into subactions, which are more specific); (b) composition (a process is composed by a set of ordered activities); and (c) synonymy (a particular designation for an action can be equivalent linguistically speaking to another designation). The terms and their definitions included in the NeOn Glossary are presented in Annex I in alphabetical order classified into process or activity.

6. Conclusion

Ontology Engineering is a discipline in constant evolution. New ontology building approaches appear (like the creation of ontology networks and the reuse of any knowledge resource available) and new applications based on semantic issues emerge (such as the Linked Data initiative). As a consequence of this evolution, the terminology used by ontology practitioners is sometimes unclear and fuzzy and can create misunderstandings. Terminological clarifications should be made explicit so as to allow a proper knowledge transmission in this engineering field. One way of facing this challenge is to normalize the terminology used in Ontology Engineering. The first attempt to this terminological normalization was the creation of a glossary on the main processes and activities involved in ontology development. This glossary should: (a) contain and unify the terminology in the ontology engineering field, particularly the terminology related to the ontology development process, as well as the explicit relations between the different terms; and (b) be built by means of general consensus among epistemological communities.

We have analysed in depth the different factors that could affect the creation of the glossary. Our study reveals four key factors to be taken into account. One is the dynamic, evolutive nature of the ontology engineering field, as already mentioned. A second factor refers to the multidisciplinary dimension of the ontology engineering field, in which people from different communities participate, with diverse knowledge background (domain experts, linguistics, software engineers, knowledge engineers, ontology engineers, etc.). This second factor implies the third one, the cross-fertilization of terminology coming from other related disciplines, such as Software Engineering and Knowledge Engineering. Finally, an additional factor is the lack of initial agreement on two fundamental terms in ontology development: *process* and *activity*.

Bearing in mind all these factors, we have designed an **iterative and collaborative methodology** to create a glossary in the ontology engineering field, particularly a glossary on the terminology involved in ontology development. To devise this methodology, we have drawn on two related works: one for creating multilingual terminological products (Cabré and Tebé 2005) from which we have taken advantage of its general steps, and the other for defining a glossary in the area of interoperability of enterprise applications and software (Velardi et al. 2006) from which we have reused and adapted some specific tasks. We have also considered domain knowledge organization, as suggested by the Termontography method (Temmerman and Kerremans 2003).

In addition, it is worth mentioning that our iterative and collaborative methodology is based on a middle-out approach in which ontology practitioners think firstly on the names for particular actions in ontology development, then decide on explicit definitions for those names; and secondly, they take terms they know and introduce them in the new field, after delimiting the meaning intension and deciding on the most suitable definitions. In addition, this methodology is also grounded in a consensus reaching process.

To support our methodology, new technologies, in particular wikis, to allow people geographically distributed to contribute the glossary creation, have been used. This kind of multi-user collaborative framework facilitates and accelerates the feedback and revision processes and reduces the time needed to complete the whole process.

As a result of applying our iterative and collaborative methodology, the **NeOn Glossary of Processes and Activities** has been obtained. This glossary identifies and defines all the current processes and activities potentially involved when developing ontology networks. In addition, the glossary also includes the relations between the different concepts and terms, which will undoubtedly benefit the global understanding of the terminology in this field. It is important to mention that the glossary is a key asset in the NeOn Methodology framework (Suárez-Figueroa 2010; Suárez-Figueroa et al. 2012). Currently this glossary is being used by numerous ontology practitioners, although it cannot be thought of as completely extended to the whole ontology community. However, we consider this effort as the first step for solving the lack of a standard glossary in the ontology engineering field, in contrast with the software engineering field that boasts the IEEE Standard Glossary of Software Engineering Terminology (IEEE 610.12-1990).

In this paper we describe the methodology and provide an explicit solution to the problem of how to create a glossary (a) reusing some existing terms, (b) involving different communities, and (c) applying mechanisms and tools to achieve consensus.

Our methodology presents both lights and shadows. On the one hand, the fact that (a) the glossary creation has involved people with different backgrounds coming from many varied disciplines, and (b) the ontology engineering terminology has influences from closed engineering fields (such as software and knowledge engineering) is a means of enriching the terminology included in our glossary. In addition, the use of collaborative technology to support the methodology has benefitted the implication of people geographically disperse, thus avoiding the need of having only physical meetings for the glossary creation. The iterative nature of our methodology has allowed the gradual improvement and revision of the terminology being defined in the glossary. However, there are also some shadows in the process. First, the multidisciplinary dimension of the ontology engineering field has made it difficult at some points to put all the different viewpoints together in some specific terminological definitions. With respect to the technological support, we realized that, especially at the beginning of the process, some actors involved in the glossary creation but not accustomed to using wiki tools had minor problems on how to contribute the process. Finally, the iterative dimension of our methodology could lead to an unfinished glossary creation process. To avoid this problem we had to determine a fixed period in which the proposed methodology would be applied.

On the other hand, one important issue to be taken into account is the evolutive nature of the ontology engineering field. The current NeOn Glossary includes definitions for those processes and activities that were used in the field in the period in which the glossary was created. However, new neologisms recently emerging (such as a new sense for ontology summarization, and new actions in ontology development like ontology winnowing and ontology trimming, or new actions related to the Linked Data initiative) should be included in the revised version of the glossary. Our long term goal is to have a more complete and commonly-agreed glossary, which could become the terminological reference in the ontology engineering field. In this regard, it is important to note that the glossary is mainly addressed to the ontology engineering community, that is, terminology used in definitions are closer to ontology engineers than to, for example, software developers and final users. For future work, it could be also interesting to analyze how terms defined in the glossary have evolved both formally and semantically and verify to what extent they are used in the literature, and whether concepts have traversed other fields and their own limits. As mentioned in the introduction, these cutting-edge fields are constantly innovating in technology, and, consequently, in terminology too. However, this analysis and the subsequent results can be object of a future study.

Acknowledgments

This work has been partially supported by the NeOn project (IST-2005-027595) and the BabelData project (TIN2010-17550). We would like to thank the two anonymous reviewers for their valuable suggestions and comments. We are also grateful to our NeOn partners for their revisions. Our thanks go also to Rosario Plaza and Elena Montiel-Ponsoda for their comments. However, all remaining errors are ours alone.

Notes

1. http://www.neon-project.org/

2. http://www.jfsowa.com/ontology/gloss.htm

3. This cross-fertilization process has been especially fruitful in different areas of computer science as some authors have pointed out. We refer more interested readers to Aguado de Cea (1994/1996) and Meyer et al. (1997).

4. For a deep analysis on 'standardization' see De Vries (1997).

5. http://www.merriam-webster.com/ (accessed on 23rd October 2012).

6. http://wordnetweb.princeton.edu/perl/webwn (accessed on 23rd October 2012).

7. A neologism can be a new word (simple or complex) with a new meaning and an existing word (simple or complex) used in a new field with a new meaning.

8. http://mayor2.dia.fi.upm.es/oeg-upm/files/pdf/NeOnGlossary.pdf

9. *Non-ontological resource* is defined as a knowledge resource whose semantics has not yet been formalized by means of an ontology. Elements in this set are glossaries, dictionaries, lexicons, classification schemes and taxonomies, and thesauri.

10. *Ontological resource* is defined as a set of elements extracted from a set of available ontologies in order to solve a need. Elements from this set can be ontologies, ontology modules, ontology statements or ontology design patterns.

References

Acuña, S. T., M. López, N. Juristo, and A. M. Moreno. 1999. "A Process Model Applicable to Software Engineering and Knowledge Engineering." *International Journal of Software Engineering and Knowledge Engineering IJSEKE* 9 (5): 663–687.

Aguado de Cea, G. 1994/1996. *Diccionario Comentado de Terminología Informática.* Madrid: Paraninfo.

Antia, B. E. (ed.). 2007. *Indeterminacy in Terminology and LSP*. Amsterdam: John Benjamins Publishing.

Ahmad, K. 2000. "Neologisms, Nonces and Word Formation." In *The 9th EURALEX Int. Congress*, ed. by U. Heid, S. Evert, E. Lehmann, and C. Rohrer, Vol. II, 711–730. Munich: Universitat Stuttgart.

de Beaugrande, R. 1997. *New Foundations for a Science of Text and Discourse Cognition, Communication and the Freedom of Access to Knowledge and Society*. New York: Ablex Pub. Corp.

Bloom, B. (ed.). 1956. *Taxonomy of Educational Objectives. Handbook I: Cognitive Domain*. New York: Addison Wesley Publishing Company.

Bowker, L., and J. Pearson 2002. *Working with Specialized Language. A Practical Guide to Using Corpora.* London/New York: Routledge.

Budin, G. 2010. "Socioterminology and Computational Terminology: Toward an Integrated Corpus-based Research Approach." In *Discourse, Politics, Identity*, ed. by R. DeCilia, H. Gruber, M. Krzyzanowski, F. Menz, 21–31. Tübingen: Staufenburg Verlag.

Cabré. M. T. 1993. *La Terminología. Teoría, Metodología, Aplicaciones*. Empúries (Barcelona): Antártida.

Cabré, M. T., and C. Tebé. 2005. "El Trabajo Terminológico Multilingüe de Enriquecimiento Lingüístico: una Propuesta Metodológica." *Revista española de lingüística aplicada* Volumen Monográfico: 19–41. Gil, L., and Aguado de Cea, G. (Guest editors).

Condamines, A. 1995. "Terminology: New Needs, New Perspectives." *Terminology* 2 (2): 219–238.

Condamines, A., and J. Rebeyrolle. 2001. "Searching for and Identifying Conceptual Relationships via a Corpus-based Approach to a Terminological Knowledge Base (CTKB): Method and Results." In *Recent Advances in Computational Terminology*, ed. by D. Bourigault, C. Jacquemin, and M.-C. L'Homme, 127–148. Amsterdam: John Benjamins Publishing.

Condamines, A. 2005. "Linguistique de corpus et terminologie." *Langages* 39 (157): 36–47.

Devedzić, V. 2002. "Understanding Ontological Engineering." *Communications of the ACM: Supporting Community and Building Social Capital* 45 (4): 136–144.

Felber, H. (ed.). 1979. *Theory of Terminology and Terminological Lexicography*. Vienna/New York: Springer.

Fernández-López, M., A. Gómez-Pérez, and N. Juristo. 1997. "METHONTOLOGY: From Ontological Art Towards Ontological Engineering". In *The Spring Symposium on Ontological Engineering of AAAI*, 33–40. California: Stanford University.

Gómez-Pérez, A., M. Fernández-López, and O. Corcho. 2003. *Ontological Engineering. Advanced Information and Knowledge Processing Series*. Berlin: Springer Verlag.

Gómez-Pérez, A., and M. D. Rojas-Amaya. 1999. "Ontological Reengineering for Reuse." In *The 11th European Workshop on Knowledge Acquisition, Modeling and Management (EKAW 1999)*, LNCS Volume 1621/1999, 139–156. Berlin (Germany): Springer-Verlag.

Gómez-Pérez, A., and N. Juristo-Juzgado. 1993. "Towards a Consensus on the English-Spanish Translation of Knowledge Engineering Nomenclature." In *The 3rd International Congress on Terminology and Knowledge Engineering (TKE*), 193–199. Cologne (Germany).

Gómez-Pérez, A., N. Juristo-Juzgado., and J. Pazos. 1995. "Evaluation and Assessment of Knowledge Sharing Technology." In *Towards Very Large Knowledge Bases: Knowledge

Building and Knowledge Sharing (KBKS'95), ed. by N. Mars, 289–296. Amsterdam (The Netherlands): IOS Press.

Gómez-Pérez, J. M., C. Daviaud, B. Morera, R. Benjamins, T. Pariente Lobo, G. Herrero Cárcel, and G. Tort. 2006. "NeOn Deliverable D8.1.1 Analysis of the Pharma Domain and Requirements." Accessed October 15, 2012. http://www.neon-project.org/.

Herrera-Viedma, E., F. Mata, L. Martínez, and L. G. Pérez. 2005. "An Adaptive Module for the Consensus Reaching Process in Group Decision Making Problems". In *MDAI 2005*, ed. by V. Torra, Narukawa, & Y. Miyamoto, LNAI 3558, 89–98. Berlin/Heidelberg: Springer-Verlag.

Huang, C. R., N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari, and L. Prévot (ed.). 2010. *Ontology and the Lexicon: A Natural Language Processing Perspective*. Cambridge: Cambridge University Press.

Ibekwe SanJuan, F., A. Condamines, and M. T. Cabré, M. T. (guest editors). 2005. *Application-driven Terminology Engineering.* Special issue of *Terminology* 11 (1).

IEEE 1074-2006. *IEEE Standard for Developing a Software Project Life Cycle Process* (Revision of IEEE Std. 1074–1997).

IEEE 1074-1997. *IEEE Standard for Developing Software Life Cycle Processes* (Revision of IEEE Std. 1074–1995; Replaces IEEE Std. 1074.1-1995).

IEEE 1074.1-1995. *IEEE Guide for Developing Software Life Cycle Processes*.

IEEE 610.12-1990. *IEEE Standard Glossary of Software Engineering Terminology* (Revision and redesignation of IEEE Std. 792–1983).

ISO 1087-1:2000. *Terminology Work. Vocabulary. Part 1: Theory and Application*.

Iglesias, M., C. Caracciolo, Y. Jaques, M. Sini, F. Calderini, J. Keizer, F. Le Hunte Ward, M. Nissim, and A. Gangemi. 2006. "NeOn Deliverable D7.1.1 WP7 User Requirements." Accessed October 15, 2012. http://www.neon-project.org/.

Jacquemin, C. 2001. *Spotting and Discovering Terms through Natural Language Processing*. Cambridge (MA): MIT Press.

Kerremans, K., R. Temmerman, and P. De Baer. 2008. "Construing Domain Knowledge via Terminological Understanding." *Linguistica Antverpiensia* 7: 177–192.

Kalfoglou, Y., and M. Schorlemmer. 2003. "Ontology Mapping: The State of the Art." *The Knowledge Engineering Review* 18 (1): 1–31.

Kingston, J. 1994. "Linking Knowledge Acquisition with CommonKADS Knowledge Representation." In *BCS SGES Expert Systems 1994 Conference*. Cambridge: St John's College.

Kotis, K., and G. Vouros. 2004. "The HCONE Approach to Ontology Merging." In *The Semantic Web: Research and Applications. First European Semantic Web Symposium (ESWS 2004)*. Heraclion, Crete, Greece.

Landau, S. I. 1984. *Dictionaries: The Art and Craft of Lexicography.* Cambridge: Cambridge University Press.

Leuf, B., and W. Cunningham. 2001. *The Wiki Way.* Boston, MA: Addison-Wesley.

L'Homme, M. C. 2004. "Sélection des termes dans un corpus d'informatique: comparaison de corpus et critères léxico-sémantiques» In *Euralex International Congress*, 583–593. Lorient (France).

Malaisé, V., P. Zweigenbaum., and B. Bachimont. 2005. "Mining Defining Contexts to Help Structuring Differential Ontologies." *Application-driven Terminology Engineering Special issue, Terminology*, ed. by F. Ibekwe SanJuan, A. Condamines, and M. T. Cabré, Volume 11 (1).

Meyer, I., V. Zaluski, and K. Mackintosh. 1997. “Metaphorical Internet Terms: A Conceptual and Structural Analysis.” *Terminology* 4 (1): 1–33.

Montiel-Ponsoda, E., G. Aguado de Cea, A. Gómez-Pérez, and W. Peters. 2011. “Enriching Ontologies with Multilingual Information.” *Journal of Natural Language Engineering* 17 (3): 283–309.

Munn, K., and B. Smith (eds). 2008. *Applied Ontology: An Introduction*. Frankfurt: Ontos Verlag.

Nazar R., J. Vivaldi., and L. Wanner. 2012. “Automatic Taxonomy Extraction for Specialized Domains Using Distributional Semantics.” *Terminology* 18 (2): 188–225.

Nuopponen, A. 2007. “Terminological Modelling of Processes.” In *Indeterminacy in Terminology and LSP*, ed. by B. E. Antia, 199–216, Amsterdam: John Benjamins Publishing.

Pearson, J. 1998. *Terms in Context*. Amsterdam: John Benjamins Publishing.

Pinto, H. S., C. Tempich, and S. Staab. 2004. “DILIGENT: Towards a Fine-grained Methodology for DIstributed, Loosely-controlled and EvolvInG Engineering of oNTologies.” In *The 16th European Conference on Artificial Intelligence (ECAI 2004)*, ed. by Ramón López de Mantaras and Lorenza Saitta, 393–397. Valencia (Spain): IOS Press.

Rey, A. 2005. “The Concept of Neologism and the Evolution of Terminologies in Individual Languages.” *Terminology* 11 (2): 311–331.

Sager, J. 1993. *Curso práctico sobre el procesamiento de la terminología*. Madrid: Fundación Germán Sánchez Ruipérez.

Saint, S., and J. R. Lawson. 1994. *Rules for Reaching Consensus: A Modern Approach to Decision Making.* San Francisco: Jossey-Bass.

Schumann, A. K. 2012. “Towards the Automated Enrichment of Multilingual Terminology Databases with Knowledge-Rich Contexts.” In *Proceedings of the Conference Dialogue 2012, ed. by* A. E. Kibrik et al., Vol. 1, 559–567. Moscow: Rossijskij Gosudarstvennyj Gumanitarnyj Universitet.

Sierra, G., R. Alarcón, C. Aguilar, and C. Bach. 2008. “Definitional Verbal Patterns for Semantic Relation Extraction.” *Terminology* 14 (1): 74–98.

Staab, S., H. P. Schnurr, R. Studer, and Y. Sure. 2001. “Knowledge Processes and Ontologies.” *IEEE Intelligent Systems* 16 (1): 26–34.

Stojanovic, L. 2004. “Methods and Tools for Ontology Evolution.” Ph.D. Thesis presented at the University of Karlsruhe (Germany).

Stojanovic, L., N. Stojanovic, and S. Handschuh. 2002. “Evolution of the Metadata in the Ontology-based Knowledge Management Systems”. In *The 1st German Workshop on Experience Management: Sharing Experiences about the Sharing of Experience*. Berlin.

Studer, R., V. R. Benjamins, and D. Fensel. 1998. “Knowledge Engineering: Principles and Methods.” *Data & Knowledge Engineering* 25 (1–2): 161–197.

Suárez-Figueroa, M. C. 2010. “NeOn Methodology for Building Ontology Networks: Specification, Scheduling and Reuse”. Ph.D. Thesis presented at Universidad Politécnica de Madrid (Spain). http://oa.upm.es/3879/

Suárez-Figueroa, M. C., A. Gómez-Pérez, E. Motta, and A. Gangemi (eds). 2012. *Ontology Engineering in a Networked World*. Berlin: Springer.

Temmerman, R. 2000. *Towards New ways of Terminology Description: The Sociocognitive Approach*. Amsterdam: John Benjamins Publishing.

Temmerman, R., and K. Kerremans. 2003. “Ontology Building and the Sociocognitive Approach to Terminology Description.” In *International Congress of Linguists. CIL 17*, ed. by E. Hajicová, A. Koteovcová, and J. Mírovský. 1–10. Prague: Matfyzpress.

The Compact Edition of the Oxford Dictionary. 1971/1987. Oxford: Oxford University Press.
Trimble, L. 1985. *English for Science and Technology*. Cambridge: Cambridge University Press.
Velardi, P., R. Poler, and J. V. Tomás-Miquel. 2006. "Methodology for the Definition of a Glossary in a Collaborative Research Project and its Application to a European Network of Excellence." In *Interoperability of Enterprise Software and Applications*, ed. by D. Konstantas, J. -P. Bourrières, M. Léonard, and, N. Boudjlida, 311–322. London: Springer-Verlag.
Viégas, F. B., M. Wattenberg, and K. Dave. 2004. "Studying Cooperation and Conflict between Authors with History Flow Visualizations." In Proceedings of the *SIFCHI Conference on Human Factors in Computing Systems*, 575–582. New York, NY, USA.
de Vries, Henk. 1997. "Standardization-What's in a Name." *Terminology* 4 (1): 55–83.
West J. A., and M. L. West. 2008. *Using Wikis for Online Collaboration: The Power of the Read-Write Web*. San Francisco: Jossey-Bass.
Wüster, E. 1968. *The Machine Tool: An Interlingual Dictionary of Basic Concepts*. London: Technical Press.

Appendix I. NeOn Glossary of Processes and Activities

<u>Processes</u>:

– ***Ontology Aligning.*** It refers to the activity of finding the correspondences between two or more ontologies and storing them. A synonym for this activity is Ontology Mapping.
– ***Ontology Design Pattern Reuse.*** It refers to the process of using available ontology design patterns in the solution to different modelling problems during the development of new ontologies.
– ***Ontology Module Reuse.*** It refers to the process of using available ontology modules in the solution of different problems.
– ***Ontology Reengineering*** (Gómez-Pérez and Rojas-Amaya 1999). It refers to the process of retrieving and transforming a conceptual model of an implemented ontology into a new, more correct and more complete conceptual model, which is re-implemented.
– ***Non-Ontological Resource Reengineering.*** It refers to the process of retrieving and transforming a non-ontological resource[9] (data bases, controlled vocabularies, etc.) into an ontology.
  This process could be compared with the ontology learning activity with the difference that in this activity the knowledge is only transformed into conceptual structures, whereas in the process of reengineering non-ontological resources the sources can be transformed into conceptual structures and instance data.
– ***Non-Ontological Resource Reuse.*** It refers to the process of taking the available non-ontological resources (data bases, controlled vocabularies, etc.) for the development of ontologies.
– ***Ontological Resource Reuse.*** It refers to the process of using available ontological resources[10] (ontologies, modules, statements, or ontology design patterns) for solving different problems (e.g. the development of different ontology-based applications, the activity of ontology aligning (as background knowledge), etc.).
– ***Ontology Reuse.*** It refers to the process of using available ontologies for solving different problems.

– ***Ontology Statement Reuse.*** It refers to the process of using available ontology statements in the solution of different problems.
– ***Ontology Validation.*** It refers to the ontology evaluation process that compares the meaning of the ontology definitions against the intended model of the world aiming to conceptualize. It answers the question "Are you producing the right ontology?".

Activities:

– ***Ontology Annotation.*** It refers to the activity of enriching the ontology with additional information, e.g. metadata or comments.
– ***Ontology Assessment.*** It refers to the activity of checking an ontology against the user's requirements, such as usability, usefulness, abstraction, quality.
– ***Ontology Comparison.*** It refers to the activity of finding differences between two or more ontologies or between two or more ontology modules.
– ***Ontology Conceptualization.*** It refers to the activity of organizing and structuring the information (data, knowledge, etc.), obtained during the acquisition process, into meaningful models at the knowledge level and according to the ontology requirements specification document. This activity is independent of the way in which the ontology implementation will be carried out.
– ***Ontology Configuration Management*** (Gómez-Pérez et al. 2003). It refers to the activity of recording all the versions of the documentation, software and ontology code, and of controlling the changes.
– ***Control*** (Gómez-Pérez et al. 2003). It refers to the activity of guaranteeing that the activities scheduled in the ontology development process are completed and performed in the manner intended.
– ***Ontology Customization.*** It refers to the activity of adapting an ontology to a specific user's needs.
– ***Ontology Diagnosis.*** It refers to the activity of identifying parts of the ontology directly responsible for incorrectness and incompleteness. Ontology diagnosis is triggered by ontology validation.
– ***Ontology Documentation.*** It refers to the collection of documents and explanatory comments generated during the entire ontology building process.
  Examples of the documents external to the implemented ontology include ontology requirement specification document, sources used for acquiring knowledge, ontology conceptualization document, design and decision criteria, etc.
  The information inside the implemented ontology includes natural language comments, ontology metadata, and implementation code.
  In summary, anything that could be useful to help users, who did not build the ontology, to understand and learn how the ontology was built. Note that the level of granularity of descriptions can help the understanding of the ontology.
– ***Ontology Elicitation.*** It refers to the knowledge acquisition activity in which conceptual structures (e.g. T-Box) and their instances (e.g. A-Box) are acquired from domain experts and other knowledge resources.
– ***Ontology Enrichment.*** It refers to the activity of extending an ontology with new conceptual structures (e.g. concepts, roles and axioms).
– ***Ontology Environment Study.*** It refers to the activity of analyzing the environment in which the ontology is going to be developed.
– ***Ontology Evaluation.*** It refers to the activity of checking the technical quality of an ontology against a frame of reference.

- ***Ontology Evolution*** (Stojanovic 2004). It refers to the activity of facilitating the modification of an ontology by preserving its consistency.
  Ontology Evolution can be seen as a consequence of different activities during the development of the ontology.
- ***Ontology Extension.*** It refers to the ontology enrichment activity for stretching the ontology in width.
- ***Ontology Feasibility Study*** (Gómez-Pérez et al. 2003). It refers to the activity of answering questions such as "Is it possible to build the ontology? and/or "Is it suitable to build the ontology?".
- ***Ontology Formalization*** (Gómez-Pérez et al. 2003). It refers to the transformation of a conceptual model into a formal or semi-computable model according to a knowledge representation paradigm (e.g. description logics, frames, and rules).
- ***Ontology Forward Engineering*** (Gómez-Pérez and Rojas-Amaya 1999). It refers to the activity of outputting a new implementation of the ontology on the basis of the new conceptual model.
- ***Ontology Implementation.*** It refers to the activity of generating computable models according to the syntax of a formal representation language (e.g. RDF(S), OWL, and FLogic).
- ***Ontology Integration.*** It refers to the activity of including one ontology in another ontology.
- ***Knowledge Acquisition for Ontologies.*** It refers to the activity for capturing knowledge (e.g. T-Box and A-Box) from a variety of sources (e.g. documents, experts, data bases, etc.). We can distinguish between Ontology Elicitation, Ontology Learning and Ontology Population.
- ***Ontology Learning.*** It refers to the knowledge acquisition activity that relies on (semi-) automatic methods to transform unstructured (e.g. corpora), semi-structured (e.g. folksonomies and html pages) and structured data sources (e.g. data bases) into conceptual structures (e.g. T-Box).
- ***Ontology Localization.*** It refers to the adaptation of an ontology to a particular language and culture.
- ***Ontology Mapping.*** It refers to the activity of finding the correspondences between two or more ontologies and storing/exploiting them. A synonym for this activity is Ontology Aligning.
- ***Ontology Matching.*** It refers to the activity of finding or discovering relationships or correspondences between entities of different ontologies or ontology modules.
  Ontology Matching can be seen as the first stage of Ontology Aligning.
- ***Ontology Merging.*** It refers to the activity of creating a new ontology or an ontology module from two or more, possibly overlapping, source ontologies or ontology modules.
- ***Ontology Modification*** (Stojanovic 2004). It refers to the activity of changing the ontology without considering the consistency.
- ***Ontology Modularization.*** It refers to the activity of identifying one or more modules in an ontology with the purpose of supporting reuse or maintenance.
  A distinction between Ontology Module Extraction and Ontology Partitioning can be made.
- ***Ontology Module Extraction.*** It refers to the activity of obtaining from an ontology some concrete modules that could be used for a particular purpose (e.g. to contain a particular sub-vocabulary of the original ontology).
- ***Ontology Partitioning.*** It refers to the activity of dividing an ontology into a set of (not necessary disjoint) modules that together form an ontology but that can be treated separately.

- ***Ontology Population.*** It refers to the knowledge acquisition activity that relies on (semi-) automatic methods to transform unstructured (e.g. corpora), semi-structured (e.g. folksonomies and html pages) and structured data sources (e.g. data bases) into instance data (e.g. A-Box).
- ***Ontology Pruning.*** It refers to the activity of discarding conceptual structures (e.g. part of T-Box) of a given ontology that are not or no longer relevant.
  Pruning can be used in combination with ontology learning methods to discard potentially irrelevant learned concepts/relations.
- ***Ontology Quality Assurance*** (Gómez-Pérez et al. 2003). It refers to the activity of assuring that the quality of each and every process carried out and each and every product built (ontology, software and documentation) is satisfactory.
- ***Ontology Repair***. It refers to the activity of solving errors (incompleteness, incorrectness) in the ontology. This activity is triggered by ontology diagnosis.
- ***Ontology Requirements Specification.*** It refers to the activity of collecting the requirements that the ontology should fulfil (for example, reasons to build the ontology, identification of target groups and intended uses). Such requirements may be reached through a consensus process.
- ***Non-Ontological Resource Reverse Engineering.*** It refers to the activity of analyzing a non-ontological resource in order to identify its underlying components and creating a representation of the resource at higher levels of abstraction.
- ***Non-Ontological Resource Transformation.*** It refers to the activity of generating an ontological model at different levels of abstraction from the non-ontological resource.
- ***Ontology Restructuring*** (Gómez-Pérez and Rojas Amaya 1999). It refers to the activity of correcting and reorganizing the knowledge contained in an initial conceptual model, and detecting missing knowledge.
  This activity contains two different tasks: analysis and synthesis. The goal of the analysis is to evaluate the ontology technically, that is, to check that the hierarchy of the ontology and its classes, instances, relations and functions are complete (contain all the definitions required for the domain of chemical substances), consistent (there are no contradictions in the ontology and with respect to the knowledge sources used), concise (there are no explicit and implicit redundancies) and syntactically correct. On the other hand, the synthesis task seeks to correct the ontology after the analysis phase and to document any changes made.
- ***Ontology Reverse Engineering*** (Gómez-Pérez and Rojas-Amaya 1999). It refers to the activity of outputting a possible conceptual model on the basis of the code in which the ontology is implemented.
- ***Scheduling*** (Gómez-Pérez et al. 2003). It refers to the activity of identifying the different activities and processes to be performed during the ontology development, their arrangement, and the time and resources needed for their completion.
- ***Ontology Search.*** It refers to the activity of finding candidate ontologies or ontology modules to be reused.
- ***Ontology Selection.*** It refers to the activity of choosing the most suitable ontologies or ontology modules among those available in an ontology repository or library, for a concrete domain of interest and associated tasks.
- ***Ontology Specialization.*** It refers to the ontology enrichment activity for extending the ontology in depth.
- ***Ontology Summarization.*** It refers to the activity of providing an abstract or summary of the ontology content.

The summary can include, for example, a couple of top level concepts in the ontology class hierarchy (perhaps a graphical representation of these top-level concepts and the links between them).

– ***Ontology Translation.*** It refers to the activity of changing the representation formalism or language of an ontology to another. Ontology translation can be part of an ontology reengineering process.
– ***Ontology Update.*** It refers to minor changes carried out in an ontology that could not be considered an upgrade.
– ***Ontology Upgrade.*** It refers to the activity of replacing an ontology with a new version.
– ***Ontology Verification.*** It refers to the ontology evaluation activity that compares the ontology against the ontology requirement specification document (ontology requirements and competency questions), thus ensuring that the ontology is built correctly (in compliance with the ontology requirements specification).
  It answers the question "Are you producing the ontology right?".
– ***Ontology Versioning*** (Stojanovic 2004). It refers to the activity of handling ontology changes by creating and managing different versions of the ontology.

Authors' addresses

Mari Carmen Suárez-Figueroa, Guadalupe Aguado de Cea, Asunción Gómez-Pérez
Ontology Engineering Group (OEG).
Facultad de Informática, Universidad Politécnica de Madrid (UPM).
Campus de Montegancedo s/n.
28660 Boadilla del Monte. Madrid. Spain

{mcsuarez, lupe, asun}@fi.upm.es

About the authors

Mari Carmen Suárez-Figueroa is Teaching Assistant at the Universidad Politécnica de Madrid (UPM). She got the PhD in Artificial Intelligence in UPM in June 2010. She has received an Outstanding Award granted by the Commission on Graduate Doctorate from the UPM. She is also a senior researcher at the Ontology Engineering Group (OEG). Her research activities are focused on Ontology Engineering, the Semantic Web, and Linked Data. Particularly, her research lines include methodologies for ontology network development, ontology network development in different domains, ontology evaluation and quality, and ontology design patterns. In these areas, she has participated in several European research projects (NeOn, SEEMP, OntoGrid, Knowledge Web, PIKON, Esperonto, and OntoWeb), and Spanish research projects (mIO! and BuscaMedia). She is one of the editors of the book entitled "Ontology Engineering in a Networked World" published by Springer in 2012. She co-organized the Terminology and Knowledge Engineering Conference (TKE 2012) and several international workshops (Workshop on Ontology Patterns (WOP 2012), Workshop on Ontology Engineering in a Data-driven World (OEDW 2012), etc.) and tutorials.

Guadalupe Aguado de Cea received her PhD in English Philology by Universidad Complutense de Madrid (UCM) and holds a MA in Translation by the UCM. She lectures in Specialized Communication and Terminology. Her current research activities include, among others: Terminology and Ontologies, Multilinguality, and Natural Language Processing (NLP), in which she has participated in different projects concerning terminology, ontologies and multilinguality. She is the President of the Committee on Terminology, AEN-CTN-191, AENOR, and the representative for Spain at the ISO TC 37.

Prof. Asunción Gómez-Pérez is Full Professor at UPM, director of the Artificial Intelligence department, director or the OEG and PhD in Computer Science (1993). Before joining UPM, she was visiting (1994–1995) the Knowledge Systems Laboratory at Stanford University. She has coordinated in more than 15 EU projects. She has done extensive research on ontologies, Linked Data and the Semantic Web. She has published more than 150 papers and two books on Ontological Engineering. She was program chair of many international conferences.