NeOn: Lifecycle Support for Networked Ontologies

Integrated Project (IST-2005-027595)

Priority: IST-2004-2.4.7 – "Semantic-based knowledge and content systems"

D8.3.1 Ontologies for the Pharmaceutical Case Studies

Deliverable Co-ordinator: Jose Manuel Gómez-Pérez

Deliverable Co-ordinating Institution: Intelligent Software Components (iSOCO)

Other Authors: Carlos Buil (iSOCO); Germán Herrero, Tomás Pariente, Angel Baena (ATOS); Joan Candini, Juan Carlos Dalmacio (KIN)

Document Identifier:	NEON/2007/D8.3.1/v1.0	Date due:	August 31, 2007
Class Deliverable:	NEON EU-IST-2005-027595	Submission date:	August 31, 2007
Project start date:	March 1, 2006	Version:	v1.0
Project duration:	4 years	State:	Final
		Distribution:	Public

NeOn Consortium

This document is a part of the NeOn research project funded by the IST Programme of the Commission of the European Communities by the grant number IST-2005-027595. The following partners are involved in the project:

Open University (OU) – Coordinator Knowledge Media Institute – KMi Berrill Building, Walton Hall Milton Keynes, MK7 6AA United Kingdom Contact person: Martin Dzbor, Enrico Motta E-mail address: {m.dzbor, e.motta} @open.ac.uk	**Universität Karlsruhe – TH (UKARL)** Institut für Angewandte Informatik und Formale Beschreibungsverfahren – AIFB Englerstrasse 11 D-76128 Karlsruhe, Germany Contact person: Peter Haase E-mail address: pha@aifb.uni-karlsruhe.de
Universidad Politécnica de Madrid (UPM) Campus de Montegancedo 28660 Boadilla del Monte Spain Contact person: Asunción Gómez Pérez E-mail address: asun@fi.upm.es	**Software AG (SAG)** Uhlandstrasse 12 64297 Darmstadt Germany Contact person: Walter Waterfeld E-mail address: walter.waterfeld@softwareag.com
Intelligent Software Components S.A. (ISOCO) Calle de Pedro de Valdivia 10 28006 Madrid Spain Contact person: Jesús Contreras E-mail address: jcontreras@isoco.com	**Institut 'Jožef Stefan' (JSI)** Jamova 39 SI-1000 Ljubljana Slovenia Contact person: Marko Grobelnik E-mail address: marko.grobelnik@ijs.si
Institut National de Recherche en Informatique et en Automatique (INRIA) ZIRST – 655 avenue de l'Europe Montbonnot Saint Martin 38334 Saint-Ismier France Contact person: Jérôme Euzenat E-mail address: jerome.euzenat@inrialpes.fr	**University of Sheffield (USFD)** Dept. of Computer Science Regent Court 211 Portobello street S14DP Sheffield United Kingdom Contact person: Hamish Cunningham E-mail address: hamish@dcs.shef.ac.uk
Universität Koblenz-Landau (UKO-LD) Universitätsstrasse 1 56070 Koblenz Germany Contact person: Steffen Staab E-mail address: staab@uni-koblenz.de	**Consiglio Nazionale delle Ricerche (CNR)** Institute of cognitive sciences and technologies Via S. Martino della Battaglia, 44 - 00185 Roma-Lazio, Italy Contact person: Aldo Gangemi E-mail address: aldo.gangemi@istc.cnr.it
Ontoprise GmbH. (ONTO) Amalienbadstr. 36 (Raumfabrik 29) 76227 Karlsruhe Germany Contact person: Jürgen Angele E-mail address: angele@ontoprise.de	**Food and Agriculture Organization of the United Nations (FAO)** Viale delle Terme di Caracalla 1 00100 Rome Italy Contact person: Marta Iglesias E-mail address: marta.iglesias@fao.org
Atos Origin S.A. (ATOS) Calle de Albarracín, 25 28037 Madrid Spain Contact person: Tomás Pariente Lobo E-mail address: tomas.parientelobo@atosorigin.com	**Laboratorios KIN, S.A. (KIN)** C/Ciudad de Granada, 123 08018 Barcelona Spain Contact person: Antonio López E-mail address: alopez@kin.es

NeOn

Work package participants

The following partners have taken an active part in the work leading to the elaboration of this document, even if they might not have directly contributed to the writing of this document or its parts:

iSOCO

ATOS

KIN

Change Log

Version	Date	Amended by	Changes
0.1	01/08/07	Jose Manuel Gómez-Pérez	Layout and table of contents
0.2	12/09/2007	Carlos Buil	Integration Invoicing & Nomenclature contributions
0.3	16/09/2007	Germán Herrero	Revision of the deliverable & Nomenclature Reference Ontology description
0.4	05/10/07	Jose Manuel Gómez-Pérez	Some updates and exec summary and conclusions added
0.5	11/10/2007	Carlos Buil	QA received and document updated
1.0	15/10/2007	Jose Manuel Gómez-Pérez	Submission

Executive Summary

This document contains a description of the ontologies which have been developed so far in the context of the NeOn Pharmaceutical case studies, encompassing the following issues: i) application of the NeOn methodology to the Pharmaceutical case studies and, in return, how the case studies have contributed in practice to the development of this methodology; ii) inventory of existing knowledge resources, either ontological or non-ontological; iii) ontologies resulting from the application of this methodology to the Pharmaceutical domain; and iv) how the development of new ontologies, together with reutilization and extension of existing ones and the semantization of non-ontological resources contribute to improving the problems addressed by the case studies.

This document provides a twofold testbed for the NeOn methodology, which has been applied to the management of the knowledge lifecycle in two varied aspects of the pharmaceutical domain: electronic invoicing, i.e. improving interoperability in the exchange of business electronic documents, and semantic nomenclature of the various pharmaceutical products across the different existing repositories. We first describe the methodological approach used to develop the ontologies resulting from this deliverable and then describe its application to the two particular case studies as well as the results produced. Mechanisms like the Competency Questions have provided additional insight and refinement of user requirements particularly on knowledge resources. On the other hand, this document also intends to show the use of NeOn metamodel in the pharmaceutical ontologies described within.

We provide an extensive inventory of the pre-existing resources surveyed in order to accomplish such requirements and then refined it into the definitive set of resources that we have finally reused

via either extension or customization. These resources include ontological and non-ontological resources, as well as the most relevant standards, e.g. for exchange of electronic B2B documents like EDIFACT or UBL or the ATC and EphMRA chemicals classification systems. Using these resources as a starting point, we provide for both case studies networked ontologies that can be used as reference knowledge for the general invoicing and nomenclature cases. We also provide their specializations for particular applications like e.g. in the case of invoicing specific support to the PharmaInnova cluster of laboratories. Finally, we provide a glimpse of the use of these ontologies in the context of the software prototypes to be produced in WP8 from M24 on.

Table of Contents

List of tables

List of figures

1. Introduction

This document contains a description of the ontologies which have been developed so far in the context of the NeOn Pharmaceutical case studies. In this regard, the main issues to be addressed in this deliverable are:

- Application of the NeOn methodology to the Pharmaceutical case studies and insight into how the case studies have contributed in practice to the development of this methodology in return.
- Inventory of existing knowledge resources, either ontological or non-ontological.
- Ontologies resulting from the application of this methodology to the Pharmaceutical domain.
- How the development of new ontologies, together with reutilization and extension of existing ones and the semantization of non-ontological resources contribute to solving the problems addressed by the case studies.

The main objective of this deliverable is to co-ordinately produce a number of useful ontological resources which support the users of the Pharmaceutical case studies in a series of activities central to their operations.

First, we aim at cataloguing information about pharmaceutical products in order to provide pharmacies, through their nation and Europe-wide associations, with access to homogenized information repositories on such products. This information is highly distributed and lacking of consistency mechanisms and real-time access. The networked ontologies resulting from the work described in this document are key to leverage integration of these heterogeneous repositories. Ultimately, as a solution to this problem, we will produce a global Vademecum where the integration of the distributed databases and regulations can be done and kept up to date.

Second, we aim towards the financial side of the pharmaceutical sector, in particular electronic invoicing. Since its authorization in 2002, the use of electronic invoices for commercial transactions has grown exponentially [1] . Such take up has been accompanied with large heterogeneity of the means to represent and exchange invoice information, as well as the lack of invoice standards adopted by the main players of the sector. The networked ontologies presented in this document are key to i) providing a conceptual model of the information related with invoicing, which embraces the different existing standards, ii) ensuring consistency of exchanged invoice data with respect to the formal model of these ontologies, and ultimately iii) allowing users to easily define themselves the mapping between their invoices and a common, consensuated model, supported by the ontologies, to automate invoice exchange between business peers.

The remaining of this document is structured as follows. Section 2 describes the methodology used for the development of the ontologies presented herein; section 3 contains an inventory and analysis of the ontological and non-ontological resources which have been selected for reuse in the context of our ontologies; section 4 describes what ontologies (and according to which parameters) have been actually reused. Finally, section 5 describes the ontologies and how they are intended to be exploited in the main activities accomplished by the users of the Pharmaceutical case studies, by means of the prototypes to be delivered at the end of M24.

2. NeOn Methodology in the Pharmaceutical Case Studies

In this section we are going to describe the methodology that has been followed in the development of the ontologies for the two pharmaceutical case studies.

2.1. Methodology followed in the Pharmaceutical Case Studies

Currently there is not a methodology to work with networked ontologies. There exist methodologies for designing ontologies like Methontology [13] or DILIGENT [21] . In the context of WP5 in NeOn a methodology for covering the missing points in developing networked ontologies is being developed and represented in [14] . In this section we explain the main points of this methodology and how it is applied to the pharmaceutical use cases.
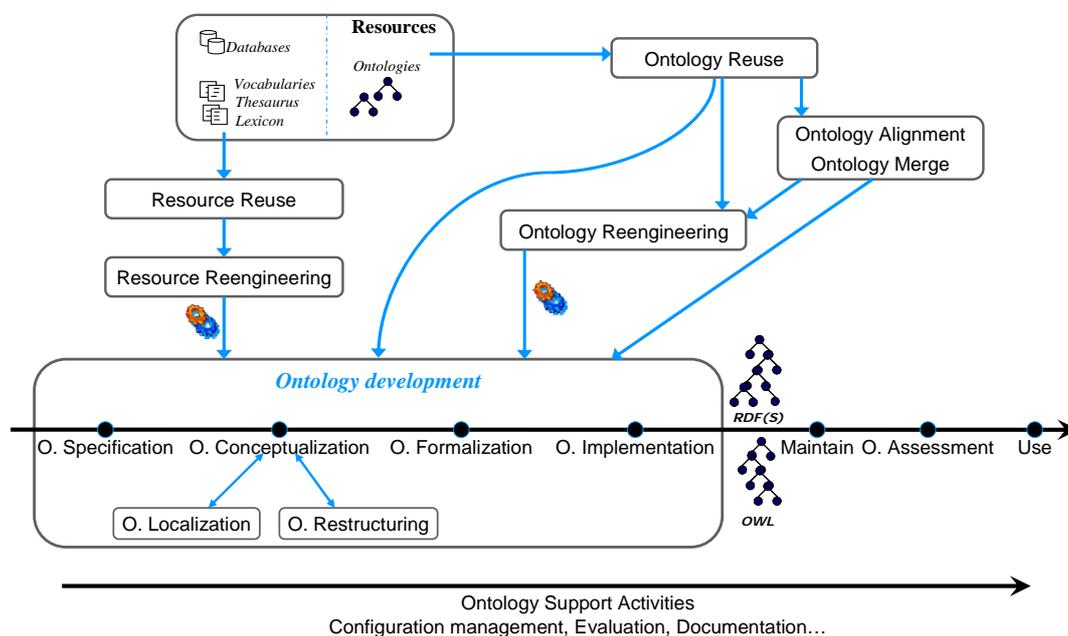
Figure 1: NeOn methodology [14]

2.1.1. Methodological approach to the ontology development process

The preliminary work plan for developing the main ontologies of the case study included 4 main tasks, following a waterfall life cycle model.

- Task 1. Knowledge acquisition for ontologies. In this task a research process over the existing resources related to the invoicing problem is done.

- Task 2. Ontology Conceptualization. During this task the initial ontologies are created, using the resources previously gathered.

- Task 3. Ontology validation. In this task a validation process of the created management ontology network is done. The validation process checks the consistency of the ontology network.

- Task 4. Ontology specialization. In this task the ontologies were used in the particular scenarios of each case study.

These were the main tasks that of course were specialized for each case study. However, after several bi-lateral ad-hoc meetings between ATOS, ISOCO and UPM, the initial work plan for developing the ontology network in the case studies was modified. During such meetings, iSOCO and ATOS explained the use cases and UPM presented methodological guidelines. Based on the guidelines and suggestions provided by UPM, the first step was identifying which kind of ontology network lifecycle model is the most appropriate to the case studies.
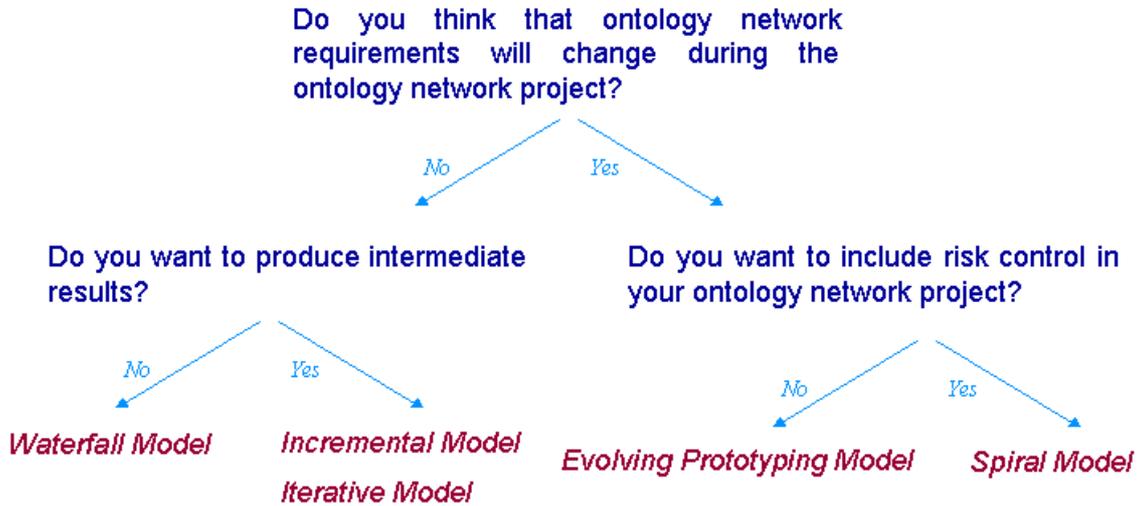
Figure 2: Tree for Selecting the Ontology Network Lifecycle Model [14]

2.1.2. Ontological activities performed

Figure 3 shows the main activities followed by the pharmaceutical case studies. The activities in and its order this figure are described in [14] . These activities have been performed in both use cases but the importance of them varies depending on the type of ontology developed. It is important to highlight the activities ontology environment study, ontology enrichment and the selection of the standards that are part of the specific ontologies.
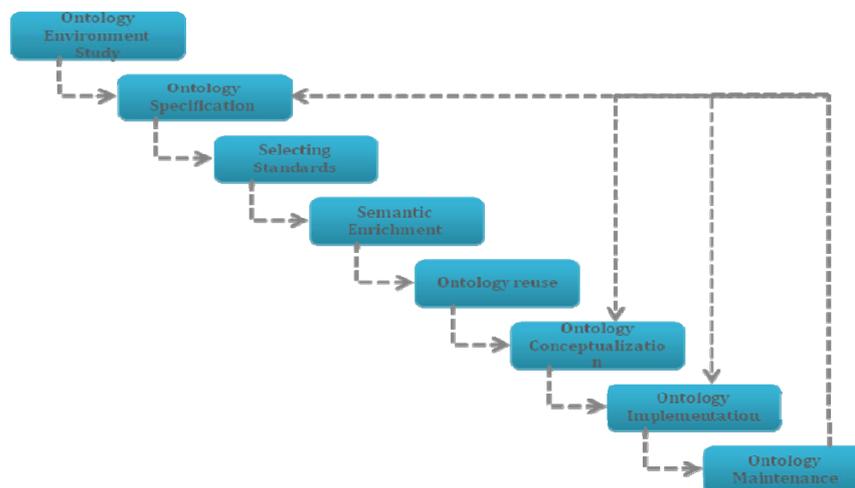
Figure 3: Ontological activities

2.1.3. Overview of the Competency Questions

The specification activity states why the ontology is being built, what its intended uses are and who the end-users are. For specifying the ontology requirements we will use the competency questions

techniques proposed in [15] Before identifying them, we identify the intended uses of the ontology and their users.

2.1.3.1.   Intended uses of the ontology

The development of the network of ontologies is motivated by scenarios related to the application that will use the ontology. Such scenarios describe a set of the ontology's requirements that the ontology should satisfy after being formally implemented. The motivating scenarios are described in [12] .

2.1.3.2.   Intended users of the ontology

In order to get a fully usable ontology it is necessary to know who is going to finally use the ontologies developed. In order to address this problem, competency questions regarding the intended users of the ontology are mandatory. In [1] and [12] a complete description of the users of these applications is presented.

2.1.3.3.   Competency Questions

Competency questions are natural language questions used to determine the scope of the ontology to be built. These questions and their answers are both used to extract the main concepts and their properties, relations and formal axioms of the ontology. The competency questions play the role of a type of requirement specification against which the ontology can be evaluated. Specific competency questions can be composed into more general questions that are answered by composing answers associated to the specific competency questions. The list of competency questions is based in the following topics:

2.2.   Applying the Methodology to the Invoicing Case Study

In this section we describe the application of the methodology proposed in [14] and briefly described in the previous section.

2.2.1. Methodological approach to the ontology development process

The ontology network life cycle model chosen in the invoicing case study is finally the incremental model. After meetings at iSOCO with the UPM team we chose the networked ontology life cycle as shown in Figure 2. In this figure we first chose no to the question "Do you think that ontology network requirements will change during the ontology network project?" The answer "no" to this question is due to the requirements specified in [1] . The main problematic of the use case is to deal with the heterogeneity of the invoices emitted and received, and how to deal with the incoming data. These requirements remain invariable in the life cycle of the ontology.

The answer to the second question, "Do you want to produce intermediate results?" is yes. During the evolution of the use case several prototypes and intermediate ontologies will be produced and they will be used along the project.

2.2.2. Ontology requirements for the invoicing case study

The requirements of the invoice reference ontology have been extracted from two different sources. The first source is [1] in which it is described the invoicing case study. The invoicing case

study deals with the heterogeneity of the invoices emitted and received by pharmaceutical organizations. The set of pharmaceutical organizations ranges from laboratories which provide drugs and medicines to other companies to providers that provide any type of good that a laboratory needs. In this range of companies it is possible to also find pharmacies, wholesalers or the public administration. Each organization uses its own invoice model and therefore they generate different invoices (instances of its own model). Also, the technologies used to create the invoices are different among companies that participate in the invoicing process. They can use comma separated value (CSV), iDOC (format developed by SAP[1]), EDIFACT[2], etc., and each one with their own implementation (the format is a standard but not the use of it). Therefore it is needed to solve the heterogeneity problem due to the need of interoperability between different organizations.

The second source is a list of competency questions [15] This list has been created from the use case requirements specified in [1] with the goal of refining such requirements. In this document the problems that are faced in the use case and the requirements that the system should provide in order to solve this problem are described. This deliverable provides a list of competency questions that the end users have answered and shows the need for different ontologies in the use case.

2.2.3. Activities followed to build the ontologies

Based on the methodological work being done in WP5, the activities carried out for building the invoice management ontology network are the following:

1. **Ontology Elicitation**. In this activity the pharmaceutical domain has been analysed. It is also described how the invoicing process is done and what actors and requirements are needed. It is important to highlight that in the requirement analysis it is described what happens to an invoice from the time it is emitted to the moment it is validated by the receiving company, including the actors that participate in this process, the requirements of each company in each cluster (laboratories, wholesalers and providers) and the NeOn technology that it is going to be used. It can be found in [1] .

2. **Ontology Specification,** using competency questions, of the necessities that the ontology has to satisfy in the new application. The list of competency questions and its description are included in 2.1.3.3.

3. **Knowledge resources reuse (Search of existing resources)**. In the community every day there are more and more ontologies available. These ontologies can be reused and imported as modules into the ontology network to be built. There exist also upper ontologies which contain a generic description of concepts. These descriptions will be used in the invoicing case study by reusing some of them included in upper ontologies such as SUMO [7] , Cyc[3] or DOLCE [7] . Also more specific ontologies related to the invoicing process that the companies follow and pharmaceutical ontologies are reused.

   From the previous competency questions the need for use of time ontologies or invoicing technologies vocabularies are identified. A set of time ontologies and its reuse is presented in [8] . In the invoicing case study a similar approach will be followed based in the necessities of the final users. In the same way resources about specific invoice technologies will be reused and included in the final invoice reference ontology.

---

[1] http://www.sap.com/index.epx

[2] http://www.edifactory.de/messages.php?s=D05A

[3] http://www.cyc.com/cycdoc/vocab/merged-ontology-vocab.html

The resources used for creating the invoicing ontology network can be organized in the following groups:

Upper level ontologies and related projects. The motivation for using upper level ontologies comes from the need of reuse of the main reference ontology for invoicing. The purpose of this ontology is that it can be instantiated for different sectors of the industry. The first instantiation is for the pharmaceutical sector, laboratories mainly, but it will also be extended for providers of these laboratories or wholesalers. These providers provide from chemical products to energy or clean products so they need different instantiations of the invoice reference ontology.

- *Invoicing resources*. These resources are mainly technologies for electronic invoicing. The technologies are the Universal Business Language (UBL), EDIFACT, xCBL or other proprietary solutions used in the industry.

- *Projects whose main goal is to integrate the invoice vocabulary into ontologies*. The problem of integrating invoice vocabulary like the one provided by UBL into ontologies is not new. There exist approaches like the ONTOLOG project[4] or the XBRL Ontology project[5]. The XBRL Ontology Specification provides main concepts and properties for describing financial and economic data on the Semantic Web. This last approach creates from scratch the invoice ontology but it is still in an early stage. Also exists other projects about integration like [22] , HARMONISE[6] or [23] in which semantic interoperability is achieved that will guide us for integrating the different technologies.

4. **Ontology Conceptualization** (Development of the invoice ontology network) In this step we conceptualize the resources analysed in the previous activities. At this point the invoice reference ontology is created. A detailed description of this activity is done can be found in section 5.1.

5. **Ontology Specialization (Adaptation of ontology network)**. The final invoice reference ontology will be adapted to the cluster of companies that are going to use it, a laboratory for instance. The invoice reference ontology will be specialized to each cluster of companies needs (laboratories in an initial phase).

6. **Ontology Localization (Localization of ontology network)**. Localization is not a very important need of this invoice ontology network but has to be taken into account. The users of the ontology belong to different regions in Spain in which different languages are used. Spanish is the official language but in these regions there are other co-official languages therefore localization has to be taken into account.

In the same way, if the ontology is adapted to other countries localization is strongly needed. Localization will be performed initially for the different Spanish regions and adapted to include other localizations.

The ontology network has been designed in English so all concepts are written in English. This ontology will be adapted by the end users to their own language and the particular needs these end user have.

---

[4] http://ontolog.cim3.net/

[5] http://xbrlontology.com/

[6] http://www.ercim.org/publication/Ercim_News/enw51/missikoff.html

7. **Ontology Evaluation (Evaluation of the ontology network)**. The ontology network will be evaluated by the users of PharmaInnova. From this ontology network invoices will be created and evaluated by the users. The evaluation will consist mainly in two steps. First the ontology will be validated by the end users, checking if there are all the concepts that they need. The second step of the evaluation will consist in adapting the invoice reference ontology using the prototype created in the scope of WP8 and emitting invoices that will be instances of the adapted ontology. In order to be able to evaluate the ontology in this second step more time is needed because the prototype is still in a development phase.

2.2.4. Selected Activities

In the ANNEX I: Activities used in the development of the invoice reference ontology", Table 14 shows the activities used in the invoicing case study. The third column has been added to the original table and each X indicates that the corresponding activity has been used in the building process of the invoice reference ontology or it is planned to be used. The selection of the activities has been based on the experience developing the invoice reference ontology.

2.2.5. Competency Questions

In this section we describe the competency questions that lead us to the design of the invoice reference ontology.

2.2.5.1.   Intended uses of the ontology

In [1] and [12] the intended uses of the ontologies that are going to be designed are explained in detail. In these documents it is shown how the users plan to use them and the functionalities that are expected from the use of the ontologies.

2.2.5.2.   Intended users of the ontology

The users of these applications are the following:

U1. User of the invoicing application who is going to model a new invoice.

U2. User who emits invoices.

U3. User who receives invoices.

U4. User who administrates the invoicing system.

U5. Developers of invoicing applications

Also it is important to take into account the sector of the companies emitting invoices. These companies belong to one of these groups: laboratories, wholesalers or providers. The competency questions also take into account this division of companies.

2.2.5.3.   Competency Questions

The competency questions are in part based on [1] and [12] in order to know what ontologies are needed by the use case and what conceptual level do the end users need. The list of competency questions is based on the following topics:

NeOn

- Specific competency questions related to the industrial sector of the companies that participate in the invoicing process. These sectors are wholesalers, laboratories, pharmacies and providers:

- Competency questions regarding the invoicing workflow.

- Specific competency questions related to multilinguality. These competency questions help to answer the need of multilinguality of the invoice reference ontology.

- Competency questions related inference rules

- Specific competency questions related to the receiver of invoices

- Specific competency questions related to the emitter of invoices

- Competency questions related to the technology used in the invoices

- Competency questions related to time and date management

- Competency questions related to currencies

- Composed competency questions

2.2.6. Conclusions Drafted

The ontology network for the invoicing case study in WP8 must contain all the concepts related to the invoice management in the pharmaceutical industry. These concepts shall reflect all the interactions and elements that are included in the pharmaceutical invoicing process. These elements vary from the most common concepts included in every invoice generated by a company to the technologies or the workflow that is followed by every invoice in the system.

The technologies and invoicing languages that are used by the companies in this invoicing process are in part specified by the result of the competency questions. The formats used are mainly CSV (by the 60% of the companies) and FLF (used by a 10% of the companies). These two formats are very simple and the most important is to have in the invoice reference ontology all the concepts that are used by the companies. There are two key electronic invoicing languages. These two languages are EDIFACT (electronic invoicing UN standard) and UBL. These two standards are world wide extended and we consider mandatory their presence in the invoice reference ontology.

The competency questions showed us the different representations of time and date that are used by the companies. Examples obtained directly from the invoices are "to be paid 20 days before from receiving date" or "to be paid before 20/07/2008".

The only currency used in the invoices, as it is shown by the competency questions, is Euro but due to the aim of generality and interoperability of the invoice process we have also included dollars and other currencies.

The competency questions along with the requirements document [1] demonstrate the need of representing two different workflows. The first workflow refers to the invoicing process. It is necessary to know in which state of the invoicing workflow an invoice is (Competency questions CQ10 and CQ15). It is also desirable to represent the workflow of the products in order to provide to the laboratories some market intelligence to boost their supply chain.

The competency questions also show that is possible to infer knowledge from the data contained in the invoices. Information about the final price of a product or the amount of money that has to be paid with taxes can be obtained by using inference over the invoices.

The competency questions show that the differences between the invoices belonging to different sector companies (laboratories, wholesalers, etc.) are not so many. Therefore the main complexity of the interoperability in the invoicing process relies in the different technologies/languages/formats

used. In the invoice reference ontology we make a complete representation of the previously mentioned invoicing technologies/languages and all the concepts that are in the invoices.

From the competency questions we also obtained the overall architecture of the network of ontologies. This architecture is based in a federated network of ontology with one central invoice reference ontology. The reference ontology use is based on the distribution of all the companies that are part of the invoicing process. In this invoice reference ontology all the technologies and concepts that are needed in the invoicing process are included.

By creating a reference ontology in which all the possible invoice models are represented it is possible to solve the problem. Every company (laboratories, providers, wholesalers, etc.) has their common invoice specifications and every company will have a model created from the invoice reference ontology. If a company emits an invoice based on an ontology customized from the invoice reference ontology this invoice will be automatically interpretable by the system.

The solution proposed in WP8 is that each company emits the invoices in its own format as instance of its own model generated from a customized ontology network. This customized ontology network will be created from a reference ontology network which contains all the elements needed for the invoicing process and adapted (by adding the terminology they need, missing concepts, etc) by each company to its needs. A first version of the prototype that uses this invoice ontology network is described in Section 5.3. This invoice reference ontology network will be created following the methodological guidelines specified by WP5.

The ontology network for the invoicing case study in WP8 must contain all the concepts related to the invoice management in the pharmaceutical industry. These concepts shall reflect all the interactions and elements that are included in the pharmaceutical invoicing process. These elements vary from the most common concepts included in every invoice generated by a company to the technologies or the workflow that is followed by every invoice in the system. Figure 4 shows the architecture of the network of ontologies in the invoicing case study. In this figure there are only represented some of the technologies that are considered in the deliverable. An extended list of the resources analyzed can be found in Section 3.1.
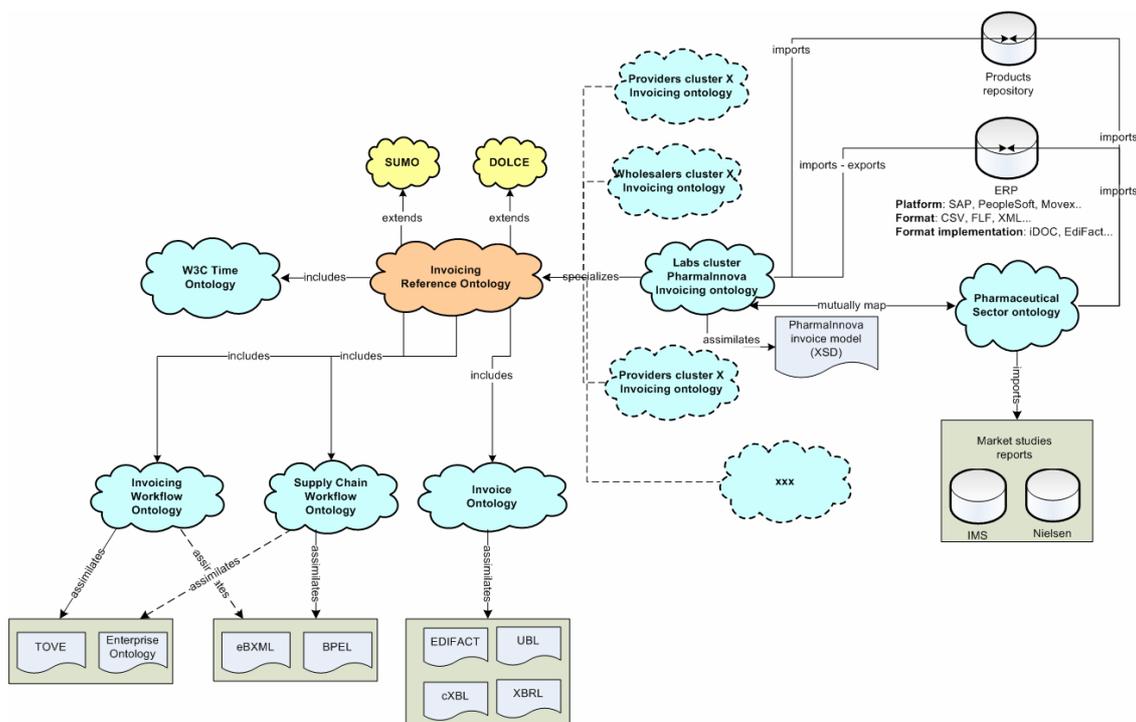
Figure 4. General View of the Invoice Management Ontology Network

In the ontology conceptualization activity which 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 according to the ontology specification document we will use the OWL NeOn metamodel [24] . This metamodel provides the means we need to conceptualize all the requirements specified in this section.

Two important considerations come up from the architecture of the invoice reference ontology. First is the need of modularization in this ontology. The invoice reference ontology will include several modules with electronic invoice technologies and languages. Therefore the ontology shall be scalable, and providing means for modularizing distributed ontologies is mandatory. For providing this feature we use the NeOn metamodel for ontology modularization [24] In the same way means for managing mappings between networks ontologies are needed. The reuse of different ontological resources (for example the XBRL ontology project[7]) may require mapping several concepts with the invoice reference ontology).

2.3.   Applying the Methodology to the Semantic Nomenclature Case Study

2.3.1. Semantic Nomenclature methodological approach to the ontology development process

One of the main goals in the Semantic Nomenclature is the development of the Nomenclature ontology network. Based on the guidelines and suggestions provided by UPM (WP5 leaders), as described in section 2.1.1, the first step was identifying which kind of ontology network lifecycle model is the most appropriate for the Semantic Nomenclature case study.

The selected ontology network lifecycle model was the iterative/incremental model. This decision was taken up based on the decision tree shown in Figure 2 suggested in [14] , derived from a previously study and identification of the different requirements or restrictions of the Nomenclature Ontology network and stated on the past experiences of our capabilities.

The main motivation of this decision are that the pharmaceutical scene is more or less static in their models, the pharmaceutical sector has a low frequency of change in knowledge level (not in data level), so is deduced that there will not be many changes in the ontology network requirements. Other main reason that motivates this decision is that it is planned to produce different versions of the networked ontology and application during the next months of the NeOn project.

One of the characteristics of the selected lifecycle model is that the ontology network is not completely developed in the first steps. This lifecycle model proposes different development iterations during the project, where in each iteration different parts of the network are improved. Also, we take the main idea of the incremental approach, "produce and deliver", when a new ontology is finished, it can be integrated in the ontology network and used.

2.3.2. Ontology requirements for the Semantic Nomenclature

During the ontology specification, should be taken into account the key ontology requirements for the Semantic Nomenclature described in [1] . The most relevant aspects of the ontology requirements are the following:

Reusability of modules

Some concrete modules of the networked ontologies should be reused from the existing resources, so the way of reusability of ontologies or modules may be facilitated. In this case study could be

---

[7] http://xbrlontology.com/

identified some modules as time, geographical, units… that are shared and reused in the reference model and in the different other models.

Mappings management

The creation of mappings between ontologies is out of the scope of this deliverable. Also, the mapping management is one of the most frequent actions that could occur in the editor role, in this case study. So, the NeOn toolkit should provide mechanisms for a graphical management of the mappings of the stakeholder ontologies.

Methodology / Model for Semantic Nomenclature

For the development of the Networked Ontologies, a methodology to describe guidelines for the editors and administrators of NeOn on how to design, implement and maintain the ontology network is needed. This methodology helps these users in several situations in the ontology's lifecycle. Also, the methodology should provide guidelines in order to facilitate the decision when a particular model could be involved in the network and when not, or if one unstructured resource could be modelled or reengineered…

Re-engineering & Semantic Enrichment

All the information of the pharmaceutical products are modelled and stored in different legacy systems and databases. Furthermore more resources like thesaurus, vocabularies and classifications are non-ontological resources. Previously to integrate these resources into the Nomenclature network, they should be re-engineered or semantically enriched. So, some methodologies or guides to re-engineering or semantic enrichment of this kind of resources could be interesting for the ontology editors.

Multilinguality

Due to the fact of the multilingualism in Spain (Spanish, Catalonian, Basque...) and the internationalization of the reference ontology a multilingual support is needed. Multilinguality could be identified at different levels: labels for concepts, relations… metadata element and ontology content… The multilinguality of the ontologies is out of the scope of this deliverable. However, new versions of the ontologies should be adapted to the multilinguality model provided by NeOn and a mechanism to link the concepts with ontologies developed in other languages comparing labels, facilitating the mapping management of concepts between two ontologies described in different languages.

Ontology Population

The core information about the pharmaceutical products is stored in several distributed databases. These databases are Access DBs, it is decided for the case study is more interested in leaving the individual records where they are, and only support their integration if is necessarily any population. R2O & ODEMapster provide the framework to upgrade relational legacy data (databases) to the ontologies, in this case, upgrade information from databases (Digitalis, BOTPlus…) to the respective ontologies. This technology is based on the declarative description of mappings between relational and ontology elements and in an exploitation of mappings by a domain independent processor.

2.3.3. Ontological activities performed in the Semantic Nomenclature

The main objective of the Semantic Nomenclature case study is focused on the integration of different and heterogeneous pharmaceutical product information repositories [1] . Information about medicine and medical and regulatory knowledge is highly distributed and in some cases inconsistent with the government regulation and database. As a solution to this problem a global Vademecum in a reference ontology is proposed, based on an ontology network, where the distributed resources of the domain are integrated and kept up to date with the government

database, containing the information about the pharmaceutical products of the Spanish pharmaceutical market.

The resultant architecture of the ontology network will consist of the reference ontology as the core component of the network, supported by general ontologies (time, location…) and connected via mappings to medical classification ontologies (ATC) and a set of local ontologies which describe the pharmaceutical product information in its own way of knowledge. This architecture is described in section 2.2.5

In broad terms, the proposed methodology, depicted in section 2.1.1, describes different activities and tasks identified in the ontology lifecycle development and each simple scenario is an instance of the needed activities needed in each specific scenario. Following the methodology depicted in D531 suggested by WP5, the building process carried out for building the Nomenclature ontology network consists of:

1. **Knowledge Acquisition and Ontology Elicitation**: the pharmaceutical domain was specified and studied in D8.1.1. This study, based on different interviews with domain experts and documentation, analyzes the Spanish pharmaceutical sector and problems of integration between the different actors when exchange information about drugs.

2. **An Ontology Specification:** via a list of Competency Questions the necessities and requirements that the Nomenclature Ontology network has to satisfy are pinpointed. Also, the intended uses and users of the ontology network are identified. The list of competency questions has been included in section 2.3.

3. **Selecting the standards that cover most of the identified necessities**. Pharmaceutical classification systems and different thesaurus, taxonomies and vocabularies are identified and reviewed in the inventory of resources of the case study. First, as is described before, the local source of knowledge is analysed, trying to find reliable sources as the government and official pharmaceutical entities in Spain. These resources are the main databases (Digitalis, Integra, BOTPlus). Next, we find out resources suggested by pharmaceutical professionals as online vademecum (Vademecum) or Catalonian source... Then, medical vocabularies, thesaurus and taxonomies (ATC, Snomed, UMLS...) from international bodies related with pharmacy are analyzed and studied trying to find how the pharmaceutical domain is described in order to find connections between them and the Spanish models. More detailed analysis of the resources is described in [1] and section 3.

4. **Semantic enrichment of the standards**. The standards and medical vocabularies that are selected to be reused in the scenario of the case study are not ontologies resources. So, these resources need a **Ontology Reengineering** and **Ontology Enrichment** to OWL ontologies before reuse them in the model

5. **Evaluating Ontology Content**. Based on the necessities of each general domain recognized in the CQ, after an **Ontology Search** in order to find candidate ontologies about general domain as Time, Measures, and Geography. This ontologies are evaluated trying to find which one covers the ontology requirements in order to match which ontology should be selected and reused (**Ontology Reuse**) in the Nomenclature ontology network.

6. **Ontology Conceptualization**, **Ontology Formalization**, **Ontology Integration** and **Ontology Evaluation** of the Nomenclature ontology network. Previously, an ontology scheduling is prepared in order to check the tasks and milestones of the ontology network.

7. **Ontology Implementation** of the Nomenclature ontology network. The language selected for describe the resources is OWL. After, some activities of **Knowledge Acquisition for Ontologies,** in this case study, by means of **Ontology Population** in some ontologies with data extracted from the resources.

8. **Maintenance**. The Nomenclature ontology network and other ontologies that are going to be supplied by the Semantic Nomenclature case study, should be maintained and supported in future months. These ontologies should evolve according to the changes and suggestions given from Spanish pharmaceutical sector and professionals in the future. Some activities related to this stage are **Ontology Documentation, Ontology Configuration Management, Ontology Assessment** and **Ontology Verification & Validation**

9. Other activity related with the ontology lifecycle and development of the Nomenclature ontology network is the **Ontology Localization**. The Nomenclature ontology network is located in Spain but it will be provided in English. The main motivations are the internationalization of the ontology and link with the main medical and pharmaceutical vocabularies of the world.
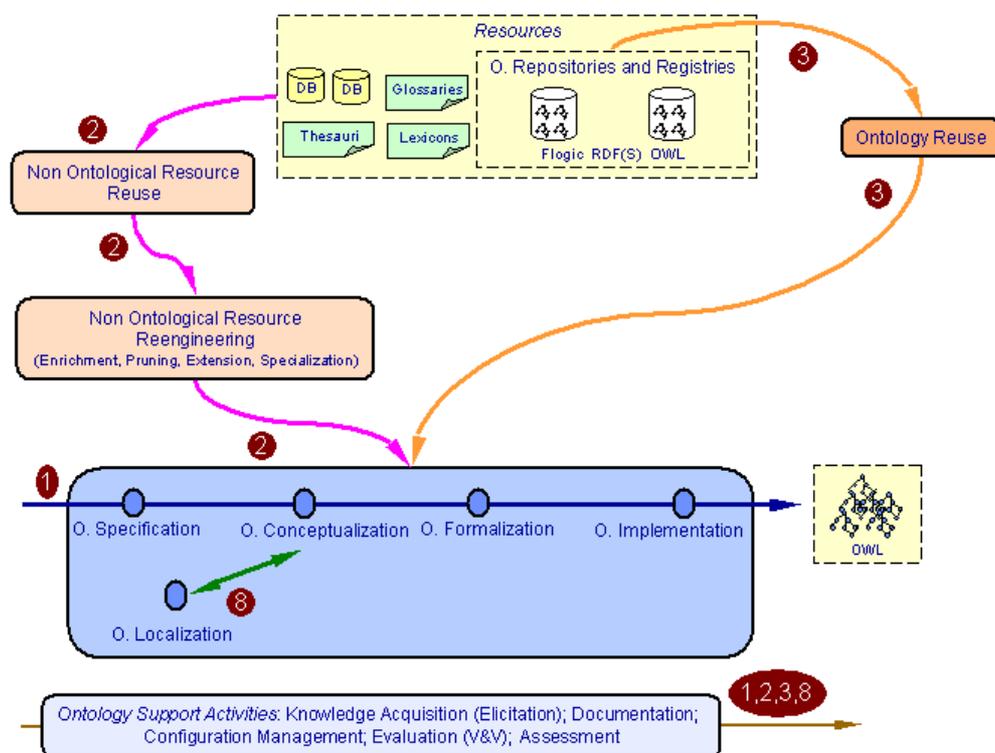
Figure 5: Semantic Nomenclature ontology activities

2.3.4. Selected Activities in Semantic Nomenclature

In the ANNEX III: Activities used in the development of the Nomenclature ontology network is included the table of 'Required-If Applicable' activities provided by WP5 in order to speed up the Nomenclature ontology network development by ontology experts. This table gather information about the required activities that should be carried out when developing the Nomenclature network ontology and the optional activities (or if-applicable) during the Nomenclature network ontology development. Not all the required activities are selected, such as Ontology control or Ontology Management Configuration, because not all the versions of ontologies, documentation have an intensive version control activities.

2.3.5. Competency Questions

The ontology specification activity, as is described in [14] , *is a collection of requirements that the ontology should fulfil, e.g. reasons to build the ontology, target group, intended uses, possibly reached through a consensus process.*

Before identifying the ontology requirements applying the competency questions technique [15] , section 2.3.3.1 and section 2.3.3.2 identifies the intended uses and the target groups of the Nomenclature Ontology network.

2.3.5.1.   Intended uses of the ontology

The development of the Reference Ontology and the Nomenclature network ontology is motivated by scenarios presented to the end-user application that will use the ontology network. Such scenarios together with the ontology's requirements are described in the deliverable [12] & [12] . The ontology network should satisfy these requirements after being formally implemented and should provide a consensual knowledge of the domain and solve the lack of communication between stakeholders in the pharmaceutical sector. The purpose of the Nomenclature Network Ontology is to provide a complete reference model about all the knowledge around the pharmaceutical products based on the main pharmaceutical classification and models used in the sector.

2.3.5.2.   Intended users of the ontology

The analysis of the scenarios and of the pharmaceutical sector described in deliverables [12] and [12] allows us to identify the different intended users of the ontology:

U1: Pharmacist   → Navigation across the ontology searching for drugs information

U2: GSCoP technician   → Searching for more information or relations about a given drug, principle, etc.

U3: GSCoP technician   → Updating and extracting the latest information in the BOTPlus database

U4: Spanish Government   → Analyzing the situation of the information about drugs or updating content.

2.3.5.3.   Competency Questions

According to the intended use of the ontology and the scenarios described for the Semantic Nomenclature application in [12] , the network of ontologies should satisfy these requirements. The competency questions are enumerated and are grouped into different concept domain: pharmaceutical product, laboratory and active ingredient. The complete list of the competency questions is included in ANNEX IV: Semantic Nomenclature Competency Questions.

Most of the specific competency questions defined in the Semantic Nomenclature are related with the Pharmaceutical Product, which is the most relevant concept in the domain. The CQs tries to extract information about the characteristics of a pharmaceutical product, such as time information (registration date, withdrawal date…), pharmacological information (dosage, route of administration…) or administrative information (national code identifier, price…). Also, the pharmaceutical product is related with other concepts like active ingredient, laboratory… The correspondent CQs are from CQ1 to CQ29. Some examples are:

> CQ1.   What is the drug commercial name?
>
> CQ13. Which is the drug composition?
>
> CQ17.   Which route of administration has the drug?

CQ18.   What is the drug pharmaceutical form?

Also, other specific competency questions related with the Laboratory, one of the actors of the pharmaceutical domain and closely related with the pharmaceutical product. The correspondent CQs are from CQ30 to CQ34, some example is:

CQ31.   Where is located the laboratory?

Several specific competency questions are related with active ingredient, which is the main substance of the pharmaceutical products. Moreover, there are some classifications of pharmaceutical products based on their active ingredient. The correspondent CQs are from CQ34 to CQ45, some example is

CQ34.   What is the national code of the active ingredient?

CQ36.   What is the ATC code of the active ingredient?

CQ37.   What is the WHO therapeutical subgroup of the active ingredient?

After this description of specific competency questions, could be composed into more general questions that are answered by composing answers associated to the specific competency questions. Composed competency questions that use a pharmacist for obtaining information about a drug are already defined in the Semantic Nomenclature list from CQ46 to CQ 49. As an example:

CQ46. Given the information of a drug, (name, national code, price…), has the nomenclature another similar drug with a lower price?

Furthermore, composed competency questions that use a GSCoP technician for obtaining information about new drugs are included in the list from CQ50 *to* CQ61. Some examples are:

CQ50.   Which are the latest drugs approved by the government?

CQ55.   Given a time interval (one week, one month…), which are the latest approved drugs?

From the competency questions, we extract the terminology (also known as predicates) and objects in the universe of the discourse (instances). The terminology will be formally represented in the ontology by means of concepts, attributes and relations. A list of these identified terms, which are grouped by the most relevant terms and concepts extracted from the CQs, is described in ANNEX V: Semantic Nomenclature Terminology, Glossary and Objects in the Universe of Discourse.

Enclosed with the terminology, a glossary describes and defines in natural language the terms identified. This glossary facilitates the study of pharmaceutical knowledge and terms for ontology experts.

Moreover, ANNEX V: Semantic Nomenclature Terminology, Glossary and Objects in the Universe of Discourse includes objects in the universe of the discourse, which are instances of the terms identified in the terminology: pharmaceutical products, laboratories, active ingredients, pharmaceutical form… These objects are useful in the selection of the standards to be reused when building the Nomenclature ontology network.

2.3.6. Architecture of the Semantic Nomenclature network of ontologies

It has conclusively been shown that the Spanish pharmaceutical sector has a lack of a reference ontology or description about all the knowledge around the pharmaceutical products. One of

objectives of this case study is covering this lack, providing a new reference ontology model based on the main schemas of the pharmaceutical databases in Spain and ontology requirements, and linking this reference ontology with the main medical vocabularies used around the world.

According to the characteristics of the scenario and the recommendations of the methodology, the use of a network of ontologies for representing the reference ontology in the pharmaceutical sector seems to be a good solution. As this scenario is described, semantic integration by means of a single, globally semantic model is too expensive to be adopted as a solution for this scenario. In contrast, a solution for the semantic integration based on a network of contextualized ontologies provides more facilities for maintaining ontologies locally consistent and easier to manage.

The Nomenclature Ontology Network is organized in three levels: the Pharmaceutical domain ontologies, the Application domain ontologies and the General ontologies. Figure 6 shows the levels of the Nomenclature Ontology Network based on reusability and usability of the levels.
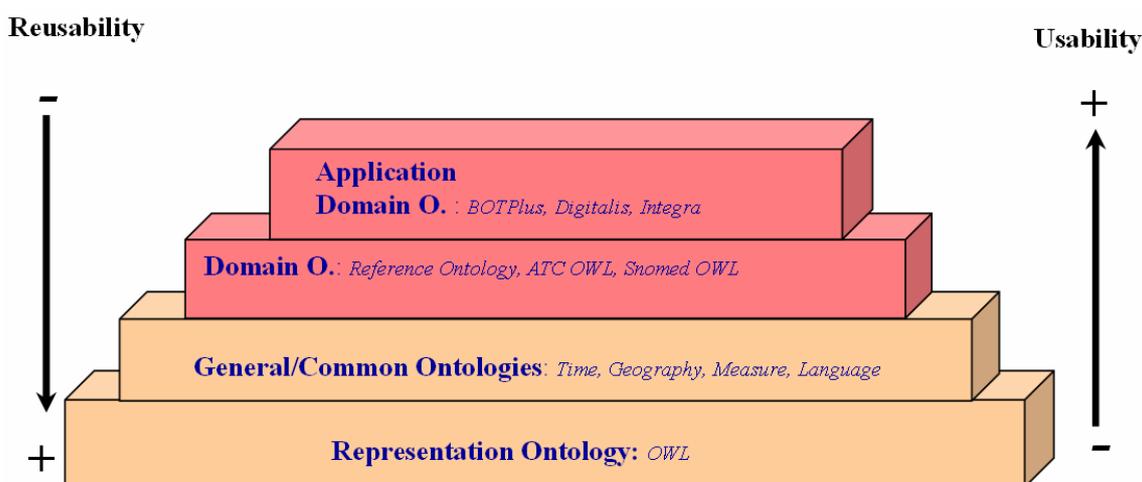
Figure 6: Modular approach of the Nomenclature Ontology Network

In the domain ontology level the ontologies or ontology modules which define several notions and concepts of the pharmaceutical domain that are substantial in the sector are included. The Pharmaceutical Reference Ontology is a compilation of the main terms and objects related with pharmaceutical products and the general aspects of them. Also, in this level ontologies are included, which provide a classification or vocabulary of these pharmaceutical terms, in this case, the ATC classification (because of the fact that is the WHO recommendation and is followed by the pharmaceutical experts in Spain and Europe) or the Snomed vocabulary.

The common ontology level groups the ontologies needed in the ontology network for describing any sort of real world objects and things, which could be of interest in some areas of discourse. Here are included ontologies as Time ontology, Geography ontology, Units ontology…

At a top level, the application domain ontologies are grouped. These ontologies are specialized in representing the knowledge of the real-world resources, in other words, they are the ontology models of the main databases which contain the information about the pharmaceutical products available in the Spanish market, as Digitalis or BOTPlus.

Figure 7 shows the appearance of the Nomenclature Ontology Network. The pharmaceutical reference ontology model is enriched with the general ontologies (Time, Location, Units…), and connected via mappings with the ontology models of the application level, as Digitalis or BOTPlus. In addition, the pharmaceutical reference ontology is related with the most important classification system ontologies as the ATC or Snomed. Finally, in new iterations of the lifecycle model, the

pharmaceutical reference ontology should integrate new resources or ontologies related with the main medical vocabularies used in the world
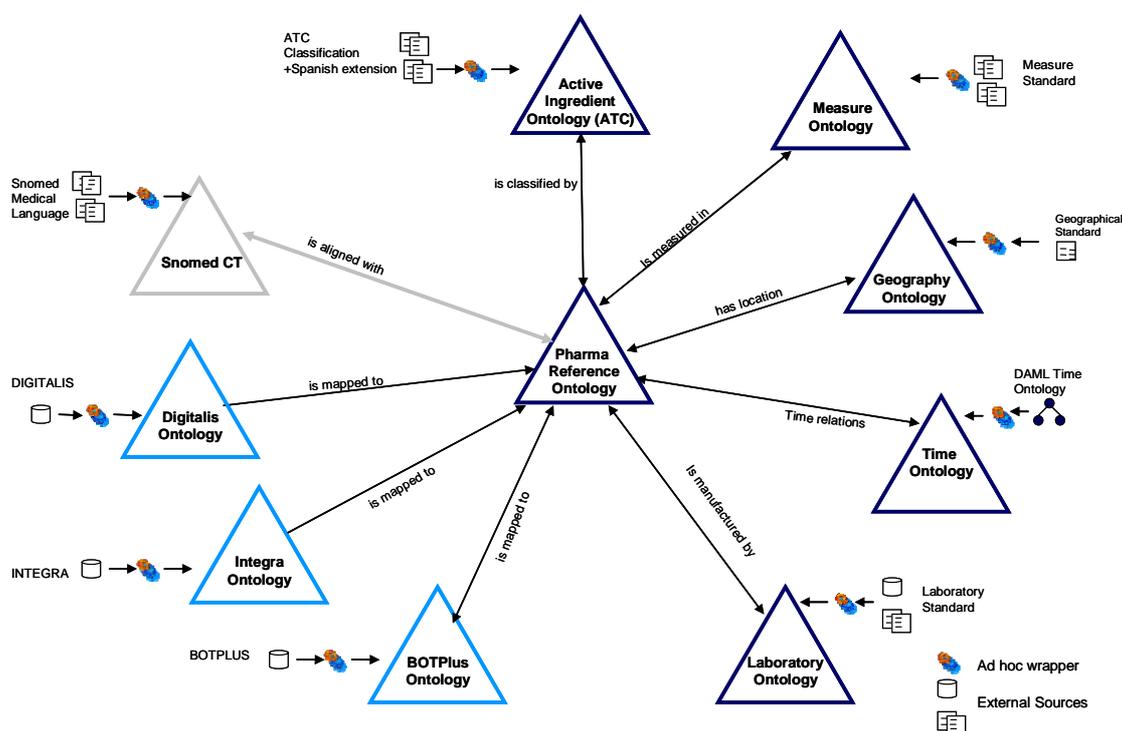
Figure 7: Nomenclature Ontology Network

Furthermore, in new iterations of the lifecycle model, the Nomenclature Ontology Network should integrate new resources or ontologies that could appear related with medical vocabularies used in the world. These ontologies may come from the current stakeholders (as ontologies of laboratory products) or external ones (ontologies from other countries or similar domains).

This ontology network outlined in the Semantic Nomenclature case study is aligned with the goals and expectations extracted from the case study scenarios depicted in [12] : integration of existing pharmaceutical resources and semi-automatic update of the BOTPlus information. These two scenarios summarize what the Semantic nomenclature tries to achieve. The ontology network architecture facilitates the aggregation of drug-related information in a semantic way because the reference ontology is mapped and related with different pharmaceutical ontologies at different levels. Also, the update of the BOTPlus database is a perfect scenario to show the possible business impact of networked ontologies, because the pharmaceutical product information gathered in the networked ontologies give an added value to the GSCoP in order to improve their commercial database reducing their effort and complementing typical pharmaceutical compendium characteristics by giving flexible, extensible and reliable information about drugs to the users of the Pharmaceutical domain

3. Inventory and Analysis of Knowledge Resources

In this section we describe the different knowledge resources related to the invoicing and semantic Nomenclature use case. These resources varies from ontological resources to non ontological including descriptions of technologies, existing ontologies (upper level ontologies, medical ontologies, time ontologies, etc) or concrete vocabularies related to the use cases.

3.1. Resources Analyzed in the Invoicing Case Study

3.1.1. Ontological Resources

The ontological resources analyzed in the invoicing case study are divided in four different types. The first type is related to the upper level ontologies that have been considered to be the basis of the invoice reference ontology. We have analysed SUMO, DOLCE and OpenCyc.

3.1.1.1. Upper level ontologies

SUMO (Suggested Upper Merged Ontology)

The **Suggested Upper Merged Ontology** or **SUMO** is an upper ontology intended as a foundation ontology for a variety of computer information processing systems. It was originally developed by the Teknowledge Corporation and now is maintained by Articulate Software.

SUMO originally concerned itself with meta-level concepts (general entities that do not belong to a specific problem domain), and thereby would lead naturally to a categorization scheme for encyclopaedias. It has now been considerably expanded to include a mid-level ontology and dozens of domain ontologies.

SUMO was first released in December 2000. It defines a hierarchy of *SUMO classes* and related rules and relationships. These are formulated in a version of the language SUO-KIF which has a LISP-like syntax. A mapping from WordNet synsets to SUMO has also been defined.

SUMO is organized for interoperability of automated reasoning engines. To maximize compatibility, schema designers can try to assure that their naming conventions use the same meanings as SUMO for identical words, (e.g.: agent, process). SUMO has an associated open source Sigma_knowledge_engineering_environment.

DOLCE

The Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) is the first module of the WONDERWEB foundational ontologies library. As implied by its acronym, DOLCE has a clear *cognitive bias*, in that it aims at capturing the ontological categories underlying natural language and human commonsense. DOLCE, however, does not commit to a strictly referential list metaphysics related to the intrinsic nature of the world. Rather, the categories it introduces are thought of as cognitive artifacts, which are ultimately depending on human perception, cultural imprints and social conventions. In this sense, they intend to be just *descriptive* (vs. *prescriptive*) notions that assist in making already formed conceptualizations explicit. DOLCE is an ontology of particulars, in the sense that its domain of discourse is restricted to them. Of course, universals are used to organize and characterize the particulars, but they are not themselves subject to being organized and characterized (e.g., by means of metaproperties). DnS (Descriptions and Situations) is a *constructivist* ontology that pushes DOLCE's descriptive stance further. DnS does not put restrictions on the type of entities and relations that one may want to postulate, either as a domain specification, or as an upper ontology, and it allows for context-sensitive '*redescriptions*' of the

types and relations postulated by other given ontologies (or 'ground' vocabularies). The current OWL encoding of DnS assumes DOLCE as a ground top-level vocabulary. DnS and related modules also exploit 'Codeps' (Content Ontology Design Patterns), a newly created tool which provides a framework to annotate 'focused' fragments of a reference ontology (i.e., the parts of an ontology containing the types and relations that underlay 'expert reasoning' in given fields or communities). Both DOLCE and DnS are particularly devoted to the treatment of social entities, such as e.g. organizations, collectives, plans, norms, and information objects.

OpenCyc

The latest version of OpenCyc, 1.0, was released in July 2006. OpenCyc 1.0 includes the entire Cyc ontology containing hundreds of thousands of terms, along with millions of assertions relating the terms to each other. The knowledge base contains 47,000 concepts and 306,000 facts and can be browsed on the OpenCyc website. The first version of OpenCyc was released in May 2001 and contained only 6,000 concepts and 60,000 facts. The knowledge base is released under the Apache License. Cycorp has stated its intention to release OpenCyc under parallel, unrestricted licences to meet the needs of its users. The CycL and SubL interpreter (the program that allows you to browse and edit the database as well as to draw inferences) is released free of charge, but only as a binary, without source code. It is available for GNU/Linux and for Windows.

3.1.1.2.   Process ontologies

TOVE [3]

The TOVE project consists in building an ontology for defining processes. This ontology has as goals 1) to provide a common terminology for the enterprise that every application can jointly understand and use, 2) define precisely the semantics of each term, 3) implement in PROLOG the semantics of the previous terms and 4) define a symbology for depicting a term or concept constructed. TOVE is based in a representation of actions composed by activities, states, time and causality. Given a set of actions that occur in different points of time in the future what are the resources and properties of resources and activities at other points in time. In TOVE activities are defined as basic transformation actions in which processes and actions can be represented. States define the parameters that have to be true in order to for the activity to be performed. A caused state defines what is true of the world once the activity has been completed. Activities, states, and the representation of time (as a continuous line) is the core of TOVE.

The Enterprise Ontology [4]

The Enterprise Ontology was developed within the Enterprise Project, a collaborative effort to provide a framework for enterprise modelling. The Ontology was built to serve as a basis for this framework which includes methods and a computer tool set for enterprise modelling. The Enterprise Ontology it includes a wide variety of terms which are widely used for describing enterprises in general. The idea is to provide one set of terms and definitions which adequately and accurately covers the relevant concepts in the enterprise modelling domain. This can be used to resolve any misunderstandings where terms are used differently. The Enterprise Ontology is proposed as one such set of terms and definitions. The EO was intended to serve as a basis for the Enterprise Tool Set. Broadly, it is intended to help ensure effective interchange of information and knowledge between different users, tasks and systems.

GLIF [5]

The Guideline Interchange Format (GLIF) is a model for representation of sharable computer-interpretable guidelines. The current version of GLIF is GLIF3. GLIF3 enables encoding of a guideline at three levels: a conceptual flowchart, a computable specification that can be verified for logical consistency and completeness, and an implemental specification that is intended to be incorporated into particular institutional information systems. GLIF3 leverages standards being developed in Health Level 7 in order to allow integration of guidelines with clinical information systems. The GLIF3 specification consists of an extensible object-oriented model and a structured syntax based on the resource description framework (RDF). Empirical validation of the ability to generate appropriate recommendations using GLIF3 has been tested by executing encoded guidelines against actual patient data. GLIF3 is can be used for broader experimentation and prototype use that want to capture the logic of clinical guidelines, to implement them in clinical systems, and thereby to provide integrated decision support to assist clinicians.

Business Process Modelling Ontology (BPMO) [6]

The Business Process Modelling Ontology defines the link between processes and organisations. BPMO is the superset of the EPC [19] and BPMN[8] ontologies, by defining the common concepts of the two ontologies. Thus, the EPC and BPMN ontologies are specializations of the BPMO and therefore import the BPMO. Differences between elements in the EPC and BPMN ontologies will be in terms of: a) terminology; b) structural properties; c) behavioural semantics. The BPMO will identify elements from EPC and BPMN with common behavioural semantics and bring them together under the same terminology, also specifying their common structural properties, if there are. For example, a "Logical Connector" in EPC has similar operational semantics to a "Gateway" in BPMN, however, their specialization into "Branch connector" and "Merge connector" is done differently for EPC and BPMN, because of different structural restrictions (e.g. EPC imposes that a connector can only link a Function to an Event or an Event to a Function, a structural restriction which does not exist eBusiness ontologies in BPMN). Some of the common concepts likely to be included in the BPMO are, for example, Process, Organisational Unit, Resource, Data, Products and Services, which are likely to appear in other business process modelling notations as well. Each of these can, in turn, be modelled in detail in specific ontologies. To understand this aspect better, the ARIS House is a useful analogy, whereby knowledge about Data, Organisation, Processes, and Products/Services is separated in "views"; similarly, a SUPER ontology can be created for each view. However, we believe that modelling only the Organisation and Process concepts in detail are valuable for SUPER. The Process concept will be defined in the Upper Level Ontology (and a mapping to BPMO will be defined), while the Organisation concept in the Organisational Ontology, which will be imported by the BPMO. The BPMO may also be developed so that, additional mappings could later on be built to other popular business process modelling notations, such as the UML2 Activity Diagram. The BPMO is a current effort towards providing a description of the BPM domain which supports BPM languages in the next future done in the SUPER[9] project.

WSMO [26]

The Web Service Modeling Ontology (WSMO) provides a conceptual framework and a formal language for semantically describing all relevant aspects of Web services in order to facilitate the automation of discovering, combining and invoking electronic services over the Web. The Web Service Execution Environment (WSMX) [25] is execution environment for dynamic discovery, invocation and composition of WSMO services.

OWL-S [27]

---

[8] www.bpmn.org/

[9] http://www.ip-super.org/

OWL-S is an ontology built on top of Web Ontology Language (OWL) by the DARPA DAML program. It replaces the former DAML-S ontology. The OWL-S ontology is a set of ontologies build in OWL that enable the users and software agents to automatically discover, invoke, compose, and monitor Web resources offering services, under specified constraints. With these ontologies is possible to represent processes and all of its elements. The ontologies in OWL-S are process model ontology, profiles ontology and grounding ontology.

3.1.1.3.   eBusiness Ontologies

Ontolog Community[10]

ONTOLOG (a.k.a. "Ontolog Forum") is an open, international, virtual community of practice devoted to advancing the field of ontology, ontological engineering and semantic technology, and advocating their adoption into mainstream applications and international standards. Within this community the UBLOntology has been developed. This ontology is an extension of SUMO [7] with elements of UBL[11]. The UBLOntology extends SUMO by mapping the Core Components Types, which are the semantic base of UBL, to ontology concepts creating classes for the core components that are missing in SUMO and creating also classes for the elements and concepts from UBL.

3.1.1.4.   Time Ontologies

In [8] there is a description of different time ontologies and a process followed to select one of them in the scope of the Esperonto[12] project. The ontologies analysed in this paper are:

- the time ontology in Upper Cyc Ontology[13] base, and is implementation in KIF and XML;

- the Unrestricted Time[14] ontology, developed at the Italian National Research Council (CNR) and codified in Ontolingua language;

- the Simple Time[15] ontology, which has been considered for developing further time ontologies, like the Reusable Time and the Kestrel Institute Time, and is implemented in Ontolingua language;

- The Reusable Time [20] , a very detailed ontology time developed at Stanford University and implemented in the Ontolingua language;

- the Kestrel Institute Time[16] ontology, implemented in DAML+OIL;

- the SRI Time ontology[17] implemented in DAML+OIL by the SRI's Artificial Intelligence Center;

- the modelling of time in SUMO [7] , developed by the IEEE Standard Upper Ontology working group, and implemented in different languages;

- the DAML time[18] ontology, implemented inside the DAML group in KIF and OWL; and

---

[10] http://ontolog.cim3.net/

[11] www.**oasis**-open.org/committees/**ubl**/

[12] http://www.esperonto.net

[13] http://www.opencyc.org/

[14] http://www.loa-cnr.it/medicine/unrestricted-time/index.html

[15] http://www-ksl-svc.stanford.edu:5915/FRAME-EDITOR/UID-1594&sid=ANONYMOUS&user-id=ALIEN

[16] http://www.daml.org/ontologies/100

[17] www.ai.sri.com/daml/ontologies/time/Time.daml

- the AKT Time[19] ontology, developed inside the AKT initiative and implemented in OCML.

3.1.2. Non-Ontological Resources

In this section we describe the non ontological resources analyzed in the case study. These resources include a big set the technologies used by the companies in their invoicing processes. These technologies vary from languages for modelling processes to invoicing languages and standards.

3.1.2.1.   eBusiness Standards

Inside the eBusiness we briefly describe electronic invoicing languages like EDIFACT (based on Electronic Data Interchange), the Universal Business language (UBL), ebXML (predecessor of UBL), the Core Components (which define common terms in UBL and EDIFACT) or XBRL.

Electronic Data Interchange (EDI) and EDIFACT

Electronic Data Interchange (EDI) refers to the computer-to-computer exchange of business information using a standardized data format [9] . Standardized EDI messages are based on common business documents such as purchase orders, invoices and bills of lading and are sent from one computer application to another over telecommunications links without human intervention or interpretation[20]. EDI is a standard which defines a set of messages, with its own terminology and its own schema. The final users can define their own EDI messages based on this main standard. Examples new business languages created from EDI are X12 which is the EDI version use in the United States or UN/EDIFACT which is the standard for electronic business message interchange main used in Europe. The structure of EDI contains:

1. **A syntax and encoding scheme for messages which specifies the structure of data.** The data should be independent of systems, machine and media constraints and should allow for human interpretation of the data transferred. As well, the data elements or groupings which are part of standard messages should be independent of each other so that one part may be changed without affecting any other part.

2. **A data dictionary.** This component of EDI standards defines the standard business data elements, such as date, time, delivery address, and currency used to create messages.

3. **Combinations of data elements to be used for standard messages.** A paper invoice, for instance, normally consists of a header portion stating the name and address of the billing party, the name and address of the paying party, the date of the invoice, an account number, etc. There is then a detail portion which consists of a series of invoice lines, each giving details of a billed transaction such as date, order number, number of units, item number, item description, unit price, and total price. There may also be a summary portion which gives totals. Each of these sections has an equivalent in EDI format with data elements combined into "segments" and segments combined into "messages".

EDIFACT

---

[18] http://www.w3.org/2006/time

[19] http://www.aktors.org/ontology/support

[20] http://www.ifla.org/VI/5/reports/rep4/rep4.htm

UN/EDIFACT stands for the United Nations rules for the Electronic Data Interchange for Administration, Commerce and Transport. They are a set of international standards, directories and guidelines for the electronic interchange of structured data, and, in particular, relate to trade in goods and services between independent computerized information systems [10] .

Example of an EDIFACT message:

> UNH+EW45852899+INVOIC:D:93A:UN:EAN007'
>
> BGM+380+FA044'
>
> DTM+137:20040623:102'
>
> RFF+ON:9953128'
>
> NAD+SU+4300929900006::9++LLUCH TRANSPORTIR, S.A.::INSCRITA REG.MERCANTIL DE BARCELONA:, TOMO 4.098, LIBRO 3.437 DE LA SEC+CL / L?ATLANTIC 112+BARCELONA++08040'
>
> RFF+VA:A43009299'

ebXML

ebXML (electronic business XML initiative) [11] initially a common effort between the UN/CEFACT (United Nations Center for Trade Facilitation and Electronic Business) and OASIS (Organization for the Advancement of Structured Information Standards) and now endorsed by other important standards organizations. It is centred on B2B scenarios, it aims to define a global electronic marketplace where enterprises find one another and conduct business process collaborations and transactions. It also defines a set of specifications for enterprises to conduct electronic business over the Internet by establishing a common standard for business process specifications, business information modelling, business process collaborations, collaborative partnership profiles, agreements and messaging.

UBL

Universal Business Language (UBL) is a library of standard electronic XML business documents such as purchase orders and invoices. UBL was developed by an OASIS Technical Committee with participation from a variety of industry data standards organizations. UBL is designed to plug directly into existing business, legal, auditing, and records management practices. It is designed to eliminate the re-keying of data in existing fax- and paper-based business correspondence and provide an entry point into electronic commerce for small and medium-sized businesses.

UBL version 2.0 was approved as an OASIS Committee Specification in October 2006 and has been publicly released. UBL is owned by OASIS and is currently available to all, with no royalty fees. The UBL library of business documents is a well-developed mark-up language with validators, authoring software, parsers and generators.

UBL provides the following:

- A library of XML schemas for reusable data components such as "Address," "Item," and "Payment" — the common data elements of everyday business documents.

- A set of XML schemas for common business documents such as "Order," "Despatch Advice," and "Invoice" that are constructed from the UBL library components and can be used in generic procurement and transportation contexts.

A standard basis for XML business schemas provides the following advantages:

- Lower cost of integration, both among and within enterprises, through the reuse of common data structures.

- Lower cost of commercial software, because software written to process a given XML tag set is much easier to develop than software that can handle an unlimited number of tag sets.

- An easier learning curve, because users need master just a single library.

- Lower cost of entry and therefore quicker adoption by small and medium-size enterprises (SMEs).

- Standardized training, resulting in many skilled workers.

- A universally available pool of system integrators.

- Standardized, inexpensive data input and output tools.

- A standard target for inexpensive off-the-shelf business software.

Core Component Types

The ebXML Core Components Technical Specification (ISO 15000-5) is a system for expressing business information in a reusable yet flexible way. Core Components (CCs) are a *bottom-up* initiative, defining terms and concepts at the discrete level, independently of the documents in which they are to appear (the document layer is handled by UBL or similar specifications). UBL is the first fully-conformant implementation of CCTS.

A CC may be **atomic** (also known as basic) or **aggregate**. An example of an aggregate component is a postal address, which makes up a coherent, abstract concept; it is composed of several atomic components such as province and postal code. The core components have no business meaning by themselves, as a number has no meaning by itself, it is necessary to point that this number is a specific measure for example in order to have some meaning. It is necessary to have a business context in which the core components acquire this meaning. The core component types are used in Business Information Entities, which can be:

- Basic Information Entities: A Basic Information Entity is a singular concept that has a unique business semantic definition. A Basic Information Entity adds semantic meaning to a single datatype or a Core Component Type (CCT).

- Aggregate Information Entity: An Aggregate Information Entity contains two or more Basic Information Entities or Aggregate Information Entities that together form a single business concept (e.g. postal address). Each Aggregate Information Entity has its own business semantic definition.

The Core Components Technical Specification is designed according to the ISO/IEC 11179 specification for metadata registries (generally data dictionaries). The ISO 11179 specifications are extremely thorough and meticulous, and there is no doubt that UBL is founded on the most rigorous semantic basis one could expect. UBL is the first international standards body implementation of the ebXML Core Components Technical Specification (CCTS 2.01, aka ISO 15000-5). The UBL library consists of ebXML CCTS Business Information Entities (BIEs). UBL XML schemas are defined through the application of UBL Naming and Design Rules (NDRs) to an underlying data model mapped to the Core Component types. UBL is also currently working with UN/CEFACT to converge the UBL library with the emerging UN/CEFACT Core Component library.

XBRL

XBRL (**Extensible Business Reporting Language**) is an emerging XML-based standard to define and exchange business and financial performance information. The standard is governed by

a not-for-profit international consortium (XBRL International Incorporated) of approximately 450 organizations, including regulators, government agencies and software vendors.

XBRL International is supported by its jurisdictions - independent bodies, generally organised on a country-specific basis, that work to promote the adoption of XBRL and the development of taxonomies that define the information exchange requirements of their particular domains. Its adoption has been quicker in Europe and Asia than in the U.S.

XBRL is a standards-based way to communicate business and financial performance data. These communications are defined by metadata set out in taxonomies. Taxonomies capture the definition of individual reporting elements as well as the relationships between elements within a taxonomy and in other taxonomies.

XML Common Business Library (xCBL)

The XML Common Business Library (xCBL) is a set of XML building blocks and a document framework that allows the creation of robust, reusable, XML documents to facilitate global trading. It essentially serves as the "mother code," providing one language that all e-marketplace participants can understand. This interoperability allows businesses everywhere to easily exchange documents across multiple e-marketplaces, giving global access to buyers, suppliers, and providers of business services. xCBL provides a migration path from EDI-based commerce because of its origins in EDI semantics. xCBL will be able to support all essential documents and transactions for global e-commerce including multi-company supply chain automation, direct and indirect procurement, planning, auctions, and invoicing and payment in an international multi-currency environment. xCBL is the result of collaboration between Commerce One and XML standards bodies, e-commerce enterprises, and hardware and software vendors, as well as analysis of existing e-commerce standards including Electronic Data Interchange (EDI), RosettaNet, and Open Buying on the Internet (OBI). The last version of xCBL dates of June 2003.

3.1.2.2.   Process Standards

Business Process Execution Language (BPEL)

The Business Process Execution Language for Web Services (WS-BPEL, BPEL4WS or BPEL[21] for short) is the de facto standard for describing Web Service Flows. It enables orchestration of Web Services using their abstract interface definition, defined in the Web Service Description Language (WSDL) [3]. A process itself is also represented as a Web Service via a WSDL file. In BPEL there are two kinds of activities, basic activities and structured activities. The latter are used to define the control flow of the process. The most essential ones are: <sequence> enabling sequential execution of activities, <flow> enabling parallel execution, and <while> supporting loops. Basic activities can be grouped into two categories: interaction activities [4] and others. The interaction activities (<invoke>, <receive>, <reply> and <pick>) enable communication with a partner Web Service referencing the portType and operation to be used. Additionally these activities reference a <partnerLink> which specifies the role the partner service and the process itself plays. The type of the <partnerLink> is defined as a WSDL extension, the so called <partnerLinkType> that specifies one or two roles and the <portType> each role has to implement. All of these interaction activities have at least one variable, input and/or output variable depending on their usage.

In a BPEL process the dataflow is realized via access to globally shared data, i.e. data passing from one to another activity is achieved by accessing the variables defined in the surrounding scope. Note, that in BPEL there are no explicit constructs to define dataflow, i.e. dataflow is

---

[21] www.ibm.com/developerworks/library/ws-bpel/

implicit. The <assign> activity can be used to copy data from one variable to another. BPEL enables synchronous and asynchronous Web Service invocation. Synchronous invocation here means that the response message of the used Web Service is received by the invoking activity, i.e. the activity waits blocking for the response. In this case the <invoke> activity has one input and one output container. Asynchronous invocation means that an <invoke> activity does not take a response directly, but the process handles the response message in a <receive> activity later on. The <correlationSet > enables the correlation of messages to the corresponding process instance and the appropriate activity. The <correlationSet> is of utmost importance when using asynchronous invocation because by defining a <correlationSet> it is assured that the response message is associated to the right process instance.

XPDL[22]

The **XML Process Definition Language** (**XPDL**) is a format standardized by the Workflow Management Coalition (WfMC) to interchange Business Process definitions between different workflow products like modelling tools and workflow engines. XPDL defines a XML schema for specifying the declarative part of workflow.

XPDL is designed to exchange the process design, both the graphics and the semantics of a workflow business process. XPDL contains elements to hold the X and Y position of the activity nodes as well as the coordinates of points along the lines that link those nodes. This distinguishes XPDL from BPEL which is also a process definition format, but BPEL focuses exclusively on the executable aspects of the process. BPEL does not contain elements to represent the graphical aspects of a process diagram.

BPMN[23]

BPMN defines a Business Process Diagram (BPD), which is based on a flowcharting technique tailored for creating graphical models of business process operations. A Business Process Model, then, is a network of graphical objects, which are activities (i.e., work) and the flow controls that define their order of performance.

A BPD is made up of a set of graphical elements. These elements enable the easy development of simple diagrams that will look familiar to most business analysts (e.g., a flowchart diagram). The elements were chosen to be distinguishable from each other and to utilize shapes that are familiar to most modellers. For example, activities are rectangles and decisions are diamonds. It should be emphasized that one of the drivers for the development of BPMN is to create a simple mechanism for creating business process models, while at the same time being able to handle the complexity inherent to business processes. The approach taken to handle these two conflicting requirements was to organize the graphical aspects of the notation into specific categories. This provides a small set of notation categories so that the reader of a BPD can easily recognize the basic types of elements and understand the diagram. Within the basic categories of elements, additional variation and information can be added to support the requirements for complexity without dramatically changing the basic look-and-feel of the diagram. The four basic categories of elements are:

- Flow Objects: A BPD has a small set of (three) core element, which are the Flow Objects, so that modellers do not have to learn and recognize a large number of different shapes. The three Flow Objects are Event, Activity and Gateway.

---

[22] http://www.wfmc.org/standards/XPDL.htm

[23] http://www.bpmn.org/

- Connecting Objects: The Flow Objects are connected together in a diagram to create the basic skeletal structure of a business process. There are three Connecting Objects that provide this function. These connectors are sequence flow, message flow and association.

- Swim lanes: Many process modelling methodologies utilizes the concept of swim lanes as a mechanism to organize activities into separate visual categories in order to illustrate different functional capabilities or responsibilities. BPMN supports swim lanes with two main constructs. The two types of BPD swim lane objects are pool and lane.

- Artifacts: BPMN was designed to allow modellers and modelling tools some flexibility in extending the basic notation and in providing the ability to additional context appropriate to a specific modelling situation, such as for a vertical market (e.g., insurance or banking). Any number of Artifacts can be added to a diagram as appropriate for the context of the business processes being modelled. The current version of the BPMN specification pre-defines only three types of BPD Artefacts, which are data object, group and annotation.

3.2.   Resources analyzed in the Semantic Nomenclature

In this section a detailed inventory of resources related with the Semantic Nomenclature case study is presented. The inventory includes both ontological and non-ontological resources. This inventory tries to analyse and consider all the possible resources which are the grounding of the knowledge of the case study. Some of the selected resources are the core information of the pharmaceutical products used by the pharmacists and domain experts in the Spanish pharmaceutical sector.

In the pharmaceutical domain at ontological level, different medical thesaurus, classifications and languages used by the international health community are identified. UMLS (a controlled compendium of medical vocabularies), LOINC, HL7, NCI, ATC classification… are some examples of the several vocabularies that try to classify different medical topics.

Moreover, the information about the pharmaceutical products is stored in some distributed and heterogeneous repositories. These resources are databases provided by the Spanish government (Digitalis, Integra) or by the GSCoP (BOTPlus). These legacy systems are the main resources of the case study and their characteristics are described in section 3.2 following the resource template of the Annex 1. With this template are pinpointed general, data, technical, administrative and practical aspects of these resources.

Both types of resources are very relevant and useful for the case study and some of them should be reused and/or reengineered in the ontology lifecycle of the Reference ontology or added via mapping to the Semantic Nomenclature network of ontologies. So, in section 3.1 and 3.2 these resources are analysed taking into account the type of resource (ontological or not-ontological), distribution, license aspects, and the relevance for the Semantic Nomenclature case study.

3.2.1. Ontological resources

3.2.1.1.   Medical Vocabularies

Snomed CT

SNOMED Clinical Terms (SNOMED CT) is a dynamic, scientifically validated clinical health care terminology and infrastructure that makes health care knowledge more usable and accessible. The SNOMED CT Core terminology provides a common language that enables a consistent way of capturing, sharing and aggregating health data across specialties and sites of care. Among the applications for SNOMED CT are electronic medical records, ICU

monitoring, clinical decision support, medical research studies, clinical trials, computerized physician order entry, disease surveillance, image indexing and consumer health information services.

Type

Ontology

Distribution / License

Rich Release Format (RRF)

Copyright, some parts are downloadable or accessible in UMLS

URL

http://www.snomed.org

Selection & Relevance to Semantic Nomenclature Case Study

It is the most internationally accepted nomenclature. It comes in English and has Spanish version, although in Spanish has some restricted user rights. It would be an important part of the Nomenclature network in the future. The SNOMED Taxonomy/ontology contains a perfect match for a Nomenclature

UMLS

UMLSKS provides access to multiple knowledge sources in the medical domain (SNOMED included).

Type

Various

Distribution / License

Various

Free use but has some restrictions. The terms are available in http://wwwcf.nlm.nih.gov/umslicense/snomed/license.cfm

URL

http://umlsinfo.nlm.nih.gov/

Selection & Relevance to Semantic Nomenclature Case Study

It is not a single resource, but a compendium of several resources in the medical domain (there are Thesaurus, nomenclatures, etc). It provides an API to access directly to the resources. Take into account in the case study.

HL7

Health Level Seven, Inc. (HL7) is an all-volunteer, not-for-profit organization involved in development of international healthcare standards. HL7 is also used to refer to some of the specific standards created by the organization. HL7's primary mission is to create flexible, low-cost standards, guidelines, and methodologies to enable the exchange and interoperability of electronic health records. The Reference Information Model (RIM[24]) expresses the data content needed in a specific clinical or administrative context and provides an explicit representation of the semantic and lexical connections that exist between the information carried in the fields of HL7 messages

Type

Distribution / License

Various (XML)

---

[24] http://www.hl7.org/library/data-model/RIM/modelpage_mem.htm

URL

http://www.hl7.org/Library/data-model/

Selection & Relevance to Semantic Nomenclature Case Study

It is not a relevant resource in the case study at the moment, but in the future could be integrated in the Nomenclature Ontology network.

MESH

MeSH is the National Library of Medicine's (USA) controlled vocabulary thesaurus. It consists of sets of terms naming descriptors in a hierarchical structure that permits searching at various levels of specificity. MeSH descriptors are arranged in both an alphabetic and a hierarchical structure. At the most general levels of the hierarchical structure are very broad headings such as "Anatomy" or "Mental Disorders". More specific headings are found at more narrow levels of the eleven-level hierarchy, such as "Ankle" and "Conduct Disorder." There are 22,997 descriptors in MeSH. In addition to these headings, there are more than 151,000 headings called Supplementary Concept Records (formerly Supplementary Chemical Records) within a separate thesaurus. There are also thousands of cross-references that assist in finding the most appropriate MeSH Heading.

Type

Thesaurus

Distribution / License

Various (XML)

No license is required to obtain the data via FTP. The data are available to all requesters, both within and outside the United States. There is no charge for obtaining the file. Users are required to complete an online registration form before receiving the data.

URL

http://www.nlm.nih.gov/mesh/filelist.html

Selection & Relevance to Semantic Nomenclature Case Study

It is not a relevant resource in the case study in the first iterations of the ontology network lifecycle, but in the future should be considered in order to enrich the Nomenclature ontology network.

3.2.1.2.   Health Standards and classifications

In this section the two important classifications of drugs provided by non-profit organisations as WHO and EPhMRA are described. Both classifications provide guidelines and a hierarchy for classifying the pharmaceutical products, which their ingredients are chemical substances, according to their therapeutical indications.

ATC Classification

The Anatomical Therapeutic Chemical Classification System is used for the classification of drugs. It is controlled by the WHO Collaborating Centre for Drug Statistics Methodology, and was first published in 1976. Drugs are divided into different groups according to the organ or system on which they act and/or their therapeutic and chemical characteristics.

The ATC classification system was originally based on the same main principles as the Anatomical Classification (AC-system) developed by the European Pharmaceutical Market Research Association (EPhMRA) and the Pharmaceutical Business Intelligence and Research Group (PBIRG).

Medicinal products are classified according to the main therapeutic use of the main active ingredient, on the basic principle of only one ATC code for each pharmaceutical formulation (i.e. similar ingredients, strength and pharmaceutical form). The main pharmaceutical databases in Spain classify the products according to this standard.

The ATC code provided by this Anatomical, Therapeutic and Chemical classification system of active ingredients compiles information about the organ or system on which they act and the pharmacologic effects, therapeutic indications and the chemical structure of the drug.

The ATC code is classified in different levels:

- 1st level: Anatomical level: organ or system on which the drug acts.

- 2nd level: Therapeutic subgroup

- 3rd level: Pharmacologic or therapeutic subgroup

- 4th level: Chemical, pharmacologic or therapeutic subgroup

- 5th level: Active ingredient or Chemical substance

In example, Table 1 analyzes an ATC code of the Ibuprofen active ingredient (ATC Code: M01AE01) in order to show the signification of each one or their digits means and are organized

Substance	Anatomical subgroup (1 digit)	Therapeutic subgroup (2 digits)	Pharmacologic subgroup (1 digit)	Chemical subgroup (1 digit)	Active ingredient (2 digits)
Ibuprofen *M01AE01*	M *Muscular-Skeletal System*	M*01* *Antiinflammatory and Antirheumatic products*	M01*A* *Antiinflammatory and Antirheumatic products Non-Steroids,*	M01A*E* *Propionic acid derivatives*	M01AE*01* *Ibuprofen*

Table 1: Ibuprofen ATC code description

The ATC classification is searchable and free of charge via web25, but also, the database is available on subscription only, and provides the possibility to order the ATC classification index with DDDs in a XML file format.

EphMRA

The Anatomical Classification of Pharmaceutical Products has been developed and maintained by the European Pharmaceutical Marketing Research Association (EphMRA) and is therefore the intellectual property of this Association.

This system represents a subjective method of grouping certain pharmaceutical products. The products are classified according to their main therapeutic indication and each product is assigned

---

25 http://www.whocc.no/atcddd/

to one category. In the AC-system, categories are organized on a cascade of 4 levels where each sub-level gives additional details about its upper-level.

The EPhMRA classification system is used world-wide by IMS (Intercontinental Medical Statistics) in producing marketing research statistics for the pharmaceutical industry.

The EphMRA classification is distributed free of charge as PDF file in http://www.ephmra.org/main.asp?page=465 . This classification system is not relevant for the Semantic Nomenclature case study, because is not used by the main resources in the pharmaceutical sector.

3.2.1.3.   Other Ontological Resources

There are large communities of researchers working in the areas of healthcare informatics and life sciences. These communities make good use of the semantic technologies to manage their science research and it is common that some of their results are several ontologies related with health domain.

Some of these ontologies are related with medical terms, gene, body structures, diseases, anatomy, clinical activities… Also, top-level ontologies contain concepts related with the pharmaceutical domain. In table 2 some of the ontologies found in different repositories and related with the Semantic Nomenclature case study are analyzed.

NCI thesaurus is a terminology and biomedical ontology used in a growing number of NCI and other systems. It covers vocabulary for clinical care, translational and basic research, and public information and administrative activities. The NCI Thesaurus provides definitions, synonyms, and other information on nearly 10,000 cancers and related diseases, 8,000 single agents and combination therapies, and a wide range of other topics related to cancer and biomedical research. It is maintained by a multidisciplinary team of editors, who add about 900 new entries each month[26]

The GALEN ontology is a result from the OpenGALEN Foundation[27] (a non profit organisation). The main goal of the ontology is provide terminology and classifications related with the anatomy, surgical deeds, diseases, and their modifiers used in the definitions of surgical procedures. Also, the ontology provides a module for units concepts, very useful describing some characteristics of the pharmaceutical products.

Unified Medical Language System® UMLS[28] is a resource provided by the National Library of Medicine (NLM) from the USA. This system gives cross-references between more than thirty biomedical vocabularies and classifications. These cross-references are obtained from a lexical analysis of the medical terms. UMLS is organized in three sources: Metathesaurus, a specialist lexicon and a semantic network. The Metathesaurus provides a lot of concepts and their derivates from the different biomedical source vocabularies. The Specialist lexicon contains several lexical terms from the biomedical domain. The Semantic Network is in charge of a severe classification of the biomedical concepts represented in the Metathesaurus through some semantic types. In this ontology, the category concepts are related with the most relevant relations extracted from the biomedical domain: physically-related-to, spatially-relate-to, temporally-related-to, functionally-related-to and conceptually-related-to. When it is possible, these relations are appointed to the top concepts of the ontology, in order that their subconcepts in the hierarchy inherit the relation.

TOP-level ontologies, as COSMO ontology or OpenCYC ontology or the OWN ontology (Wordnet translation), describe different domains in the world trying to provide a human consensus of the

---

[26] http://www.nci.nih.gov/cancerinfo/terminologyresources

[27] http://www.opengalen.org/

[28] http://www.nlm.nih.gov/pubs/factsheets/umls.html

concepts. In this ontologies are described and classified top-level concepts of the pharmaceutical domain, as drug, chemical substance, active ingredient… Moreover, they provide descriptions and concepts related with geography, time, units… So, these ontologies enrich the network of the ontologies described in the Semantic Nomenclature case study.

Name	Type	Distribution	License	URL	Relevance to WP8
NCI	Thesaurus	OWL Full	Free	http://www.mindswap.org/2003/CancerOntology/nciOncology.owl	Medical terms, Large ontology
GALEN	Ontology	OWL, XML	Free	http://www.galen.org/	Medical terms. Models the active ingredients. Not very important for the case study
UMLSSN	Ontology	OWL	Free	http://swpatho.ag-nbi.de/owldata/swpatho1/umlssn.owl	UMLS Semantic Network.
COSMO	Ontology	OWL	Free	http://colab.cim3.net/file/work/SICoP/ontac/reference/COSMO-ontology/COSMOtopOnly475.owl	Top-level ontology
OpenCYC	Ontology	OWL	Free	http://www.opencyc.org	Upper ontology whose domain is all of human consensus reality. Models drugs and substances and contains instances of drugs.
OWN	Ontology	OWL	Free	http://www.loa-cnr.it/ontologies/OWN/OWN.owl	Wordnet translation. Not very important

Table 2: Other ontology resources table

Also, the National Center for Biomedical Ontology from the USA government supplies BioPortal. BioPortal provides access to the Open Biomedical Ontologies repository, a library of publicly accessible biomedical ontologies. It is free of charge and is accessible in http://www.bioontology.org/ncbo/faces/pages/ontology_list.xhtml

3.2.2. Non-ontological resources

In this section are described and analysed knowledge resources with non-ontological resources as databases, web pages...

DIGITALIS Database

Nomenclature Digitalis contains the list of the products of the pharmaceutical market (around 10,000 products). This Nomenclature is officially used in the invoicing of prescriptions and contains data such as the identification of the pharmaceutical product, prices, composition of the medicine, etc.

Resource: DIGITALIS db	
General aspects	
Name	Digitalis Database

Provenance	Ministry of Health Spain Government
Url	http://www.msc.es/profesionales/farmacia/pdf/NomenclatorDigitalis_2007_03.zip
Format	Access DB
How does this system relate to other systems?	This Nomenclature has the same information as the previous government invoicing nomenclature plus new 18 attributes
Is there overlap between this system and other systems in the inventory?	Integra, BOTPlus (public information about drugs)
Which kind of experts uses the system in their work?	Domain experts
Who are the typical users of the system and how do they use the system?	Pharmacists, GSCoP technicians, Laboratories, Government technicians
In what broad ways does it satisfy the needs of the case study?	Digitalis is one of the officials Nomenclature of the government. It contains information about new approved and modified drugs This nomenclature offers more variability than others, because one of its objectives is to incorporate new proposals of classification. In this sense, Digitalis would lead to ontology evolution.

Data aspects

How many drugs are described?	38913 products (May 2007)
What kind of information store?	Id (Name & national code, etc.), price, composition, dates, dispensation info
Which languages use for describing medicines?	Spanish
What kind of classification system?	ATC classification
Maintained by whom?	Spanish government (Instituto de Información Sanitaria) and Regional Government
How widely used?	Professionals in general
How updated?	Monthly (Internet)
Where does data come from?	Government (Laboratories provide info to government)
How is data collected?	Info collected via electronic form and drug analysis
How frequently?	Monthly
For how long?	-
Who stands behind the data validity?	Spanish government (AGEMED)
Is the data easily accessible/available?	Yes
What are the access rights?	Public access

Technical aspects

How can the data be accessed or exposed?	Users can download the database from the Ministry of Health website and access to the information via AccessDB forms.

Administrative aspects	
Who has authority for the system?	Spanish Government
Who can change the data?	Spanish Government
Who can modify the system?	Spanish Government

Table 3: Digitalis analysis

INTEGRA Database

Nomenclature Integra stores information about pharmaceutical products classified by the ATC system[29] with content identification, DDD[30] and administrative characteristics. Integra is not another medicine list, but a tool created to help pharmaceutical professionals in two ways:

– To accomplish studies of medicine consume

– To incorporate proposals of classification and/or pharmaceutical product's identification that can be used by the pharmaceutical professionals of any welfare level for the prescription of these products.

Resource: INTEGRA db	
General aspects	
Name	Integra Database
Provenance	Ministry of Health Spain Government
Url	http://www.msc.es/profesionales/farmacia/pdf/Integra_2007_02.zip
Format	Access DB
How does this system relate to other systems?	This Nomenclature has the same information as the previous government nomenclature but has some differences.
Is there overlap between this system and other systems in the inventory?	Digitalis, BOTPlus (public information about drugs for hospital use)
Which kind of experts uses the system in their work?	Domain experts, Hospitals
Who are the typical users of the system and how do they use the system?	Pharmacists, GSCoP technicians, Laboratories, Government technicians, Hospitals
In what broad ways does it satisfy the needs of the case study?	Integra is one of the official Nomenclatures provided by the Spanish government. It contains information about drugs and health material used in Spanish hospitals.
Data aspects	

---

[29] ATC system: *Anatomical, Therapeutic, Chemical classification system*

[30] DDD: *Defined Diary Doses*

How many drugs are described?	8917 products
What kind of information store?	Id information (Name & national code…) reference price, composition, dates, dispensation info, administrative info
Which languages use for describing medicines?	Spanish
What kind of classification system?	ATC classification
Maintained by whom?	Spanish government (Instituto de Información Sanitaria) and Regional Government
How widely used?	Professionals in general
How updated?	Monthly (some days after Digitalis)
Where does data come from?	Government (Laboratories provide info to government)
How is data collected?	Info collected from Digitalis DB
How frequently?	Monthly
For how long?	-
Who stands behind the data validity?	Spanish government (AGEMED)
Is the data easily accessible/available?	Yes
What are the access rights?	Public access
Technical aspects	
How can the data be accessed or exposed?	Users can download the database from the Ministry of Health website and access to the information via AccessDB forms.
Administrative aspects	
Who has authority for the system?	Spanish Government
Who can change the data?	Spanish Government
Who can modify the system?	Spanish Government

Table 4: Integra Analysis

BOTPlus Database

The General Spanish Council of Pharmacists provides to its members a software tool called BOTPlus, developed by ATOS. The database associated with BOTPlus contains homogenous and updated information about medicines and sanitary products. It is a reference to the drug information for professionals. It offers information on diseases, symptoms, epidemics, treatments, detection of problems related to the medicines, etc. The information is codified in different tables in an Access database, so it will be possible to register and share data between different pharmacists in Spain, on any customer who requires it, guaranteeing a common codified structure.

This tool stores all the pharmaceutical information in a huge MS Access database with several tables. In contrast with other tools, BOTPlus introduces the concept of navigation across the information versus the query point of view.

GSCoP manifested interest in following the NeOn case study, but they are not NeOn partners, so their final involvement is not granted

Resource: BOTPLUS db	
General aspects	
Name	BOTPlus
Provenance	General Spanish Council of Pharmacists GSCoP
Url	http://pfarmals.portalfarma.com/default.asp
Format	Access DB & web
How does this system relate to other systems?	This Nomenclature has the same information as government nomenclature plus more added information related with the medical and pharmaceutical domains (diseases, treatments, symptoms, etc.)
Is there overlap between this system and other systems in the inventory?	Digitalis, Integra (public information about drugs for hospital use)
Which kind of experts uses the system in their work?	Domain experts
Who are the typical users of the system and how do they use the system?	Pharmacists, GSCoP technicians, Hospitals
In what broad ways does it satisfy the needs of the case study?	BOTPlus is the main tool used by community pharmacists. It provides a not only a nomenclature, but huge amount of information about related drugs, treatments, incompatibilities, patients history, etc.
Data aspects	
How many drugs are described?	88371 products
What kind of information store?	Id information (Name & national code), reference price, composition, dates, dispensation info, administrative info, relations, precautions, indications, contraindications, activity, etc. Also information about diseases, treatments, symptoms, patients, etc.
Which languages use for describing medicines?	Spanish
What kind of classification system?	Owned classification
Maintained by whom?	GSCOP technicians (data level), Atos Origin (application level)
How widely used?	Pharmacists (professionals in general)
How updated?	Monthly (some days after Digitalis)
Where does data come from?	Government, Laboratories, external sources… Heterogeneous and unformatted sources
How is data collected?	Info collected manually
How frequently?	Monthly

For how long?	-
Who stands behind the data validity?	GSCoP technicians and domain experts
Is the data easily accessible/available?	Partially on internet
What are the access rights?	Partial public access. More detailed info is only accessible by registered users
Technical aspects	
How can the data be accessed or exposed?	Internet queries (public part) AccessDB provided by the GSCoP (complete application)
Administrative aspects	
Who has authority for the system?	GSCoP
Who can change the data?	GSCoP technicians
Who can modify the system?	GSCoP & Atos Origin

Table 5: BOTPlus analysis

CEDIMCAT

Pharmaceutical Nomenclature similar to the previous ones, but provided by the Catalonian Government. The purpose is provided information in Catalonian about pharmaceutical products.

Resource: CEDIMCAT	
General aspects	
Name	CEDIMCAT Centre d'Informació de Medicaments de Catalunya
Provenance	Generalitat Catalunya (Regional Government of Catalonia)
Url	http://www.cedimcat.info/html/ca/Du30/index.html
Format	Web
How does this system relate to other systems?	This Nomenclature has similar information as the government nomenclatures, but applied to the Catalonia's pharmacists
Is there overlap between this system and other systems in the inventory?	Digitalis, Integra
Which kind of experts uses the system in their work?	Domain experts
Who are the typical users of the system and how do they use the system?	Pharmacists, GSCoP technicians, Hospitals in Catalonia
In what broad ways does it satisfy the needs of the case study?	Cedimcat is a web that provides a lot of information about drugs in Catalonia, the information is provided in Catalan (multilingualism)
Data aspects	

How many drugs are described?	-
What kind of information store?	Id information, reference price, composition, dates, dispensation info, administrative info, etc.
Which languages use for describing medicines?	Catalan
What kind of classification system?	Owned classification
Maintained by whom?	Generalitat Catalunya
How widely used?	Catalonian Professionals in general
How updated?	Monthly (some days after Digitalis)
Where does data come from?	Spanish Government, Laboratories
How is data collected?	Info collected from Spanish Ministry of Health
How frequently?	Monthly
For how long?	-
Who stands behind the data validity?	Health Catalonian Council
Is the data easily accessible/available?	Yes
What are the access rights?	Public access
Technical aspects	
How can the data be accessed or exposed?	Web browser
Administrative aspects	
Who has authority for the system?	Health Catalonian Council
Who can change the data?	Health Catalonian Council
Who can modify the system?	Health Catalonian Council

Table 6: CEDIMCat analysis

3.2.2.1.   Inventory analysis and conclusions

In this section the inventory of resources available in the Semantic Nomenclature case study are depicted in more detail. These resources are the grounding of the knowledge of the case study. These resources are described in different format or have different nature; they are databases, thesaurus, vocabularies, classifications, ontologies… So, these resources provide knowledge and classifications in the pharmaceutical domain and the databases provide the data.

The performance of a good methodology for exploit these resources will provide us a network of ontologies for solving and cover the lacks of the pharmaceutical domain. This methodology should allow to reuse and reengineering different kind of resources and exploit them in the best way

according to the problems of the pharmaceutical domain and the semantic web application solution proposed.

4. Ontology Selection and Reuse

In this section we describe the resources we finally chose for use in the case study and the motivations for selecting them. We will mainly describe the ontologies used and why we selected these ontologies. The non-ontological resources that are now part of the implementation of the use case too.

4.1.   Ontology Selection and Reuse in the Invoicing Case Study

In this section we describe the ontologies that are part of the invoice reference ontology. The final invoice reference ontology is an ontology based in an upper level ontology (DOLCE) which contains concepts about several electronic invoicing technologies (UBL, EDIFACT and a proprietary XML format), process concepts (Enterprise Ontology, TOVE) and a specialization of the invoice reference ontology in a real invoice model, the PharmaInnova model.

4.1.1. Ontological resources selected

The list of the considered resources is in section 3 and herein we explain the motivations we had for selecting them.

4.1.1.1.   Selection of the Upper Level Ontology

The selection of the upper level ontology is based in two main reasons, first the concepts that better fit into the representation of the invoice and processes concepts and second the implementation of these upper level ontologies. We preferred a standard ontology language like OWL instead others.

The first upper level ontology analyzed is OpenCyc[31]. OpenCyc is implemented in a proprietary format. This ontology is offered to the users through a web server where the user can query the ontology, make assertions. This server can also be downloaded and installed in a local machine and offers the possibility of exporting some elements to OWL. The importing can be done to a subset of the concepts but not globally to the ontology. We discarded OpenCyc because the impossibility of exporting all of it to OWL. OpenCyc contains all the concepts needed by the invoice reference ontology but the impossibility of opening it with the NeOn toolkit decided us to discard it.

SUMO is the second upper level ontology analyzed. This ontology is under the GNU general public license and owned by IEEE. SUMO is implemented in KIF although there are versions in OWL, Loom and Protégé. Furthermore the project UBLOntology extends SUMO with UBL terminology and concepts and it is implemented in KIF and Protégé format. In the UBLOntology the concepts needed to define an invoice and other business related concepts have been added to SUMO. The addition of concepts has been done or by adding new classes to SUMO or by mapping the invoice concepts when necessary. The SUMO ontology contains almost all the concepts needed by the invoice reference ontology. The UBLOntology not only contains most of the concepts needed but contains one of the main business technologies we plan to add to the reference ontology. Therefore we discarded to extend the UBLOntology with new invoice technologies. We also discarded the SUMO upper level ontology due to the same reasons. It was not possible to get a full OWL version of SUMO from the original sources of the ontology.

---

[31] http://www.opencyc.org/

DOLCE is the third upper level ontology we took into account in the development process of the invoice reference ontology. DOLCE has been developed in OWL and contains most of the classes needed by the use case. DOLCE offers the needed extension point from which we can add invoicing elements, time elements and it provides concepts for extending the ontology with process and workflows relations. We finally chose DOLCE because it has all the elements we need.

4.1.1.2.   Selection of the Process Ontology

The selection of the process ontology has been based on the combination of [3] [4] and [6] The process ontology used adds elements of all these ontologies in order to create a process ontology that is able to represent different kind of workflows. We will use this ontology for representing the invoicing workflow and the products workflow.

4.1.1.3.   Selection of the Time Ontology

The time ontology used is the W3C time ontology and it has been imported into the invoice reference ontology. We based our decision of choosing the OWL time ontology instead of the other ontologies on the results obtained through the competency questions previously answered. The competency questions specify that in the invoicing domain (when do we have to pay? When will the goods arrive?) is needed to model time stamps, years, months, days, hours, minutes or relative time stamps (30 days from the reception of the materials). The ontologies that best fit with our necessities are the OWL time ontology and the SUMO time ontology. We finally selected the OWL time ontology because it covers all the needs of the case study, it is a standard of the W3C for representing time and rejected SUMO because we already chose an upper level ontology for representing our domain.

4.1.2. Non-Ontological resources selected

4.1.2.1.   Core Component Types

The OASIS[32] cover pages web page about the CCTs says "Using Core Components as part of the ebXML framework will help to ensure that two trading partners using different syntaxes [*e.g.*, XML and United Nations/EDI for Administration, Commerce, and Transport (UN/EDIFACT)] are using Business Semantics in the same way on condition that both syntaxes have been based on the same Core Components. This enables clean mapping between disparate message definitions across syntaxes, industry and regional boundaries." Therefore the CCTs initiative tries to solve the problem of interoperability between organizations when these organizations interchange business messages. The core component types provide a basic semantic standardisation. In the final invoice reference ontology the Core Component Types are included and the conceptualization of the technologies present in this ontology will include the core components as it is specified by the OASIS organization.

4.1.2.2.   Electronic Invoicing technologies

We selected UBL and EDIFACT as main invoicing technologies to be represented in the invoice reference ontology. Both technologies are internationally recognised standards. EDIFACT is the standard developed by the United Nations and also adopted by the International Organization for Standardization (ISO) as the ISO standard ISO 9735. Therefore we consider it mandatory to have the conceptualization of EDIFACT in the invoice reference ontology. UBL was developed by the OASIS Technical Committee with participation from a variety of industry data standards

---

[32] http://xml.coverpages.org/ni2002-12-19-a.html

organizations and based in XML. It is widely used by the industry and can easily be adopted due to the use of XML. We consider also mandatory to include UBL in the invoice reference ontology.

We discarded the other electronic invoice technologies because they are less relevant than the two previously mentioned technologies. XBRL is an emerging technology but it is still not as wide spread as the previous two meanwhile the last version of xCBL dates from 2003.

Regarding the process standards analysed we do not include any of them. We provide in the invoice reference ontology means for representing the necessary workflows but we consider that the languages for instantiating these workflows are out of the scope of this use case.

4.2. Ontology selection and reuse in the Semantic Nomenclature

In this section is presented how we have selected pharmaceutical and health standards and the most suitable ontologies or ontology modules from the inventory that cover the ontology requirements identified with the competency questions in the Semantic Nomenclature case study.

4.2.1. Selecting General Ontologies

We have grouped some competency questions related to different identified groups as Time, Measure and Location.

Also these ontologies selected to be reused (time, location, measure) could be reengineered in order to formalize, align and extend these ontologies with the requirements needed in the case study.

4.2.1.1. Time ontology

Using the competency questions, we identify different temporal properties needed in the case study in the next list:

CQ4.   What is the drug registration date?

CQ5.   What is the drug withdrawal date?

CQ28. What is the last modification date?

CQ48. Which are the latest drugs approved by the government?

CQ49. Which are the latest withdrawal drugs by the government?

CQ53. Given a time interval (one week, one month...), which are the latest approved drugs?

CQ54. Given a time interval (one week, month...), which are the latest active ingredients approved?

CQ55. Which are the latest alerts about drugs in the last month?

CQ56. Which are the modified leaflets in the last month?

CQ57. Which are the new leaflets in the last week? (PDF, HTML...)

A Time Ontology is an agreed time model implemented in a machine-readable language. There are several time ontologies to be used for describing date information. They are:

 – *The time ontology in Upper Cyc Ontology*[33], *which is included in the Cyc knowledge base* [16]

---

[33] http://www.cyc.com/cyc-2-1/cover.html

- *The Unrestricted Time ontology[34], developed at the Italian National Research Council (CNR).*

- *The Simple Time ontology[35], which has been considered for developing further time ontologies, like the Reusable Time and the Kestrel Institute Time.*

- *The Reusable Time [17] , a very detailed ontology time developed at Stanford University.*

- *The Kestrel Institute Time[36] ontology.*

- *The SRI Time ontology[37] developed by the SRI's Artificial Intelligence Center.*

- *The modelling of time in SUMO[38], developed by the IEEE Standard Upper Ontology working group.*

- *The DAML time ontology[39], implemented inside the DAML group.*

- *The AKT Time ontology[40], developed inside the AKT initiative.*

Table 7 analyzes different time-ontologies in order to select and re-use one in the reference ontology according to the temporal properties (detailed in the first column) extracted from the competency questions.

	Cyc's Upper Ontology	Unrestricted Time Ontology	Simple Time Ontology	Reusable Time Ontology	Kestrel Time Ontology	SRI's Time Ontology	SUMO Time Ontology	DAML Time Ontology	AKT Time Ontology
Time Points	X	X	X	X	X	X	X	**X**	X
Time Interval	X				X	X	X	**X**	X
Absolute and Relative Time			X	X				**X**	X
Relations between time intervals					X		X	**X**	
Different temporal granularities	X					X	X	**X**	X

Table 7: Selecting Time Ontology

---

[34] http://ontology.ip.rm.cnr.it/onto/ON9.3-OL-HTML/unrestricted-time/

[35] http://www-ksl-svc.stanford.edu/

[36] http://www.kestrel.edu/DAML/2000/12/TIME.daml

[37] http://www.ai.sri.com/daml/ontologies/sri-basic/1-0/Time.daml

[38] http://www.ontologyportal.org/

[39] http://cs.yale.edu/homes/dvm/daml/time-page.html

[40] http://dream.inf.ed.ac.uk/projects/dor/akt/akt.html

We use the criteria presented in [18] and we choose the DAML Time ontology to model the time in Nomenclature Network Ontology. DAML Time ontology is implemented in OWL[41].

4.2.1.2.   Location ontology

As in the time ontology, using the competency questions, we identify some geographical properties needed in the case study:

CQ30. Where is located the laboratory?

CQ60. Which is the price of the medicine in the X autonomous region?

There are some location ontologies in the ontology repositories to be used for describing geographical and location information. They are:

- − CYC as geographic Ontology
- − SUMO as geographic Ontology
- − WSML location Ontology[42]
- − DAML location Ontology[43]
- − Minimalistic Location Ontology[44]

	CYC Geographic Ontology	SUMO Geographic Ontology	WSML Location ontology	DAML Ontology	Minimalistic Location Ontology (Simile)
Country	X	X	X	X	**X**
State		X	X		**X**
City	X	X	X		**X**
Location	X		X	X	**X**
Address	X	X	X		**X**

Table 8: Selecting Location Ontology

They are two main candidates to be selected and reused in the ontology network: WSML Location Ontology and the Minimalistic Location Ontology. Either one like the other satisfies the ontology requirements detected in the competency questions of the Nomenclature case study. Also, both ontologies need ontology reengineering to OWL ontologies, so seems more easy reengineering the Minimalistic Location Ontology from the Simile Project because is in RDFS format.

4.2.1.3.   Unit Ontology

As in the previous ontologies (time, location), using the competency questions, we identify some measure and unit properties of the pharmaceutical products needed in the case study.

CQ13. Which is the drug composition?

---

[41] http://www.w3.org/TR/owl-time/

[42] http://www.wsmo.org/ontologies/location

[43] http://www.daml.org/experiment/ontology/location-ont

[44] http://simile.mit.edu/2005/05/ontologies/location

CQ33. Which is the unit content of the medicine?

There are not much Unit ontologies to be used for describing this type of information. After realize ontology search in different ontology repositories, the most relevant ontologies are:

- Simplified Galen Ontology[45]
- Units.owl[46] from the Vrije University of Brussels
- Units.owl[47] ontology described in the Semantic Web for Earth and Environmental Terminology (SWEET)
- Unit_ontology.owl[48] provided by the UMD Astronomy Information and Knowledge Group
- Unit.owl[49] developed by CoDAMoS (**Co**ntext-**D**riven **A**daptation of **Mo**bile **S**ervices) strategic basic research project.

	Galen Ontology	Units Vrije Ontology	SWEET Units Ontology	UMD Unit Ontology	CoDAMoS Units ontology
Milligrams	**X**			X	
Millilitres	**X**				
Grams	**X**	X	X	X	X
Temperature units	**X**	X		X	X
Dose units	**X**				

Table 9: Selecting Unit Ontology

They are not many ontologies dedicated to units of measure (distance, weight, distance, temperature…) but in the Galen ontology are detailed some concepts of different kind of units (Composite, Primitive). Galen should be reengineered in order to transform the conceptual model or extend (more types of units) into a new and more complete if it is necessary. Also, Galen is one of the main ontology models related with healthcare domain and could enrich the Nomenclature ontology network.

4.2.2. Selecting Medical and Pharmaceutical resources

Table 10 summarizes and matches terminology from Competency Questions against the standards and medical languages in order to select the most appropriate standards and taxonomies for reuse it in the pharmaceutical reference ontology. In the table are included non-ontological resources such as the main pharmaceutical products databases in Spain, the most important international health vocabularies, thesaurus, taxonomies and classifications recommended by the WHO.

---

[45] http://www.cs.man.ac.uk/~horrocks/nsd07/galen.owl

[46] http://ssel.vub.ac.be/viewcvs/viewcvs.py/*checkout*/PlatformKit/platformkit-java/Units.owl

[47] http://sweet.jpl.nasa.gov/ontology/units.owl

[48] http://archive.astro.umd.edu/ont/unit_ontology.owl

[49] http://archive.astro.umd.edu/ont/unit_ontology.owl

	Medical Languages			Thesaurus & Taxonomies		Classification Systems		Pharmaceutical DBs / Vademecum		
	UMLS	*SNOMED*	*HL7*	*NCI*	*MESH*	*ATC*	*EphMRA*	*Digitalis*	*Integra*	*BOTPlus*
Drug	X	X	X	X	X			**X**	**X**	**X**
Composition								**X**	**X**	**X**
Pharmaceutical Form		X						**X**	**X**	**X**
Laboratory								**X**	**X**	**X**
Active Ingredient		X				**X**	X	**X**	**X**	**X**
Clinical Aspects										**X**
Degree of coverage										**X**
Pharmaceutical Spanish sector needs						**X**		**X**	**X**	**X**

Table 10: Selecting Health and Pharmaceutical resources

According to the relevance for the case study and the Spanish pharmaceutical sector, the most important resources for develop the Nomenclature network ontology are the databases provided by the Spanish government and GSCoP and the ATC classification used for classify drugs and suggested by the WHO.

The relevance and importance of the databases is clear, because in their tables are detailed all the data and characteristics (codes, ingredients, name, price, DDD…) about all the pharmaceutical products, including the withdrawal products, authorised in Spain. Digitalis and Integra databases are provided by the government and are the first which gather the latest information about the products, and BOTPlus collects information about the different aspects of the pharmaceutical domain and is the database used by the pharmacists in Spain. These resources need a semantic enrichment in order to model their conceptual model into ontologies.

The relevance of the ATC classification for active ingredients is based on their use in the previous pharmaceutical databases for classify the products. Also, the ATC classification is provided and supported by the WHO and is widely adopted in Europe. These resource need semantic enrichment too, because is a non-ontological resource, they are

The other pharmaceutical resources, medical languages (UMLS, Snomed, HL7) vocabularies (MeSH), thesaurus (NCI), taxonomies and classifications (EphRMA) are resources that corresponds to the domain level of the ontology network, but they are not the priority in the case study due to these vocabularies are not used regularly by the Spanish pharmaceutical experts.

5. Networked Ontologies in the Pharmaceutical Case Studies

In this section we describe how the methodology specified in Section 2 is used in order to create the ontologies that are used in the case studies. We describe how the ontological and non-ontological resources (described in Section 3) are reused and integrated in the ontologies and how these resources help to fulfil the requirements specified and refined by the competency questions of Section 2.

5.1.   Networked Ontologies in the Invoice Case Studies

In this section we describe the steps that we followed in order to create the invoice reference ontology. These steps are the extension of DOLCE whenever it was necessary, mapping of concepts between the UBL ontology and DOLCE, addition of the necessary process concepts to DOLCE and finally the creation of the PharmaInnova ontology from the invoice reference ontology. In the figures below the classes highlighted with a green box are classes from DOLCE meanwhile the classes highlighted with a red box are invoice concepts added to DOLCE.

5.1.1.   Networked Ontologies in the Invoice Case Studies

The extension of DOLCE has been done in five different stages. During the first stage we extended DOLCE with the concepts and UBL terminology contained in the UBL ontology. We searched the common classes of both ontologies (ontology aligning) and extending from these classes the concepts missing. In the second stage we added the EDIFACT concepts and terminology to the invoice reference ontology. In the third stage we added the time ontology we previously selected and in the fourth stage we added the process concepts that are used in the invoicing process. Finally in the fifth stage we adapted the invoice reference ontology to the PharmaInnova invoice model.

5.1.1.1.   Core Component Types extension over DOLCE

The extension of DOLCE with the Core Components Types specification has been based in the previous extension that was done to SUMO by the UBLOntology project. This extension is described in the Ontolog Wiki[50]. The mapping consists in mapping the core components concepts to the concepts existing in SUMO. From each core component a concept in this ontology is identified and if it is not possible to identify any concept a new one is created. We follow the same approach in the extension of DOLCE for creating the invoice reference ontology. We try to identify each core component concept in DOLCE. If the concept exists we do nothing and if it does not exist we create a new one.

5.1.1.2.   Main extension point for invoicing technologies

The goal of the invoice reference ontology is to add as many invoice related technologies as there exist. In this deliverable we extend DOLCE with two widely extended electronic invoicing standards and with one proprietary solution based in XML. To facilitate the extension of the invoice reference ontology with more technologies we add a class from which all the specific concepts related to the specific technologies will be added. This class is "TransactionEntity" and the subclasses that contain   each   technology   are   "UBLTransactionEntity",   "ProprietaryTransactionEntity",

---

[50] http://ontolog.cim3.net/cgi-bin/wiki.pl?CctRepresentation

"EDIFACTTrransactionEntity", etc. The hierarchy of the "TransactionEntity" class is represented in Figure 8. We create "TransactionEntity" as subclass of "depiction". In DOLCE "depiction" is a realization of a representation. "TransactionEntity" contains the concrete representations of different invoice technologies and languages, therefore we classify this class as subclass of "depiction".
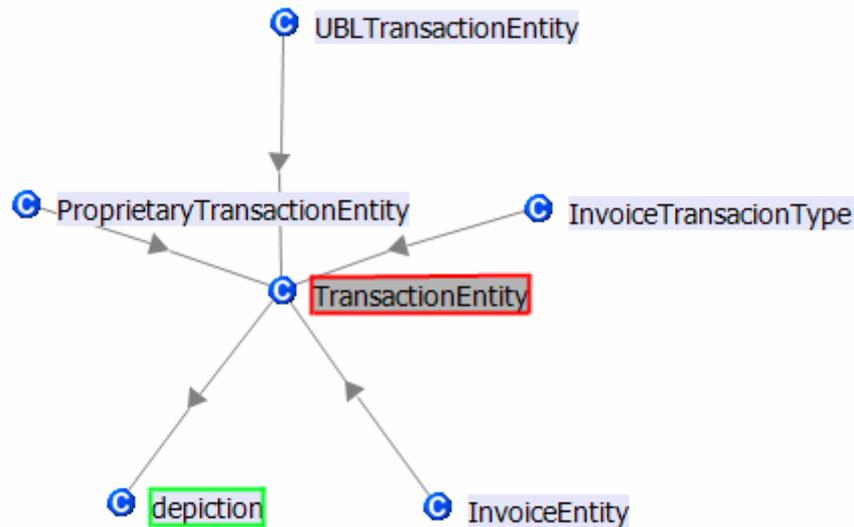

Figure 8: Extension point for invoicing technologies

5.1.1.3.   UBL extension over DOLCE

The extension of DOLCE by adding UBL invoice related concepts and UBL terminology has been based in the in the UBL ontology which is an extension of SUMO with UBL concepts and terminology. The UBL Ontology project shares some goals with this deliverable, among others the aim of representing UBL within an upper level ontology. Therefore we used the bases of this extension of SUMO adapting it for DOLCE.

The first extension of DOLCE was by adding the "Text-Abstract" class as a subclass of "information-object". The definition of "information-object" is "*Information objects are social objects. They are realized by some entity. They are ordered (expressed according to) by some system for information encoding. Consequently, they are dependent from an encoding as well as from a concrete realization. They can express a description (the ontological equivalent of a meaning/conceptualization), can be about any entity, and can be interpreted by an agent. From a communication perspective, an information object can play the role of "message". From a semiotic perspective, it plays the role of "expression".*" We created in DOLCE "Text-Abstract" as a subclass of "information-object". We consider "Text-Abstract" and all its subclasses as a social object with which the financial agents interact, the "Text-abstract" instances are realized by entities (emitters, receivers), they are expressed by a specific encoding (UBL, EDIFACT, CSV, etc.) and they express an invoice. "Text-Abstract" also plays the role of "message" between two organizations. The class "Text-Abstract" and its subclasses are represented in Figure 9.
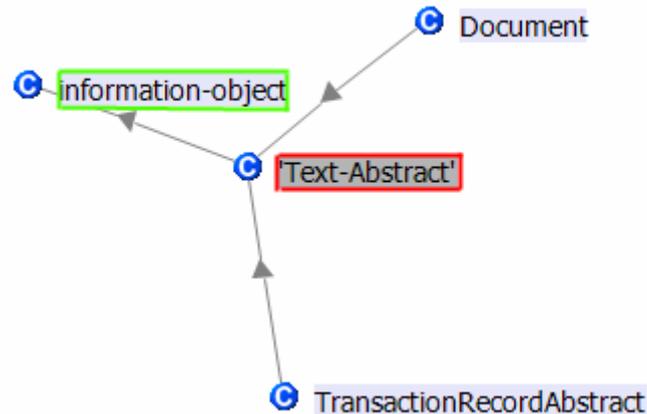
Figure 9: "Text-Abstract" hierarchy

"Text-Abstract" is the super class of "Document" and "TransactionRecordAbstract". These two subclasses are the core of the extension of DOLCE for the invoicing process. Defined in the UBLOntology a "TransactionRecordAbstract" is "*a conceptual object which contains the information regarding the whole or some part of a transaction event (or series of related events). It will have some form of physical representation, even if only in the mind of one of the participating agents*". This class contains the abstract representations of the elements that participate in an invoice. The particular realizations of these concepts are done by "TransactionEntity". "TransactionEntity" is one of the types of the UBL specification that represents all the elements that participate in a transaction. From this class we extend all the specific elements of the technologies that are part of the invoice reference ontology, including UBL, EDIFACT and the proprietary solutions of each partner. These elements are specific messages of EDIFACT or UBL.
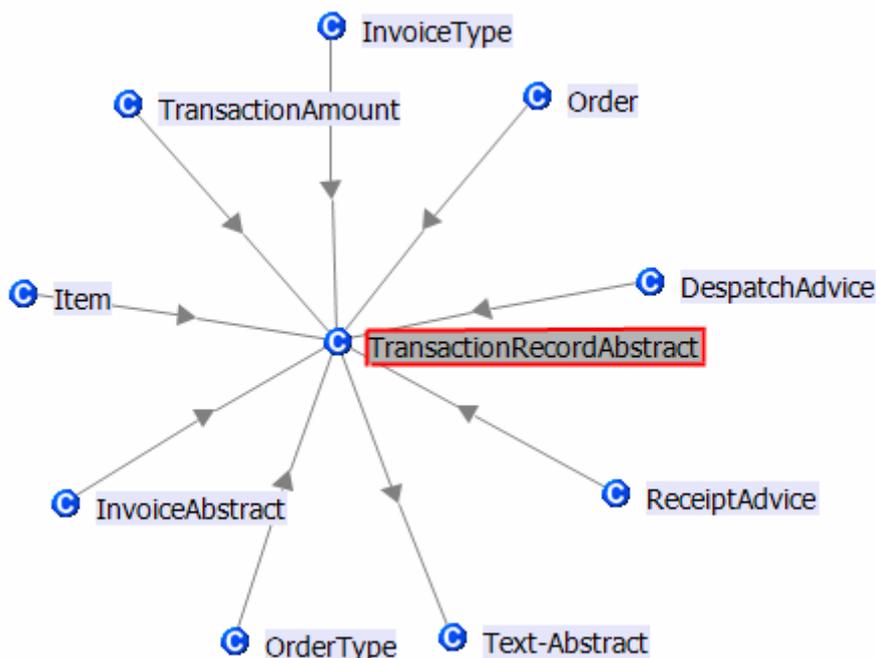
**Figure 10: TransactionrecordAbstract hierarchy**

We extend "non-agentive-social-object" by adding to it the subclass "Currency". It is imperative to have the representation of currency in an invoice. "non-agentive-social-object" is "*A social object*

that is not agentive in the sense of adopting a plan or being acted by some physical agent". A currency is always a social object which is used by agents. This hierarchy is represented in Figure 11.
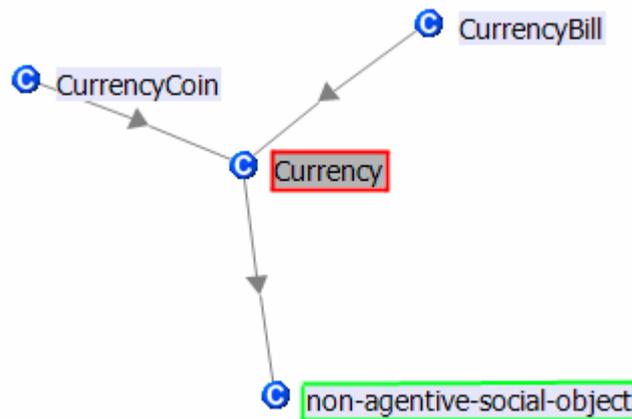

Figure 11: Currency hierarchy

In Figure 12 is represented the extension of "material-artifact" by adding to it the subclass "Product". "material-artifact" is "*a physical object that shows or is known to have an artifactual origin that counts in the tasks an ontology is supposed to support*". In our case a product fulfils this definition because product represents the physical object that is being sold or bought and it has an artifactual origin.
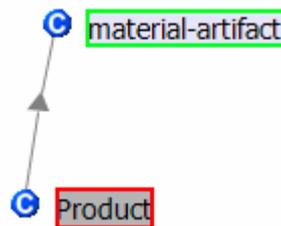

Figure 12: Product hierarchy

We extend "social-relationship" by adding to it the subclass "Transaction". This extension is shown in Figure 13. We create "Transaction" as a subclass of "social-relationship" due to its definition "*A social description defining roles for the interaction of rational agents*". We consider a transaction a social description in which two or more rational agents interact. The concept "TransactionAmount" is categorized as a subclass of "TransactionRecordAbstract" in a different manner than "Transaction" due to "TransactionAmount" is an amount not an interaction. "TransactionAmount" is also an entity that has part of a transaction event (as a subclass of "TransactionrecordAbstract"). "TransactionAmount" is a sum of money that is part of a transaction -- as the price of an individual item, the extension of a line item, a tax, a shipping charge, or whatever.
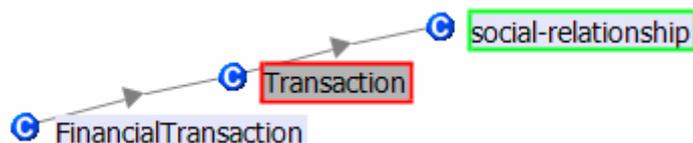

Figure 13: Transaction hierarchy

We add the class "PaymentTerms" is classified as a subclass of "plan". "PaymentTerms" is considered as a sequence-dependent specification like cooking recipes or scripts in the

UBLOntology. "Plan" in DOLCE is a method for executing or performing a procedure. It matches partially the definition but it is the best classification possible to get. The extension is shown in Figure 14.
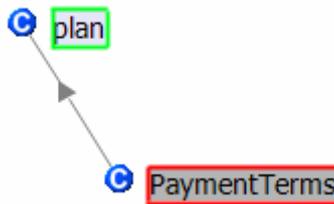

Figure 14: PaymentTerms hierarchy

We extended the class "description" with the subclass "InvoiceAttribute". The extension of "description" is shown in Figure 15. "InvoiceAttribute" is in UBL a business entity. This entity contains as subclasses dates, quantities and codes and are used in the different concepts of the technologies described in the ontology. We classified it as subclass of "description". "description" is "*A description is a social object which represents a conceptualization (e.g. a mental object or state), hence it is generically dependent on some agent and communicable. Descriptions define or use concepts or figures, are expressed by an information object and can be satisfied by situations.*" "InvoiceAttribute" matches with this definition due to them represent a conceptualization and are dependent from agents. Also they have to be satisfied by situations.

Figure 15: InvocieAttribute hierarchy

5.1.1.4.   EDIFACT extension in DOLCE

The extension of EDIFACT has been done based in the invoice message specified by EDIFACT. This message is divided in segments and groups. A segment is a set of primitive types that represent the header of the message, the information about the currency used, the terms of payments etc. The primitive types a for instance amount of the invoice. These primitive types are represented by the core components that have previously been represented in the invoice reference ontology. A group in the INVOICE message of EDFICAT is a set of segments. At the beginning of the INVOICE message there are several mandatory segments (header of the invoice, beginning of the message and date of the invoice). Once this header is ended a set of groups containing specific information of the type of goods is represented.

In the invoice reference ontology we represent first as a subclass of "diagrammatic-object" the segments that contain the primitive types of the INVOIC message of EDIFACT. The class we create is "EDIFACTSegment" and it is shown in Figure 16. We represent it as a subclass of "information-object". "information-object" is "*Information objects are social objects. They are realized by some entity. They are ordered (expressed according to) by some system for information encoding. Consequently, they are dependent from an encoding as well as from a concrete realization*". An EDIFACT segment is an information object that has to e realized in an invoice by a code. Every segment is composed by the core component types specified in the EDIFACT specification.
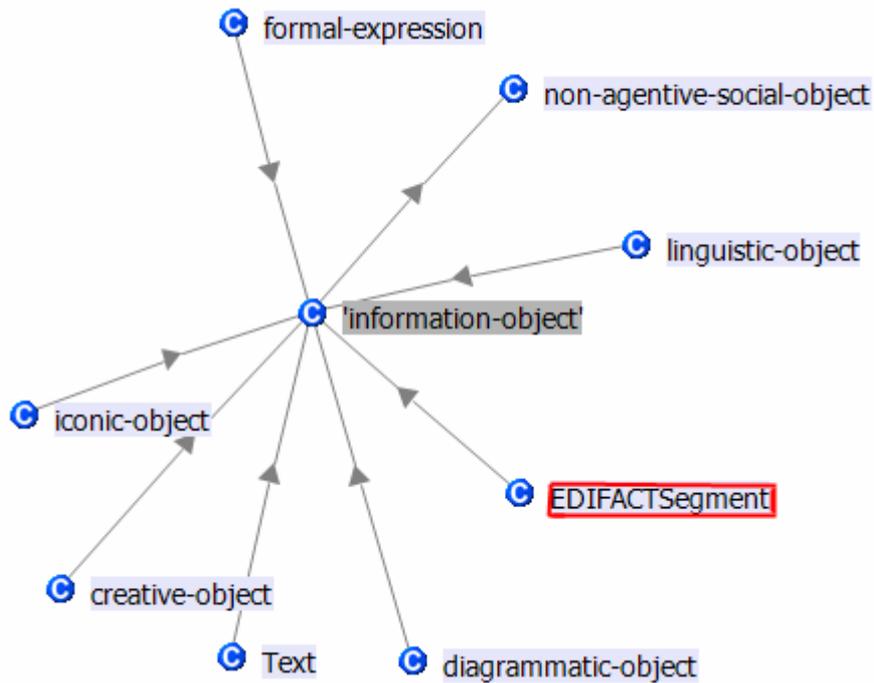

Figure 16: EDIFACT segment

We represent the groups of segments that compose the invoice message in EDIFACT as subclass of the main extension point for the invoice technologies, the class "TransactionEntity" (shown in Figure 17). The groups are subclasses of "EDIFACTSegmentGroup" and each group contains the relations specified by EDIFACT to the previously described segments.
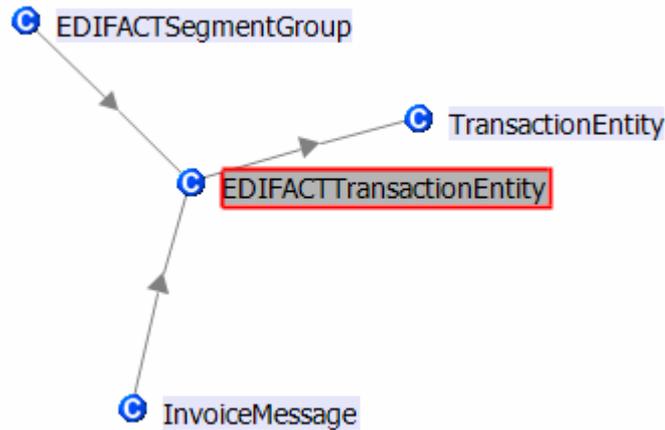
Figure 17: EDIFACT segment group representation

5.1.1.5.   Process ontology extension in DOLCE

The process ontology is divided in four main classes. Its hierarchy is shown in Figure 18. The main classes are explained next.
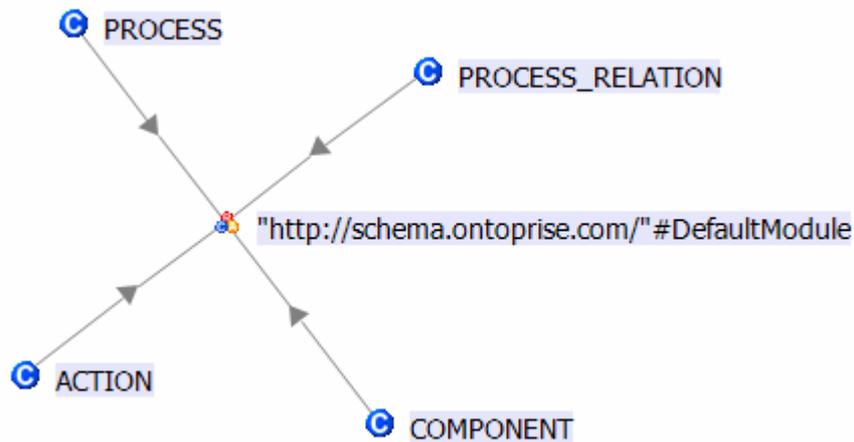
Figure 18: Process ontology

Components are the entities that can be used as resources within a process. A component can be an agent resource, an attribute, a bag or a tool. A process can use to perform its goal other agents as a resource to achieve its goal, tools used to perform an action, bags used to group other entities or attributes treated as resources used by agents. We extend DOLCE with the class Components and its subclasses as a subclass of "physical-realization". "Physical-realization" is *Any physical particular that realizes a non-physical endurant. Such physical particulars can be either physical endurants, physical qualities, physical regions, perdurants with at least one physical participant, or a situation with one physical entity in its setting. Ultimately, a physical realization depends on at least one physical endurant (each of the others physical entity types depend on a physical endurant to be considered as such).*" We consider that "Components" is a "physical-realization" because all are physical particulars as they are defined in "Physical-realization". The extension is shown in Figure 19.
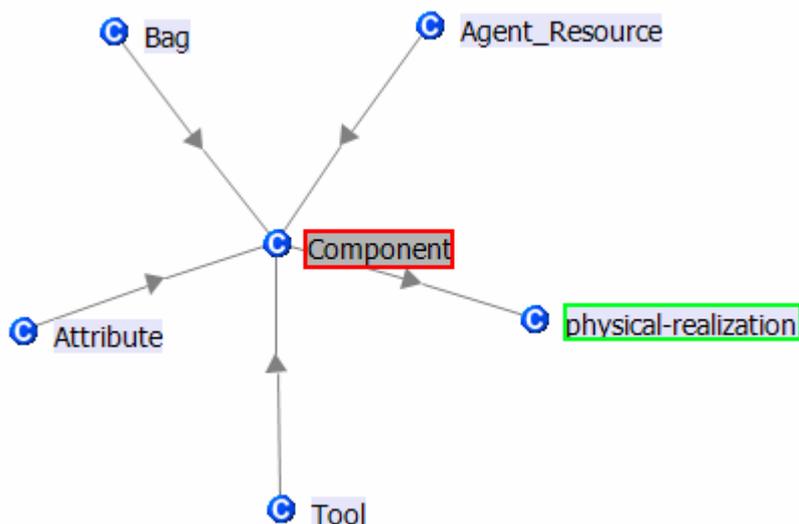
Figure 19: Component extension in DOLCE

Process Relations are the entities that contain process-level relations which can take place in a given process between the different actors participating in it. We classified process relations a subclass of Relation in DOLCE. Relation in DOLCE is *"A non-social relation(ship): formal, linguistic, etc. It is considered here a theory, because relations are established in order to give an ordering to some reality."* Both classes based in their definitions represent relations in a different level. The class "Process-relation" is a specification of the class "relation" of DOLCE. Both represents relations but "Process-relation" represents relations in a more specific domain. It is shown in Figure 20.
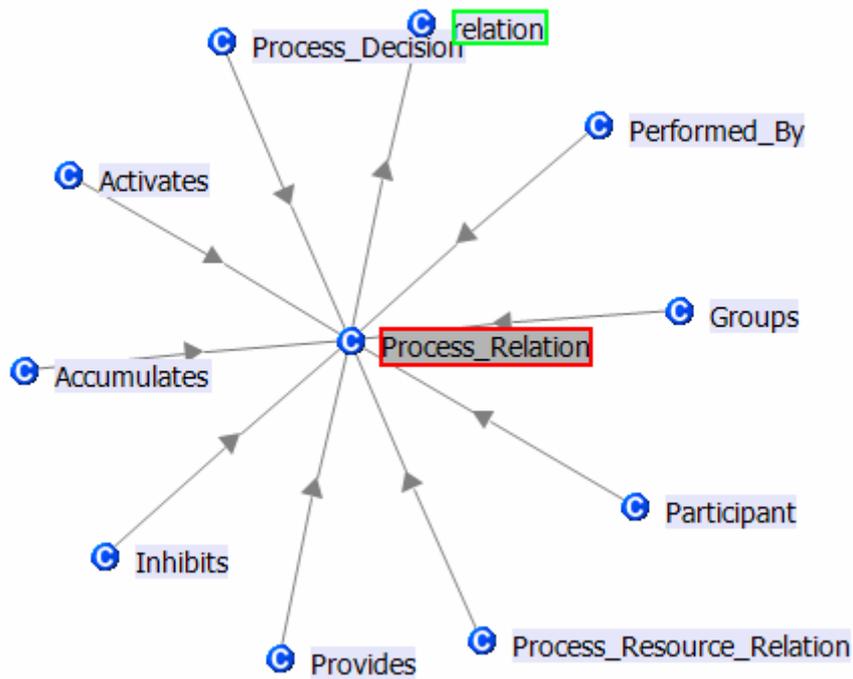
Figure 20: Process Relation hierarchy

We classify the Process entity like the Process class that is in the DOLCE upper level ontology, both represent the same. From the Process class of DOLCE we create the subclasses "Change", "Join", "Locate", "Metaprocess", "Split" and "Transaction" (Figure 21).
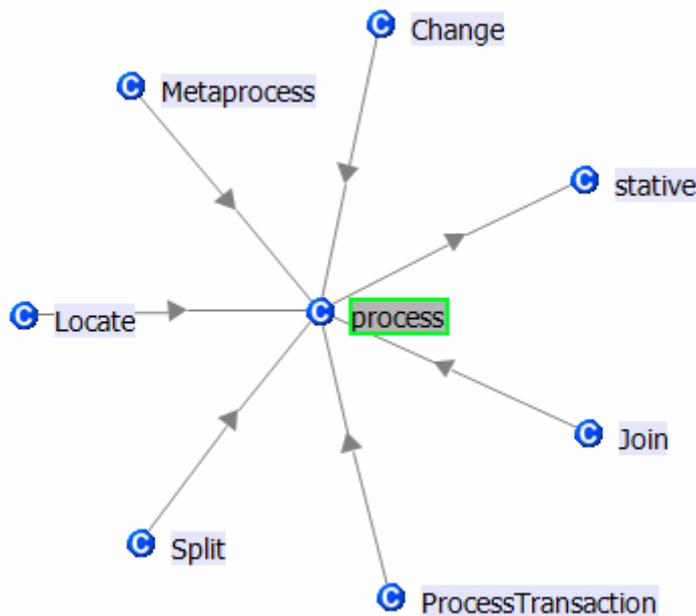
Figure 21: New process hierarchy

The Process_Action class in the process ontology classifies the actions that can be performed by the agents that participate in the processes. These actions are divided in Atomic Actions and Iterative Actions. We classify Process Action as subclass of "event". Event is defined as "*Eventive*

occurrences (events) are called achievements if they are atomic, otherwise they are accomplishments. Further developments: being 'achievement', 'accomplishment', 'state', 'event', etc. can be also considered 'aspects' of processes or of parts of them.". "Accomplishment" in DOLCE represents the same as "Process_Action" in the process ontology. Both classes represent events that happen and produce results. "Process_Action" is represented in Figure 22. In the previous definition it is stated that an "achievement" is an atomic action and "accomplishment" is composed by iterative actions. We added as superclass of "Iterative_Action" "accomplishment" and "achievement" as superclass of "Atomic_Action".
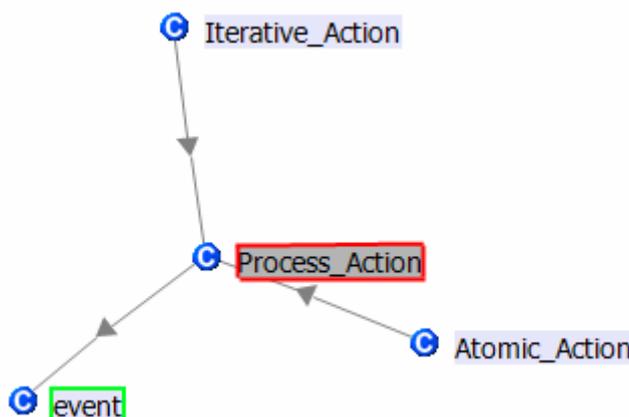
Figure 22: Process Action hierarchy

5.1.1.6.   Time ontology extension in DOLCE

The extension of the time ontology has been done as a subclass of "temporal-region". Temporal region in DOLCE is "*A region at which only temporal qualities can be directly located. It assumes a metrics for time.*" The time ontology extends this temporal region by specifying the metrics for time and a more specialized time representation. The extension is shown in Figure 23.
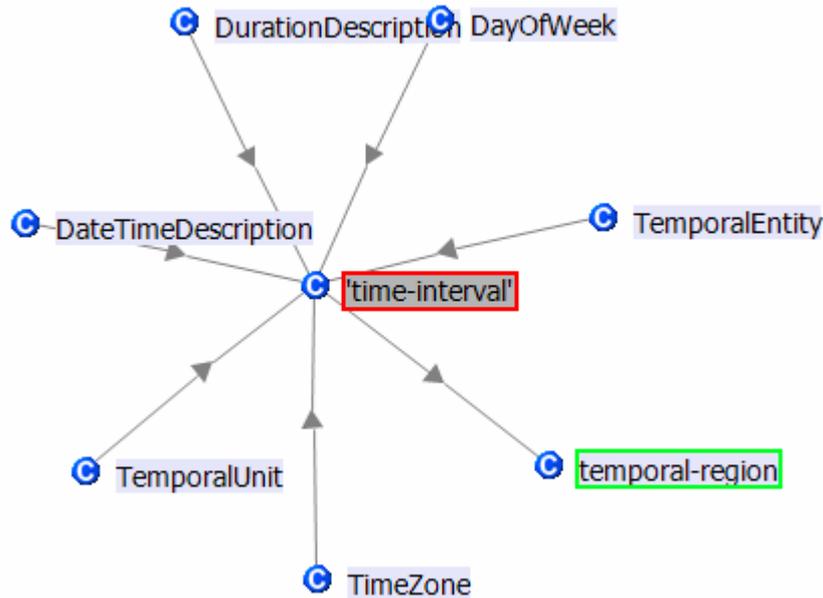
Figure 23: Time Ontology

5.1.1.7.   Invoice representation in DOLCE

The representation of an invoice is done by adding a specific subclass to the class "InvoiceAbstract". This class Invoice Abstract is a subclass of "TransactionRecordAbstract" which represents information regarding the whole or some part of a transaction event. Therefore an invoice is a sub concept of the superclass. Invoice abstract contains as subclasses the representation of the specific technologies invoices representations like "PharmaInnovaInvoice" (Figure 24). These classes contain all the possible relations and attributes that this invoice can have.
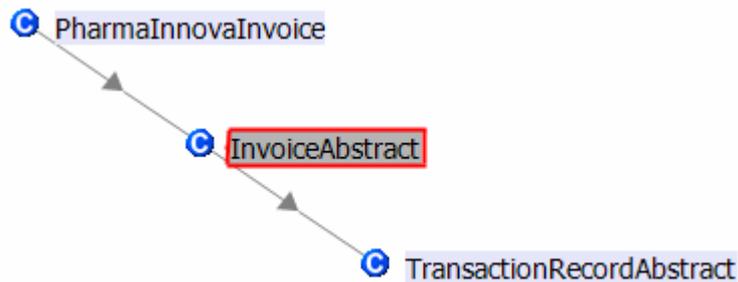
Figure 24: Invoice class

5.2.   Semantic Nomenclature Network of Ontologies

5.2.1. Naming conventions

5.2.2. Naming conventions

The naming conventions in the Nomenclature Ontology Network, concerning the classes, the relations, and the attributes in the ontology, are as follows: the label of a class is composed of one

or more words, written with capital first letters for each of the words, and without any intervals or alphanumeric symbols between them (in case a class label is a two-word one, for instance `BOTPlus:PharmaceuticalProduct`). The labels of relations and attributes follow the same rule, except for the non-capital first letter of the relation/attribute (e.g. `BOTPlus:isManufacturedBy`).

5.2.3. Current Status

The Nomenclature Ontology Network is in a process of constant development and improvement on the basis of the lifecycle ontology network regarding its usage and scenarios.

5.2.4. Pharmaceutical *Reference Ontology*

Classes: 15 classes

GSCoP_Council, Laboratory, Medical_Agency, Agemed_Agency, EMEA_Agency, FDA_Agency, Medical_Product, Medical_Product_Consumer, Hospital, Pharmacy, Medical_Product_Data_Base, Bot_Plus_Data_Base, Digitalis_Data_Base, Integra_Data_Base, Ministery_Of_Health.

Model description and justification:

This ontology has seven root concepts. Each concept represents a generic part of the pharmaceutical sector. Most of the classes included in the ontology represent the main stakeholders which take part in the system. In this way light hierarchies have been defined in order to represent medical product consumers (such as hospitals and pharmacies) or medical agencies (such as "Agemed_Agency" and "EMEA_Agency").

Furthermore the ontology represents knowledge sources like data bases, this is the case of "BOTPlus" (managed by GSCoP council), "Digitalis" and "Integra" (both of them managed by the "Ministery of Health"). Each data base is conceptualized in a different ontology within the network ontology described in section 3.2.2 and will be mapped to their corresponding class in the reference ontology.

The main class is "Medical_Product", that class will be mapped to most of the ontologies within the network ontology. On one hand such mappings will allow the inference and comparison of a product across different classification (i.e. ATC, SNOMED…). On the other hand, access to relevant product information will be allowed too across the mappings between "Medical_Product" and the data bases seen before.

Object properties

There are object properties represent the stakeholders' tasks related to the medical products:

"makeMedicalProduct" between "Laboratory" (Domain) and "Medical_Product" (Range).

"approvesMedicalProduct" between "Medical_Agency" and "Medical_Product".

Other object properties represent the relations between stakeholders each other:

"regulatesAgemedAgency" between "Ministry_Of_Health" and "Agemed_Agency".

"regulatesCouncil" between "Ministry_Of_Health" and "GSCoP_Council".

Some others represent the relation between stakeholders and knowledge bases:

"consultsDataBase" between "Medical_Product_Consumer" and "Medical_Product_Data_Base".

"managesDataBase"   between   "GSCoP_Council"   union   "Agemed_Agency"   and "Medical_Product_Data_Base".

Finally there are object properties which link different ontologies within the network ontology:

"classifiedByATC" between "Medical_Product" and "ATC_Classified_Product" (contained in the Active_Ingredient_Ontology").

"classifiedBySnomed" between "Medical_Product" and "Snomed_Classified_Product" (contained in the Snomed_Classification_Ontology").

"storedOnDataBase" which links a medical product to the data bases.

Datatype properties

There have been defined three datatype properties with the aim of storing the name of some stakeholders:

"councilName" -> "GSCoP_Council" (domain), "String" (range).

"agencyName"-> "Medical_Agency", "String".

"laboratoryName"-> "Laboratory", "String".
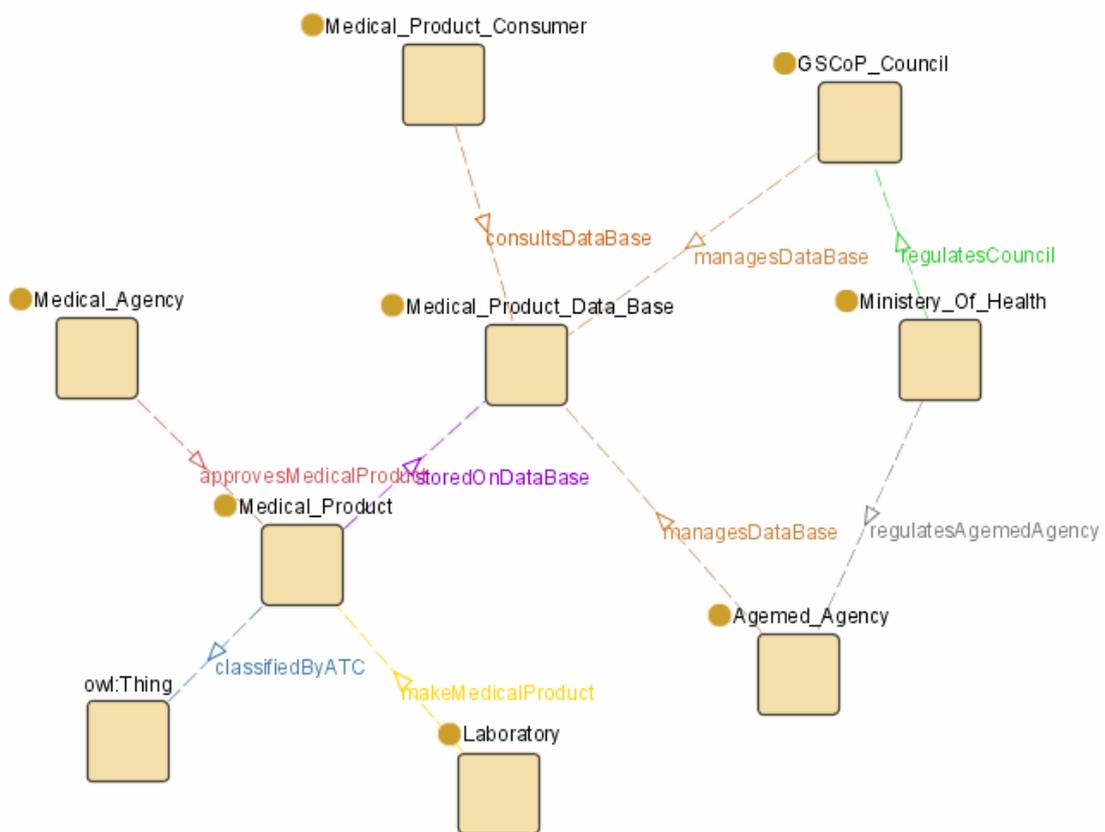
Figure 25: Pharmaceutical Reference Ontology (Primitive Class)

5.2.5. Other Ontology Models

5.2.5.1.   ATC Ontology

Classes: 122 classes

Model description and justification:

The ontology has two root concepts: "ATC_Code" and "Group_Code_Part". Despite this, it is implemented the concept "ATC_Classified_Product" that represents all the pharmaceutical products classified through the ATC code. This conceptualization of the hierarchy allows inference over the ontology model and obtains the therapeutical, anatomical, pharmacological or chemical group of one determinate pharmaceutical product from its ATC code. However, at the first version of the ontology, is modelled the two initial levels of the hierarchy (anatomical, therapeutical) due the extension of the classification. Even a complete example of all levels is modelled in order to demonstrate the mechanism of the classification.

Object properties

"Medical_Product" is related with "ATC_Code" via "hasATCCode" and is inherited in the "ATC_Classified_Product". This relation represents that pharmaceutical products are classified via their code formed by a character chain. This character code identifies the particulars subgroups that classify the pharmaceutical products.

"containsGroupPart" is the object property (Domain: "ATC_Code" Range:"Group_Code_Part") used for relating the ATC code with their correspondents anatomical, therapeutical, pharmacological and chemical subgroups.

Datatype properties

"partValue" is a datatype property (Domain: "Group_Code_Part"; Range:"String") used for representing the value of each of the subgroups or levels codified in the ATC code

Restrictions

The automatic classification of the pharmaceutical products via inference is due to all the subclasses of "Medical_Product" are defined classes. The considered restrictions in the defined classes are the necessary and sufficient condition (in each subgroup) that in its ATC code has the same value of each of the subgroups of the classification. Defining "ATC_Classified_Product" concept as subclass of "Medical_Product", when a new instance of a pharmaceutical product is created with its ATC code completed, the classification of the product based on the ATC is easier.
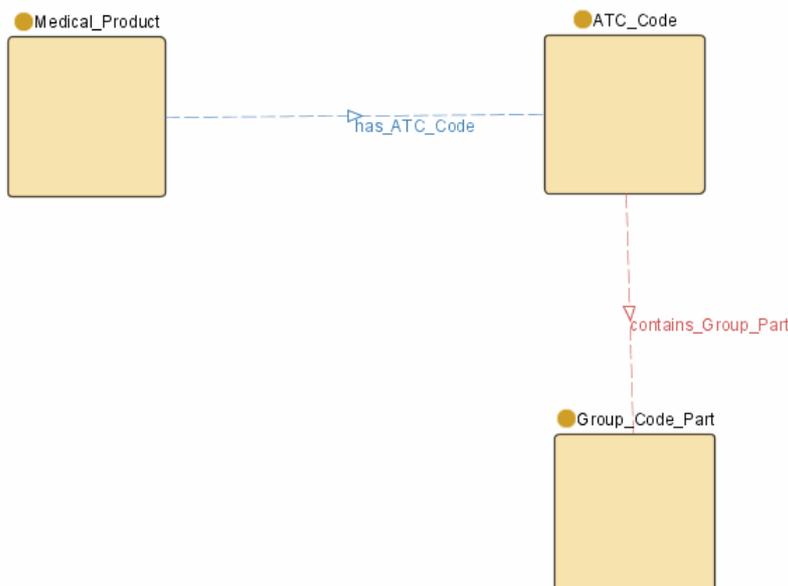
Figure 26: ATC Ontology (Primitive Class)

     5.2.5.2.   Digitalis Ontology

Classes: 12 classes:

Active_Ingredient, Chemical_Association, Composition, Dosage, Ingredient, Ingredient_AI, INSALUD_Therapeutical_Subgroup, Laboratory, Reference_Price, OMS_Therapeutical_Subgroup, Therapeutical_Subgroup, Pharmaceutical_Product, Pharmaceutical_Form, Status.

Model description and justification:

In this ontology is modelled the knowledge represented in the schema of the database Digitalis. The main concept is "Pharmaceutical_Product" that could be the point of link with the reference ontology. This link is possible via a mapping between Digitalis ontology and the Pharmaceutical Reference Ontology. Other classes represent the main concepts extracted from the tables of the DigitalisDB and the relations represented in their schema model.

In this ontology, the pharmaceutical products are classified based on the ATC classification or on the INSALUD classification (a code similar to the ATC for non-chemical pharmaceutical products).

Object properties

"has_INSALUDsubgroup"-> Domain:*"Pharmaceutical_Product"*

                    Range:"*INSALUD_Therapeutical_Subgroup*")

"hasActiveIngredient" ->   Domain:*"Pharmaceutical_Product"*,

                   Range:"*Active_Ingredient*"

"hasAssociation"->   Domain:*"Active_Ingredient"*

                   Range:"*Chemical_Association*"

"hasChemicalIngredient"->   Domain:*"Chemical Association"*

                                                         Range:"*Ingredient_AI*"

"hasOMSsubgroup"->   Domain:*"Pharmaceutical_Product"*

                                                         Range:"*OMS_Therapeutical_Subgroup*"

"hasPharmaceuticalForm"->Domain:*"Pharmaceutical_Product"*

                                                         Range:"*Pharmaceutical_Form*")

"hasReferencePrice"->   Domain:*"Pharmaceutical_Product"*

                                                         Range:"*Reference_Price*")

Datatype properties

There are several datatypes propierties specified in some of the concepts of the ontology, i.e:
*nationalCode, withdrawalDate, specificName, routeAdministration, price, bioequivalence, DDD,
regionPrice, substanceName, registrationDate...* All these datatype properties characterize each
pharmaceutical product and their components. These datatypes and their range are acquired from
the description of the Digitalis database.
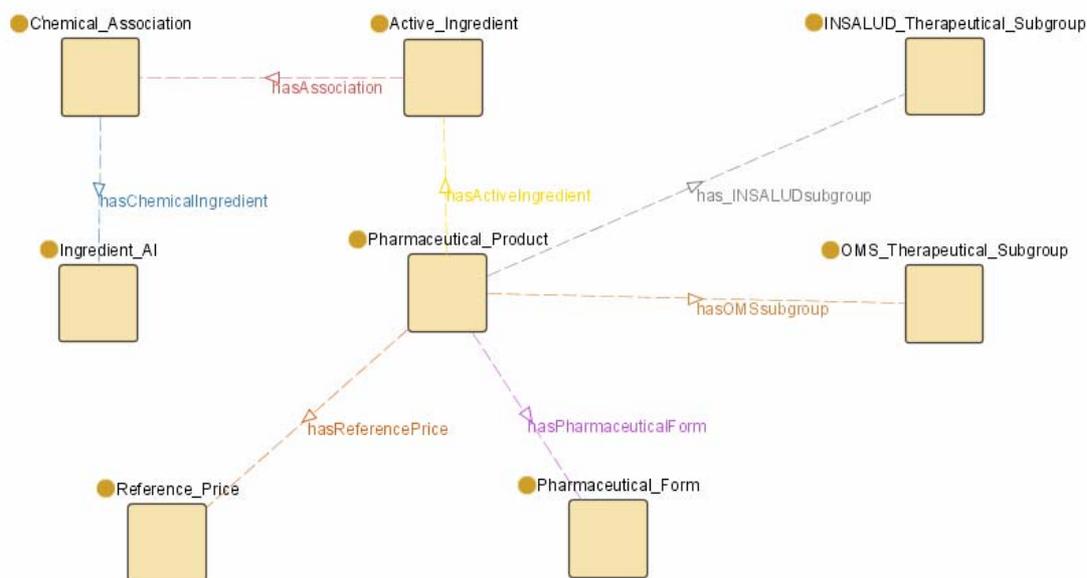
Figure 27: Digitalis Ontology

5.2.5.3.   BOTPlus Ontology

Classes: 37 classes

Model description and justification:

The BOTPlus ontology gathers the knowledge represented in the schema of the BOTPlus database. The main concept is "Pharmaceutical_Product" that could be one of the concepts that connect via mapping the BOTPlus ontology and the Pharmaceutical Reference Ontology. The BOTPlus ontology captures more data than pharmaceutical product information, as information about interactions, pathology, active ingredients… These concepts are related each other conceptualizing the relations represented in the BOTPlus schema model.

One characteristic of the BOTPlus model is that provides a classification of all pharmaceutical products based on a classification code and the specialty of the product: Human, Vet, Medical Herbs, Dermopharmacy and Parapharmacy. Figure 28 shows the hierarchy provided by the BOTPlus model for the pharmaceutical products.
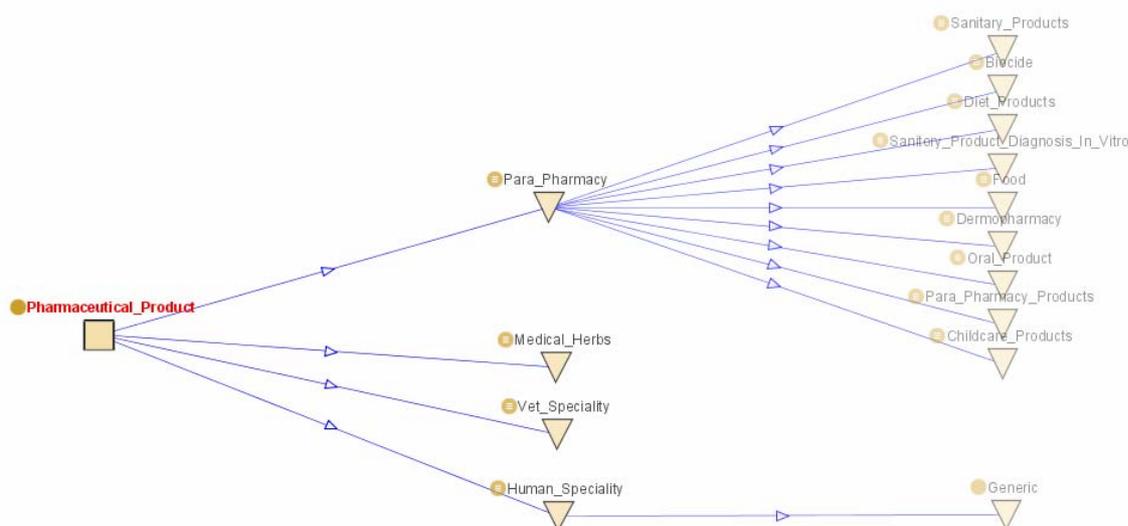
Figure 28: Pharmaceutical Product BOTPlus Hierarchy

Moreover, the pharmaceutical products have associated their ATC code if it is disposable or the therapeutical code provided by the Ministry of Health if the pharmaceutical product is a non-chemical pharmaceutical product.

The BOTPlus ontology model is linked via R2O (relational to ontology) mappings with the GSCoP database, providing mappings between the database objects and the ontology objects.

Object properties:

"hasActiveIngredient" ->	*Domain:"Ingredients_PharmaProduct"*
	Range:"Active_Ingredient"
"hasDosage" ->	*Domain:"Ingredients_PharmaProduct"*
	Range:"Dosage"
"hasIngredient" ->	*Domain:"Composition"*
	Range: "Ingredients_PharmaProduct"

"hasPharmacologicalActivity" ->   *Domain:"Active_Ingredient"*

                                       Range:"Pharmacological_Activity"

"hasRouteAdmin" ->   *Domain:"Dosage"*

                             Range:"Route_Administration"

"hasStatus" ->   *Domain:"Human_Speciality"*

                            Range:"State"

"interacts" ->   *Domain:"Active_Ingredient"*

                            Range:"Interaction"

"interactsWith" ->   *Domain: "Integration"*

                            Range: "Active_Ingredient"

"isComposed" ->   *Domain:"Human_speciality"*

                            Range:"Composition"

"isDeveloped" ->   *Domain: "Active_Ingredient"*

                            Range:"Laboratory"

"isManufacturedby" ->   *Domain: "Pharmaceutical_Product"*

                            Range:"Laboratory"

"produces" ->   *Domain: "Laboratory"*

                            Range: "Pharmaceutical_Product" "Active_Ingredient"

Datatype properties

As in the previous ontology, there are 75 datatypes properties specified in some of the concepts of the ontology, i.e: *nationalCode, withdrawalDate, specificName, routeAdministration, price, bioequivalence, DDD, regionPrice, substanceName, registrationDate...* All these datatype properties characterize each pharmaceutical product and their components. These datatypes and their range are acquired from the description of the BOTPlus database.


Restrictions

The automatic classification of the pharmaceutical products inside the hierarchy of "Pharmaceutical_Product" is due to the necessary and sufficient "hasValue" restriction defined in the subclasses over the datatype property "pharmaProductType", where each type of product has

a distinctive value to differ each type of product. Through this model, when a new instance of "Pharmaceutical_Product" is completed, the product is classified as instance of the correspondence subclass according to its type of product. Also, the hierarchy defined for "Para_Pharmacy" has an equivalent necessary and sufficient "hasValue" restriction over the property "parapharmType" defined in the subclasses.
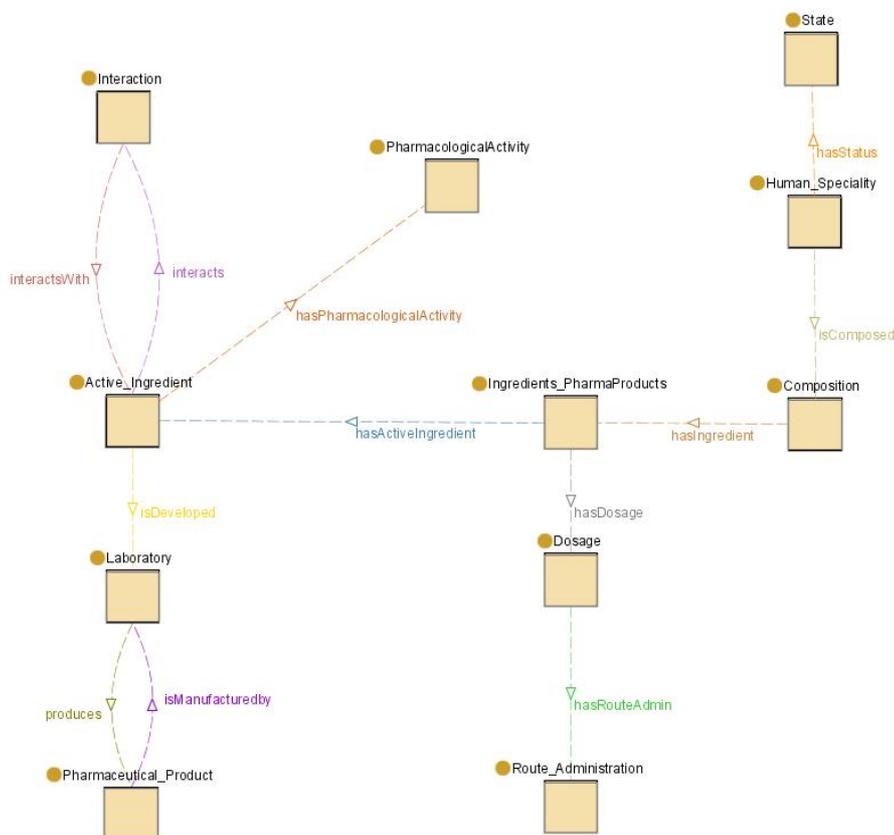
Figure 29: BOTPlus Ontology (Primitive Class)

5.3.   Pharmaceutical Ontology Networks in Use

5.3.1. Ontology-driven Invoice Mapping

The first application using the ontologies developed for the invoicing case study intends to facilitate invoice interoperability between peers using different invoice formats and models in the context of the PharmaInnova cluster.

The semantic annotation of electronic invoice data by means of an ontology formalizing the invoice model of PharmaInnova[51] is the basis for achieving efficient and easy interoperability across different invoice formats. For carrying out this task, our intention is to provide users with a GUI as close as possible to their domain and field of expertise (in this case, electronic invoicing), exploiting the underneath semantics provided by the ontologies, which will ensure consistency maintenance and constraint satisfaction in a way transparent to the user and the application developer.

---

[51] This ontology results from customizing the invoicing reference ontology described in section 5

Overall, the GUI retrieves user-defined annotations corresponding to fragments of sample electronic invoices and generalizes them, producing and storing a number of configuration parameters that define how electronic invoices of this kind can be imported into the ontology, populating it. Once all the relevant pieces of invoice information have been annotated, the knowledge required to import (and, by extension, export) all electronic invoices, represented by the sample invoice, into the ontology (and from the ontology to the concrete electronic invoice format of a particular user) is stored to be used in the future (Figure 30).

In summary, the invoicing ontologies described in this document are instrumental in this scenario to i) providing a conceptual model of the information related with invoicing, which embraces the different existing standards, ii) ensuring consistency of exchanged invoice data with respect to the formal model of these ontologies, and ultimately iii) allowing users to easily define themselves the mapping between their invoices and a common, agreed model, supported by the ontologies, to automate invoice exchange between business peers.
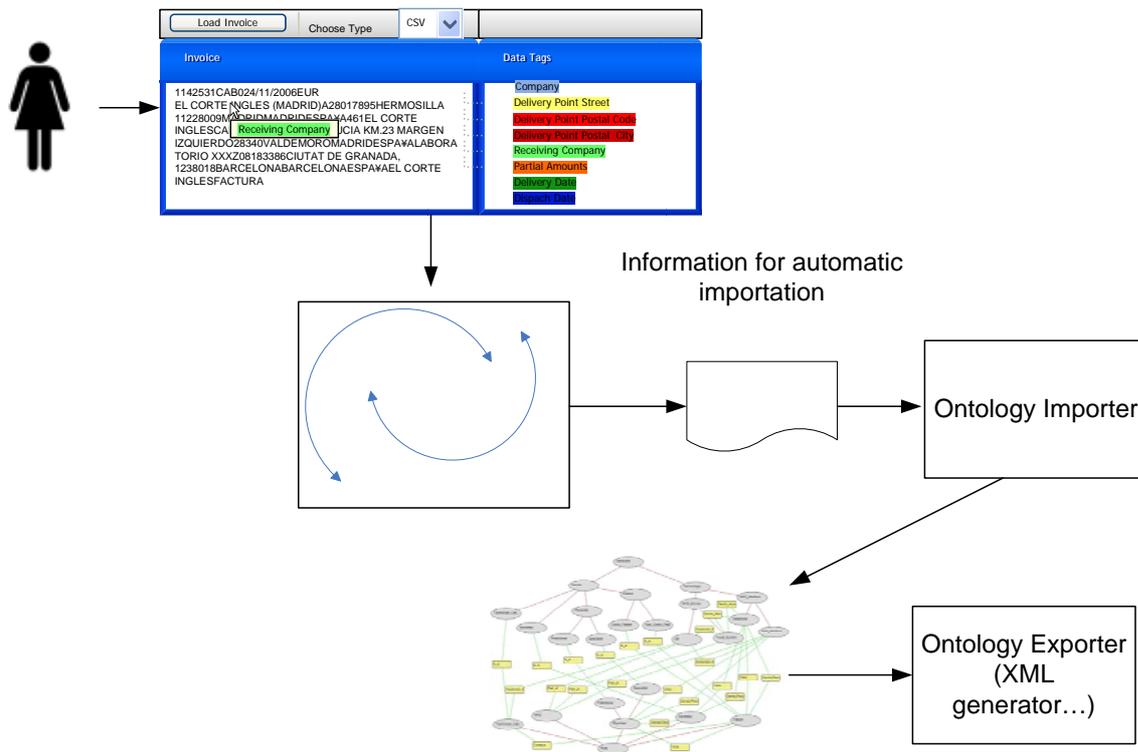
Figure 30: Overall ontology-driven invoice mapping workflow

In this section we present a sample interface which exploits the properties of the ontology to guide users through the annotation (and eventually, mapping with the ontology) of their invoices. This problem can be represented using a layered approach (Figure 31), where each layer represents a different level of abstraction: the *interface* layer enables natural user-system interaction. The *formats* layer contains the knowledge about the most representative formats for invoices, while the *ontology* layer represents a formal model for the information given in an invoice.
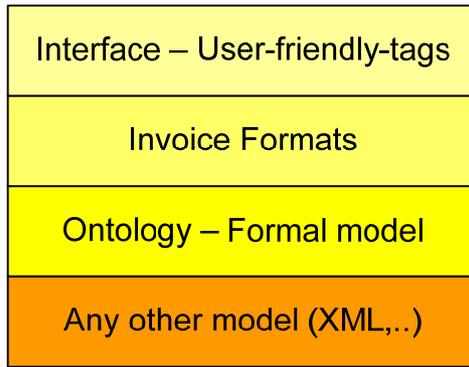
Figure 31: Ontology-driven invoice mapping abstraction layers

As shown on the right-hand side of Figure 32, at the GUI top level, friendly tags like e.g. : *Delivery Point Street*, *Delivery Point City*, *Receiving Company*, etc are used to denote invoice elements, which correspond to ontology concepts or attributes. For example, *Delivery Point Street*, *Delivery Point City* could be attributes of concept *Address* and have a relation with an instance of concept *Receiving Company*. Figure 32 also shows a sample GUI for this application. Note that this tag-based visualization approach for ontology entities in the invoicing domain could be substituted e.g. with a graphical representation of an invoice as they are traditionally displayed to humans (Figure 33).

Figure 32: Sample GUI

Figure 33: Traditional invoice representation

Users will be enabled to load two main types of invoice formats:

- **CSV (Comma Separated Values)**: Data is separated by special a character (could be a comma or any other character). EDIFact is a special type of CVS with four different separators. See example in Table 11.

- **FLF (Fixed Length Format)**: Fields of the electronic invoice are identified by its offset. There is no separation character. iDOC is the FLF implementation used by SAP. See example in.

- **XML**, e.g. the original PharmaInnova model.

CSV

```
1142531;CAB;0;24/11/2006;EUR;EL          CORTE          INGLES          (MADRID);A28017895;HERMOSILLA
112;28009;MADRID;MADRID;ESPA¥A;461;EL    CORTE    INGLES;CARRETERA    DE    ANDALUCIA    KM.23    MARGEN
IZQUIERDO;28340;VALDEMORO;MADRID;ESPA¥A;LABORATORIO          XXX;Z08183386;CIUTAT          DE          GRANADA,
123;8018;BARCELONA;BARCELONA;ESPA¥A;EL CORTE INGLES;;FACTURA;;;;;;;;;;;;;;;;;;

1142531;LIN;5;24/11/2006;PACK                              2x1                          CEPI.KIN
NAT.+MUESTRA;48566;193006;66617414;12;1;2,6;5;2,47;29,64;IVA;16;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;

1142531;PIE;999;PAGARE;PAGARE                                                                      90
DIAS;;23/02/2007;34,38;;;;;;;;;;;;29,64;;29,64;IVA;29,64;16;4,74;;;;;;;;;;;;;;;;;;;;;;34,38
```

Table 11: CSV example (fields separated by ";")

FLF

```
SEGCA1    001INDUKERN, S.A.                          A08135055         P.EMPR. MAS BLAU II, ALTA
RIBAGORZA, 6-8EL PRAT DE LLOBREGAT              BARCELONA                              ESPAÑA
08820ALMIRALL – PRODESFARMA, S.A.                    A-58869389        General  Mitre,  151
Barcelona                          Barcelona                              ESPAÑA
08022FN156653  20/12/2006157414EUR

SEGLI1    001  20062714514 CLORTALIDONA BP93              25,000 KG  19/12/2006      113,5900
2.839,75                  0003307960161                  NR.                        45086828
4,00
```

```
SEGIV1    001    2.839,75 4,00    113,59
```

Table 12: FLF example

On the right-hand side of the GUI, concepts and attributes of the ontology are shown in a user-friendly way, as indicated above. In a domain like this, where user are completely illiterate about ontologies, their benefits, and how they are defined, it is important to provide an abstraction layer which allows exploiting ontologies exclusively in terms of the user domain. Users are enabled to annotate their invoices by means of a Drag & Drop mechanism which is valid to relate invoice segments and ontology entities in 90% of the occasions. On the other hand it is important to highlight that the behaviour of the GUI is guided by the ontology, enforcing constraints like e.g. cardinality constraints. Hence, if, for example, concept *Receiving Company* of an invoice is defined as sufficient and necessary in the ontology, the GUI will prevent users from annotating other than one invoice segment with this concept.

Once that part of the text is selected, this will appear highlighted with the same colour as the corresponding ontology concept.

1142531CAB0 24/11/2006 EUR EL CORTE INGLES (MADRID)A28017895HERMOSILLA 11228009MADRIDMADRIDESPA¥A461EL CORTE INGLESCARRETERA DE ANDALUCIA KM.23 MARGEN IZQUIERDO28340VALDEMOROMADRIDESPA¥ALABORATORIO XXXZ08183386CIUTAT DE GRANADA, 1238018BARCELONABARCELONAESPA¥AEL CORTE INGLESFACTURA

Table 13: Electronic invoice annotation (left-hand side)

At the same time, when a selection is performed, the corresponding ontology entity is highlighted in the right part of the GUI. This way the user always has the feedback on what the application records. Hence actual data invoice (left), domain-based representation (right) and ontology are bound. The task of selecting data on the left side can be supported by Knowledge Tagger, an extension of GATE. With Knowledge Tagger, it is possible to identify organization names, currencies, countries, dates, etc, enabling the GUI to make suggestions to the user on how data can be mapped against the tags on the right.

Figure 34: Electronic invoice annotation (right-hand side)

Following these steps, users can use data of a sample invoice to build the mappings between all the electronic invoices of a given organization and our ontology, following an inductive approach.

Behind the graphical interface

As output, the application generates a configuration file which defines how electronic invoice data is imported into the ontology and exported from the ontology into a particular electronic invoice format. Depending on the invoice format, the information to keep is different. In the case of CSV, for example, the relevant information is the number of separators, e.g. commas, which separate the selected text from the beginning of the invoice. With FLF, the offset is important. In the case of XML, the tags around the selected text must be considered.

The importation layer selects the relevant information to be kept according to the invoice format. The Ontology Mapping module receives:

- data identifier (CSV position, XML tag, offset, etc)

- tag selected by the user

- section (header, body, summary etc.)
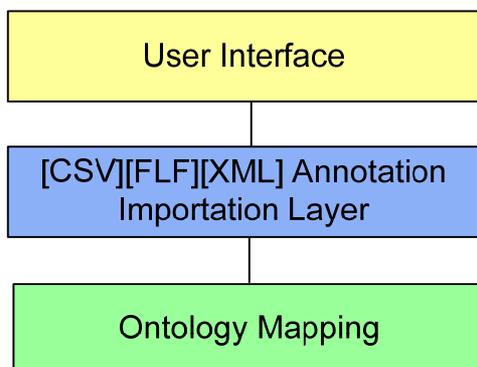
Figure 35: Ontology-driven invoice mapping architecture

The ontology mapping module contains the knowledge about the relation between the tags in the interface and the ontology. The output of this module is a text file. The information stored in the file will be the input for another component, i.e. an interpreter which imports data automatically from an invoice into an ontology.

Once these files have been produced (one for each invoice format and for each provider), each time a new invoice is received the system will be able to extract automatically the data and import it in the ontology. The ontology importer is composed by as many modules as invoice formats in order to access data in the correct way (with offset, position, tags etc).
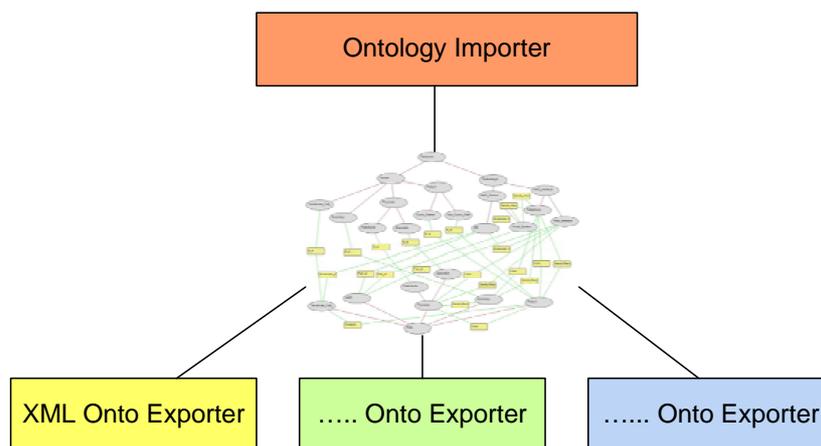
Figure 36: Invoice to ontology import/export

The ontology is the best tool to store invoice data in a semantic data structure. Furthermore, its use simplifies the development of exporters to different formats, as for instance XML, guaranteeing interoperability between different invoice formats.

5.3.2. Ontology-Based Access to Medical Product Knowledge

The aim of the Semantic Nomenclature application is to improve the medical product information management within the Spanish pharmaceutical domain which is presently decentralized in several data bases and repositories. The application uses the ontology network, seen in section 2.3.6, in order to achieve this main goal. As described in this section, the ontology network contains a main ontology, called Pharma Reference Ontology, which centralizes the key medical product information and links the rest of ontologies each others. It is especially useful in that use case to define an ontology network for allowing a more complex information retrieval as well as resolving medical product identification and different classification alignments.

The application access to the knowledge stored on three different data bases, BOTPlus, Digitalis and Integra, corresponding to the ontologies shown in Figure 7. In this way the information retrieved using a user query is extended across different stored fields in each ontologies. Furthermore there are two ontologies defined in the application which represent standard classifications for medical product, these ontologies are Active Ingredient Ontology and Snomed Ontology. The application will be able to identify an unknown medical product, accessing to the ontologies, and even to auto-classify a medical product attending to his numerical code. The connection between both ontologies within the network allows the system to compare both classifications giving the user the option to know a product classification in both standards whatever the code he knows. One of the best improvements, considering the present medical product information management, is the semi-automatic BOTPlus actualization in which information related to new medical product will be added to the knowledge base and auto-classified, it is a good practice anyway to review the automatic classification for insuring the absence of mistakes. The application can recommend a drug for a given illness by adding illness ontology due to the relation between the therapeutical use of the drugs and the disease. However it is especially important again to review the results obtained considering the possible derived consequences of a wrong inference.

The extension of the product information retrieved in the user queries is implemented via several relations between Digitalis, Integra and BOTPlus ontologies. These relations are defined across the reference ontology as object properties. In addition to this, several mappings can be defined between equivalent concepts in both ontologies with the aim of retrieving properties defined in any data base not contained in the others. In this way the maximum set of attributes and properties of a given product will be returned in each query. The Figure 37 shows the relation between Reference and BOTPlus ontologies connected via one of the mappings.
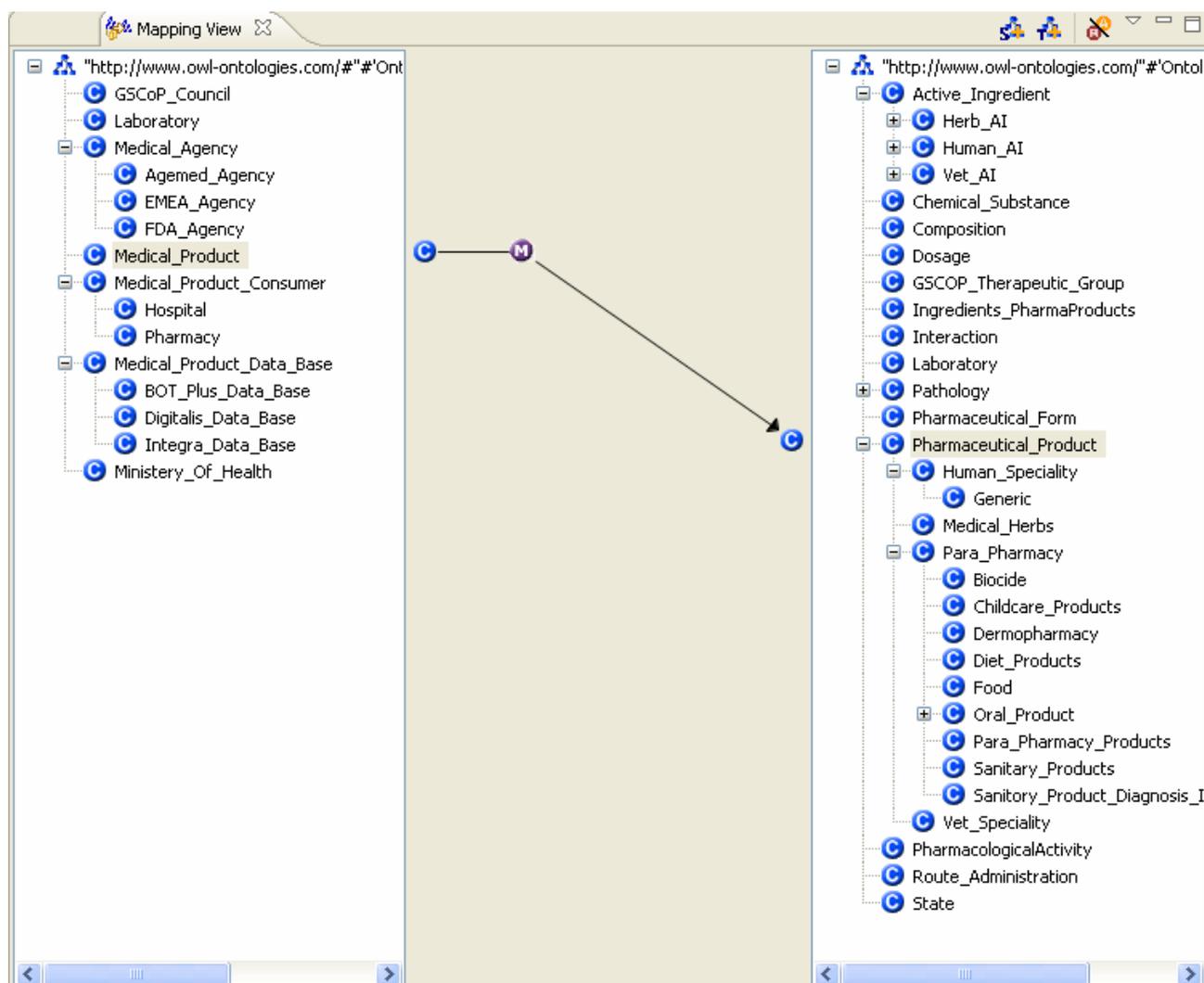
Figure 37: Mapping between Pharmaceutical Reference Ontology - BOTPlus Ontology

The automatic classification of a product is supported by Active Ingredient and Snomed ontologies, for this proposal both of them are implemented defining complete classes (classes which implement necessary and sufficient conditions) thus a medical product, represented by an instance, will be inferred in a medical subgroup attending to his numerical code. These models provide a mechanism as well for testing the product classifications. Each class defined within the ontologies implements several restrictions referring to the numeric code which describes the medical groups that a medical product belongs to.

The ontology network extends the above functionality allowing the alignment between both classifications. The alignment is provided by several mappings between both ontologies, these mappings link equivalent medical groups, thus a medical product will be auto-classified not only in a specific standard but in both of them. It is desirable a domain expert to review these mappings in order to check consistency and truthfulness on them.

As well as the alignment between both product classifications, the ontology network allows other kinds of alignment using mappings. It is the case of the relation between medical products and illnesses. Such mappings are defined between a specific illness and their associated drugs. As it is difficult to map each illness with their associated drugs, other mappings are defined between higher level concepts in each ontology, it is medical groups in the case of drugs and kind of illness. These high level mappings provide a drug recommendation for a given illness which is the medical group recommended for an illness group, both of them obtained in an automatic way across the classification.

6. Conclusions

This document contains a description of the ontologies which have been developed so far in the context of the NeOn Pharmaceutical case studies, encompassing the following issues: i) application of the NeOn methodology to the Pharmaceutical case studies and, in return, how the case studies have contributed in practice to the development of this methodology; ii) inventory of existing knowledge resources, either ontological or non-ontological; iii) ontologies resulting from the application of this methodology to the Pharmaceutical domain; and iv) how the development of new ontologies, together with reutilization and extension of existing ones and the semantization of non-ontological resources contribute to improving the problems addressed by the case studies.

We have provided a twofold testbed for the NeOn methodology, which has been applied to the management of the knowledge lifecycle in two varied aspects of the pharmaceutical domain: electronic invoicing, i.e. improving interoperability in the exchange of business electronic documents, and semantic nomenclature of the various pharmaceutical products across the different existing repositories. We have first described the methodological approach used to develop the ontologies resulting from this deliverable and then described its application to the two particular case studies as well as the results produced. Mechanisms like the Competency Questions have provided additional insight and refinement of user requirements particularly on knowledge resources. On the other hand, this document has also intended to show the use of NeOn metamodel in the pharmaceutical ontologies described within.

We have provided a extensive inventory of the pre-existing resources surveyed in order to accomplish such requirements and then refined it into the definitive set of resources that we have finally reused via either extension or customization. These resources include ontological and non-ontological resources, as well as the most relevant standards, e.g. for exchange of electronic B2B documents like EDIFACT or UBL or the ATC and EphMRA chemicals classification systems. Using these resources as a starting point, we have provided for both case studies networked ontologies that can be used as reference knowledge for the general invoicing and nomenclature cases. We have also provided their specializations for particular applications like e.g. in the case of invoicing specific support to the PharmaInnova cluster of laboratories. Finally, we have provided a glimpse of the use of these ontologies in the context of the software prototypes to be produced in WP8 from M24 on.

References

[1] Gómez-Pérez J.M, Pariente T, Daviaud C, Herrero G. *Analysis of the pharma domain and requirements*. NeOn Deliverable D8.1.1, 2006.

[2] Gómez-Pérez J.M, Buil C, Muñoz, O.*Requirements on NeOn architecture*. NeOn Deliverable D6.1.1, 2006.

[3] The TOVE Project: A Common-sense Model of the Enterprise, *Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, Belli, F. and Radermacher, F.J. (Eds.), Lecture Notes in Artificial Intelligence # 604, Berlin: Springer-Verlag, pp. 25-34.

[4] M. Uschold, M. King, S. Moralee and Y. Zorgios, *The Enterprise Ontology*. 1998 - Cambridge Univ Press.

[5] Boxwala AA, Peleg M, Tu S, Ogunyemi O, Zeng QT, Wang D, Patel VL, Greenes RA, Shortliffe EH. *GLIF3: a representation format for sharable computer-interpretable clinical practice guidelines.* Decision Systems Group, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.

[6] Dimitrov M, Simov A, Stein S and Konstantinov M., *A BPMO Based Semantic Business Process Modelling Environment*. SUPER Deliverable, D1.1 Process Modelling Ontology and Mapping to WSMO, 2007.

[7] Niles, I., and Pease, A. 2001. *"Towards a Standard Upper Ontology"*. In Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001), Chris Welty and Barry Smith, eds, Ogunquit, Maine, October 17-19, 2001.

[8] Fernández López M, Gómez Pérez A (2004) Searching for a Time Ontology for Semantic Web Applications. International Conference on Formal Ontology in Information Systems, Turín.

[9] Hutcheson, Kenneth R. (1990). Trends and Strategies in ASC X12. *Information Standards Quarterly.* 2(4):13-14.

[10]   UN/EDIFACT Rapporteurs' Team. (1990). *Introduction to UN/EDIFACT with latest news and events.* United Nations Economic Council for Europe.

[11]   Webber D. and Dutton A., *Understanding ebXML, UDDI and XML/EDI*. 2000. http://www.xmlglobal.com/downloads/ebXML_understanding.pdf

[12]   Gómez-Pérez J.M, Pariente T, Buil-Aranda C, Herrero G. *Software architecture for the NeOn pharmaceutical case studies*. NeOn Deliverable D8.2.1, 2007.

[13]   Fernández, M.; Gómez-Pérez, A.; Juristo, N. *METHONTOLOGY: From Ontological Art Towards Ontological Engineering*. Ontological Engineering on Spring Symposium Series. Stanford. 1997. Págs: 33-40.

[14]   Mari Carmen Suárez-Figueroa et al. *NeOn Development Process and Ontology Life Cycle*. NeOn Deliverable D.5.3.1, 2007

[15]   Gruninger M. R.; Fox M. *Methodology for the Design and Evaluation of Ontologies.* Workshop on Basic Ontological Issues in Knowledge Sharing held in conjuntion with IJCAI-95. Montreal, (Canada) 1995.

[16]   Lenat DB, Guha RV. *Building Large Knowledge-based Systems: Representation and Inference in the Cyc Project*. Addison-Wesley, Boston, Massachusetts 1990.

[17]   Q. Zhou & R. Fikes; *A Reusable Time Ontology*; Proceedings of the Ontologies for the Semantic Web Workshop, AAAI 2002 National Conference.

[18]   Fernández López M, Gómez Pérez A.*Searching for a Time Ontology for Semantic Web Applications*. International Conference on Formal Ontology in Information Systems, Turín, 2004

[19]   Keller, G., Nüttgens, M., & Scheer, A.-W. (1991). Semantische Prozessmodellierung auf der Grundlage "*Ereignisgesteuerter Prozessketten (EPK)*". Saarbrücken.

[20]   Q Zhou, R Fikes - Proceeding of the Ontologies for the Semantic Web Workshop, *A Reusable Time Ontology*, 2002

[21]   Tempich C, Pinto H S, Sure Y, Staab S: *An Argumentation Ontology for DIstributed, Loosely-controlled and evolvInG Engineering processes of oNTologies (DILIGENT)*. ESWC 2005: 241-256

[22]   Zao Y. *Develop the Ontology for Internet Commerce by Reusing Existing Standards*, 2003.

[23]   Gessa N, Busanelli M, De Sabbata P, Vitali F: *Extracting a semantic view from an ebusiness vocabulary*. The 8th IEEE International Conference on and Enterprise Computing, E-Commerce, and E-Services, 2006

[24]   P. Haase, S. Rudolph, Y. Wang, S. Brockmans, R. Palma, and J. Euzenat, M. d'Aquin. *NeOn Deliverable D1.1.1 Networked Ontology Model*. November 2006. Available at: http://www.neon-project.org/.

[25]   Christoph, B., Emilia, C., Dieter, F., Juan, M., G., Armin, H., Thomas, H., Michael, K., Adrian, M., Matthew M., Eyal, O., Brahmananda, S., Ioan, T., Jana, V., Tomas, V., Maciej, Z., & Michal, Z. (2005). Web service execution environment (wsmx). Retrieved April 20, 2007, from http://www.w3.org/Submission/WSMX/.

[26]   Sinuhe, A., Emilia, C., John, D., Cristina, F., Dieter, F., Birgitta, K., Holger, L., Axel, P., & Michael, S. Web service modeling ontology primer (2005). Retrieved June 12, 2007, from http://www.w3.org/Submission/WSMO-primer/.

[27]   Sean Bechhofer S, van Harmelen F, Hendler J, Horrocks I, McGuinness D L, Patel-Schneider P F, Stein L A, Franklin W. O, eds. *OWL Web Ontology Language Reference*, available at http://www.w3.org/TR/owl-ref/

7. ANNEX I: Activities used in the development of the invoice reference ontology

	Required	*If Applicable*	*Selected*
Ontology Aligning		X	X
Ontology Annotation	X		X
Ontology Assessment	X		X
Ontology Comparison	X		X
Ontology Conceptualization	X		X
Ontology Configuration Management	X		
Ontology Control	X		X
Ontology Customization		X	
Ontology Diagnosis	X		X
Ontology Documentation	X		X
Ontology Elicitation	X		X
Ontology Enrichment	*24*	X	
Ontology Environment Study	X		X
Ontology Evaluation	X		X
Ontology Evolution	X		X
Ontology Extension		X	X
Ontology Feasibility Study	X		X
Ontology Formalization	X		X
Ontology Forward Engineering		X	
Ontology Implementation	X		X
Ontology Integration	X		X
Knowledge Acquisition for Ontologies	X		X
Ontology Learning		X	
Ontology Localization		X	
Ontology Matching		X	
Ontology Merging		X	X
Ontology Modification		X	X

Ontology Modularization		X	X
Ontology Module Extraction		X	X
Ontology Partitioning		X	
Ontology Population		X	
Ontology Pruning		X	
Ontology Quality Assurance	X		*X*
Ontology Reengineering		X	
Ontology Restructuring		X	
Ontology Repair	X		*X*
Ontology Reuse	X		*X*
Ontology Reverse Engineering		X	
Ontology Scheduling	X		*X*
Ontology Search	X		*X*
Ontology Selection	X		*X*
Ontology Specialization		X	X
Ontology Specification	X		*X*
Ontology Summarization		X	
Ontology Translation		X	
Ontology Update		X	
Ontology Upgrade	X		*X*
Ontology Validation	X		*X*
Ontology Verification	X		*X*
Ontology Versioning	X		*X*

Table 14: Activities used in the invoice use case

8. ANNEX II. Invoicing competency questions

CQ1.   Is possible to identify the activity of one invoice emitter by looking at the invoice model? In general not

CQ2.   How many different concepts are the different invoices emitted by e.g. wholesaler and a laboratory? 5 - 10 concepts

CQ3.   What are the differences between the model of the invoices emitted by e.g. wholesaler and a laboratory? Mainly in bank/financial and time information

CQ4.   What concepts are mandatory for a wholesaler/provider/laboratory? Two types of information: first regarding the identification of companies (names, addresses, bank accounts, etc) and second information about the amounts of the products in the invoice and their prices.

CQ5.   Is required any specific type of information about the activity of the emitter of the invoice in the invoice? No

CQ6.   Does the company id identify the type of organization that emits/receives invoices? No

Competency questions regarding the invoicing workflow:

CQ7.   What is necessary to identify the emitter of the invoice? NIF/CIF

CQ8.   What is necessary to identify the receiver of the invoice? NIF/CIF

CQ9.   What is necessary to identify the products in the invoice? Product description

CQ10. What is the location of the products that the emitter is sending in the invoice? In the product lines

CQ11. What is the name of the emitter of the invoice? The supplier name

CQ12. What is the id of the emitter of the invoice? NIF/CIF

CQ13. What is the name of the receiver of the invoice? The customer name

CQ14. What is the id of the receiver of the invoice? NIF/CIF

CQ15. What is the address of the emitter of the invoice? The supplier fiscal address

CQ16. What is the address of the receiver of the invoice? The customer fiscal address

CQ17. What is the status of the invoice? It depends of the state of the invoice: imported, emitted, in process, accepted, in creation or disused.

Competency questions regarding multilinguality:

CQ18. What is the language of this invoice? Currently only Spanish

CQ19. Where the emitter of the invoice is from? Spain

Competency questions regarding inference rules:

CQ20. What is the total discount applied to this invoice? Discounts in payment date

CQ21. Is possible to apply any special price to this invoice? Yes, if you describe the concept in the description line (see screenshots)

CQ22. Is possible to apply any business rule in this invoice? No, only the related to amount to pay and supplier code.

CQ23. What is the unitary price before applying discounts? The net price

Specific competency questions related to the receiver of invoices:

CQ24. What product have we received? Answered in the specific invoice received

CQ25. Did we order this product? Answered in the specific invoice received

CQ26. What quantity of product X have we bought? Answered in the specific invoice received

CQ27. Who is sending this invoice? Answered in the specific invoice received

CQ28. The emitter of the invoice, did he apply any discount? Answered in the specific invoice received

CQ29. How much is the total price of the invoice? Answered in the specific invoice received

CQ30. How many products are in the invoice? Answered in the specific invoice received

CQ31. What are the products in this invoice? Answered in the specific invoice received

CQ32. Where did the emitter send the product X? Answered in the specific invoice received

CQ33. When do we have to pay? Answered in the specific invoice received

CQ34. Where do we have to pay? Answered in the specific invoice received

CQ35. Did everything arrive? Answered in the specific invoice received

CQ36. What are the details of the invoice? Answered in the specific invoice received

CQ37. What products have we received during the last week? Answered in the specific invoice received

CQ38. In what format is this invoice? Answered in the specific invoice received

CQ39. What fields does this invoice include? See the screenshots

Specific questions regarding the technology used by the emitters:

CQ40. What invoicing technologies are using the emitters of this invoice? ERP's, and small products for invoicing like Facturaplus or Contaplus, and customize applications

CQ41. What technologies use to use the emitters of the invoice? ERP's, and small products for invoicing like Facturaplus or Contaplus, and customize applications

CQ42. In percentage, can you classify the invoicing technologies of each emitter? 65% CSV and FLF, 25% xml, 10% EDI

CQ43. To what industrial sector does this invoice belong (chemical, logistic, estate agency, etc.)? Pharmacy, depending of the invoice

CQ44. Is possible to classify the technologies depending on the business type of the emitter of the invoice? No

Specific competency questions related to the emitter of invoices:

CQ45. What product did we sell in invoice X? Answered in the specific invoice received

CQ46. What amount of product X did we sell in invoice X? Answered in the specific invoice received

CQ47. What are the details of the product sold in invoice X? Answered in the specific invoice received

CQ48. What companies are buying the product X? Answered in the specific invoice received

CQ49. Have we sold any other product in invoice X? Answered in the specific invoice received

CQ50. How much is each product in invoice X? Answered in the specific invoice received

CQ51. What amount of this product have we sold in invoice X? Answered in the specific invoice received

CQ52. Where are we going to send the products in invoice X? Answered in the specific invoice received

CQ53. Where is our warehouse? Answered in the specific invoice received

CQ54. Where is the product that we are selling in invoice X? Answered in the specific invoice received

CQ55. What are the details of our company? Answered in the specific invoice received

CQ56. What are the details of the bank account of our company? Answered in the specific invoice received

CQ57. When do we have to deliver the product sell in invoice X? Answered in the specific invoice received

CQ58. What taxes do we have to charge in invoice X? Answered in the specific invoice received

CQ59. Do we have to apply any special price in invoice X? Answered in the specific invoice received

CQ60. Do we have to apply any specific rule in invoice X? Answered in the specific invoice received

Specific Competency questions related to time and date management:

CQ61. When do we have to deliver the goods sell in invoice X? The goods have to be delivered before the invoice arrive to the customer

CQ62. When do we have to pay the invoice X? It depends on the agreement

CQ63. When are the goods arriving bought in invoice X? In the agreed date

CQ64. When did we receive invoice X? After receiving the goods

CQ65. When did we send invoice X? After receiving the goods

CQ66. When was invoice X send? Answered in the specific invoice

CQ67. What is the expiry date of invoice X? There is not expiry date

CQ68. How much money did we earn during the week? Depending of the received invoices

CQ69. How much money did we earn during the month? Depending of the received invoices

CQ70. How many products did we sell during the week? Depending of the received invoices

CQ71. How many products did we sell during the month? Depending of the received invoices

CQ72. How much money did we spend during the week? Depending of the received invoices

CQ73. How much money did we spend during the month? Depending of the received invoices

CQ74. How many products did we buy during the week? Depending of the received invoices

CQ75. How many products did we buy during the month? Depending of the received invoices

Specific Competency questions related to currencies:

CQ76. In what currency are the receivers paying in invoice X? EURO

CQ77. In what currency want the emitters to be pay in invoice X? EURO

CQ78. What taxes are applied in the invoice X? IVA, IGIC, or RE

CQ79. How much is the amount of the taxes in invoice X? 4%, 16% ….

CQ80. How much is the amount without taxes in invoice X? Answered in the received invoice

CQ81. How much is the amount with taxes in invoice X? Answered in the received invoice

CQ82. How much is the total amount in invoice X? Total line amount – Total Discounts + Taxable amount

CQ83. How much are the taxable base in invoice X? Answered in the received invoice

Competency questions in each group and between groups could be composed into more general questions. The following non-exhaustive list presents some examples.

Composed competency questions that use:

CQ84. Given a set of invoices of different companies, Is possible to identify/different the common concepts used? Yes

CQ85. Given a set of invoices of the same company, what format it is used? In the same company, the same format: XML, CVS, FLF, etc.

CQ86. Given a set of invoices of different companies, what are the common elements in these invoices? The elements are all common

CQ87. Given the information of a company, what products did it buy? Answered in the received invoice

CQ88. Given a set of invoices from different wholesaler, laboratories or providers, Is it possible to determine the main differences? Yes, but there are just a few of them.

CQ89. Given the information of a company, what products have it sell? Depending of the received invoices

CQ90. Given the information of a company, what products have it bought? Depending of the received invoices

CQ91. Given the information of a product, how many units have been sell? Depending of the received invoices

9. ANNEX III: Activities used in the development of the Nomenclature ontology network

Next table shows a summary of selected activities identified in the Semantic Nomenclature case study.

	Required	*If Applicable*	*Selected*
Ontology Aligning		X	
Ontology Annotation		X	
Ontology Assessment	X		*X*
Ontology Conceptualization	X		*X*
Ontology Configuration Management	X		*X*
Ontology Control	X		
Ontology Customization		X	
Ontology Documentation	X		*X*
Ontology Enrichment		X	*X*
Ontology Environment Study	X		*X*
Ontology Evaluation	X		*X*
Ontology Evolution		X	
Ontology Feasibility Study		X	
Ontology Formalization	X		*X*
Ontology Forward Engineering		X	
Ontology Implementation	X		*X*
Ontology Integration	X		*X*
Knowledge Acquisition for Ontologies		X	*X*
Ontology Learning		X	
Ontology Localization		X	*X*
Ontology Matching		X	
Ontology Modification		X	
Ontology Modularization		X	
Ontology Module Extraction		X	

NeOn

Ontology Population		X	*X*
Ontology Quality Assurance	X		*X*
Ontology Reengineering		X	*X*
Ontology Reuse	X		*X*
Ontology Scheduling		X	
Ontology Search	X		*X*
Ontology Selection	X		*X*
Ontology Specification	X		*X*
Ontology Summarization		X	
Ontology Translation		X	
Ontology Upgrade		X	
Ontology Validation	X		*X*
Ontology Verification	X		*X*

Table 15: Activities used in the Nomenclature use case

10. ANNEX IV: Semantic Nomenclature Competency Questions

Specific competency questions related with the Pharmaceutical Product are:

CQ1.   What is the drug commercial name?

CQ2.   What is the drug main active ingredient (molecule)?

CQ3.   What is its Spanish national code?

CQ4.   What is the drug registration date?

CQ5.   What is the drug withdrawal date?

CQ6.   What is the drug reference price?

CQ7.   How much euros costs it?

CQ8.   Which is the drug laboratory manufacturer?

CQ9.   Which one is the drug therapeutical WHO group?

CQ10.   What is the drug commercial price?

CQ11.   Which is the drug generic name?

CQ12.   Which is the drug defined daily doses DDDs?

CQ13.   Which is the drug composition?

CQ14.   Is it a narcotic?

CQ15.   Which are the drug contraindications?

CQ16.   What is the drug dosage?

CQ17.   Which method of administration has the drug?

CQ18.   What is the drug pharmaceutical form?

CQ19.   What indications does the drug have?

CQ20.   Is the drug financed by Social Security (Spanish Health Care System)?

CQ21.   How much money contributes the drug to the National Institute of Social Security?

CQ22.   What government therapeutical subgroup belongs?

CQ23.   What WHO therapeutical subgroup belongs?

CQ24.   What is the pharmacetutical product state?

CQ25.   Is the drug a bioequivalent medicine?

CQ26.   What is the last modification date?

CQ27.   Which is the unit content of the medicine?

CQ28.   What kind of medical content is?

CQ29.   Which dispensation condition has the drug?

Specific competency questions related with the Laboratory are:

CQ30.   Who is the contact of the laboratory?

CQ31.   Where is located the laboratory?

CQ32.   What is the national code of the laboratory?

CQ33.   What medicines are manufactured by the laboratory?

NeOn

Specific competency questions related with active ingredient

> CQ34.   What is the national code of the active ingredient?
>
> CQ35.   What is the main substance of the composition?
>
> CQ36.   What is the ATC code of the active ingredient?
>
> CQ37.   What is the WHO therapeutical subgroup of the active ingredient?
>
> CQ38.   What is the national code of the substance?
>
> CQ39.   Which pharmaceutical activity has the active ingredient?
>
> CQ40.   Which pharmaceutical indication has the active ingredient?
>
> CQ41.   Which pharmaceutical contraindication has the active ingredient?
>
> CQ42.   Which pharmaceutical precaution has the active ingredient?
>
> CQ43.   Which pharmaceutical activity has the active ingredient?
>
> CQ44.   What medical speciality has the medicine?
>
> CQ45.   What medical pathology is associated with the medicine?

After this description of specific competency questions, could be composed into more general questions that are answered by composing answers associated to the specific competency questions.

Composed competency questions that use a pharmacist for obtaining information about a drug:

> CQ46.   Given the information of a drug, (name, national code, price...), has the nomenclature another similar drug with a lower price?
>
> CQ47.   Given a particular active ingredient, which is the most appropriate drug?
>
> CQ48.   Given information from two particular drugs, is there any kind of incompatibility?
>
> CQ49.   Which are the latest drugs approved by the government?

Composed competency questions that use a GSCoP technician for obtaining information about new drugs

> CQ50.   Which are the latest drugs approved by the government?
>
> CQ51.   Which are the latest withdrawal drugs by the government?
>
> CQ52.   Given information from a specific pathology, which is the most appropriate and cheapest drug?
>
> CQ53.   Diseases outbreaks?
>
> CQ54.   Cheapest drug?
>
> CQ55.   Given a time interval (one week, one month...), which are the latest approved drugs?
>
> CQ56.   Given a time interval (one week, month...), which are the latest active ingredients approved?
>
> CQ57.   Which are the latest alerts about drugs in the last month?
>
> CQ58.   Which are the modified leaflets in the last month?
>
> CQ59.   Which are the new leaflets in the last week? (PDF,HTML...)
>
> CQ60.   Given a specific drug, are disposible the different leaflet (PDF,HTML)?
>
> CQ61.   Similar drugs search?

11. ANNEX V: Semantic Nomenclature Terminology, Glossary and Objects in the Universe of Discourse

Semantic Nomenclature Terminology

From the competency questions are extracted the terms that will be formally represented in the ontology as concepts, attributes and relations. A list of these identified terms are grouped by the most relevant terms and concepts extracted from the CQs.

- *Terms related to the pharmaceutical products*

    Pharmaceutical product

    - *Name of pharmaceutical product*
    - *National Code*
    - *Marketing Authorisation Holder*
    - *Marketing Authorisation Number*
    - *Date of first authorisation number/renewal of the authorisation*
    - *Date of revision of the text*
    - *Dosimetry*
    - *Reference Price*
    - *Laboratory price*
    - *Commercial price*

    Clinical Aspects

    - *Therapeutic indications*
    - *Posology*
    - *Method of administration*
    - *Contraindications*
    - *Contraindications*
    - *Special warnings*
    - *Precautions for use*
    - *Interactions with other medicine products*
    - *Pregnancy*
    - *Effects on ability to drive*
    - *Undesiderable effects*
    - *Overdose*

    Pharmaceuticals particulars

    - *List of excipients*
    - *Incompabilities*
    - *Shelf life*
    - *Precautions for storage*
    - *Precautions for disposal*

- *Terms related to the substances*

    Composition

    - *Excipients*

    Active Ingredient

    - *ATC code*

- *DDD*
- *WHO therapeutical subgroup*

Pharmaceutical Form

Pharmaco properties

- *Pharmacotherapeutic group*
- *ATC Code*

– *Terms related to Laboratories*

Laboratory

- *Laboratory location*
- *Laboratory contact*

Semantic Nomenclature Glossary

Concept	Definition in natural language	Competency Questions
Drug / Medication	A medication is any drug taken to cure or reduce the symptoms of an illness or ongoing medical condition. Commercial medications are produced by pharmaceutical companies and are often patented. Those that are not patented are called generic drugs.	CQ1, CQ32, CQ44, CQ45, CQ47, CQ49, CQ50. CQ53, CQ58, CQ59
Generic Drug	A drug which is produced and distributed without a brand name. A generic must contain the same active ingredients as the original formulation.	CQ11
Composition	The combining of distinct parts or elements to form a whole (related to ingredients)	CQ13
Excipients	An excipient is an inactive substance used as a carrier for the active ingredients of a medication	
Pharmaceutical Form	The way the drugs are delivered to the patient.	CQ18
Doses / Dosage	The smallest amount of a substance required to produce a measurable effect on a living organism	CQ16
Therapeutic indications	A valid reason to use a certain test, medication, procedure, or surgery. How substances interact with living organisms to produce a change in function.	CQ19, CQ38
Posology	Drug dosification	
Route of administration	the path by which a drug, fluid, poison or other substance is brought into contact with the body	CQ17
Contraindications	a condition or factor that increases the risks involved in using a particular drug, carrying out a medical procedure or engaging in a particular activity.	CQ15, CQ39
Warnings	A precautionary statement describing a potential hazard.	
Precautions for use	Refers to the avoiding of the pharmaceutical products	CQ40

Interactions / Interaction with other medicinal products	Action that occurs as two or more pharmaceutical products have an effect upon one another	CQ46
Undesirable effects	An unintended consequence specifically arising from drug therapy	
Overdose	An excessive dose, especially of a narcotic.	
Pharmaceutical activity	Describes the beneficial or adverse effects of a drug on living matter	CQ42
Pathology	Diagnosis of disease through examination of organs, tissues, cells and bodily fluids	CQ43, CQ50
Pharmacotherapeutic group	The third level of the ATC code based on the therapeutic/pharmacological subgroup of the substance	CQ9, CQ21,
ATC code	The Anatomical Therapeutic Chemical Classification System is used for the classification of drugs. It is controlled by the WHO Collaborating Centre for Drug Statistics Methodology. Drugs are divided into different groups according to the organ or system on which they act and/or their therapeutic and chemical characteristics.	CQ22, CQ25
List of excipients	List of ingredients of the drug	
Incompatibilities	Relation that exists when opposites cannot coexist, in this case, between drugs or active ingredients	
Shelf life	The length of time a product may be stored without becoming unsuitable for use or consumption.	
Special precautions for disposal	Recommendation and guidelines provided to the pharmacist for disposing special pharmaceutical products or drugs	CQ35
Date of first authorisation/renewal of the authorisation	Date when the drug is authorised by the pharmaceutical government agency for marketing in the pharmaceutical sector	CQ4, CQ28, CQ47, CQ53, CQ54
Date of withdrawal	Date when the drug is banned by the pharmaceutical governement agency for marketing in the pharmaceutical sector	CQ5, CQ49
Date of revision of the text	Date when the patient information leaflet (PIL) of a pharmaceutical product, containing information about medical conditions, available services and treatments, is checked by the pharmaceutical governement agency.	
Marketing authorisation holder	Company authorised for marketing the pharmaceutical product	
Aportacion (contribution)	Contribution of the drug to the National Health System	CQ20, CQ36
Substance asociation	Pharmaceutical substances grouped	
Laboratory	Drugs Laboratory and manufacturers	CQ8, CQ29, CQ31
Laboratory location	Where is located the Laboratory	CQ30
Reference price	Maximum quantity financed by the National Health System in the price of a pharmaceutical product	CQ6, CQ44, CQ50
Laboratory price	Price of the pharmaceutical product from the distributor or laboratory to the pharmacy	CQ7
Marketing Price	The price the laboratory recommends that the pharmacy sell it for	CQ10, CQ52

Regional Price	The price the regional government recommends that the pharmacy sell it for	
Active Ingredient	An active ingredient, also active pharmaceutical ingredient (or API), is the substance in a drug that is pharmaceutically active. Some medications may contain more than one active ingredient. The traditional word for the API is pharmacon	CQ2, CQ45, CQ54
Situation	Position or status with regard to conditions and circumstances. State	CQ23
Substance	Substance is any drug, chemical, or biologic entity, as well as any material capable of being self-administered or abused because of its physiologic or psychologic effects	CQ26, CQ27
Chemical substance	Material with a definite chemical composition	
National Drug Code	Pharmaceutical products are identified and reported using a unique number provided by the national pharmaceutical agency	CQ3
Dosage Unit	A measured quantity of a medicine	CQ16
Content unit	the way the drugs are delivered to the patient (capsule, ampule, cream…)	CQ33
DDD	Defined daily doses (DDDs) are a WHO statistical measure of drug consumption. DDDs are used to standardise the comparative usage of various drugs between themselves or between different healthcare environments.	CQ12
Pharmaceutical Equivalent	Pharmaceutical products which contains the same active ingredients and are equivalent in concentration, dosage and route of administration. Two pharmaceutical products would be expected to be pharmaceutical equivalent, for all intents and purposes, the same.	CQ24
Dosage form	A dosage form of a drug is traditionally composed of two things: The API, which is the drug itself; and an excipient	

Table 16: Semantic Nomenclature Glossary

Objects in the Universe of the Discourse

Next table shows some examples of objects, which are instances of the terms identified in the terminology: pharmaceutical products, laboratories, active ingredients, pharmaceutical form… These objects are useful in the selection of the standards to be reused for building the Nomenclature Reference Ontology.

Pharmaceutical Product	Active Ingredient	Laboratory	Pharmaceutical Form
GRADALIN C 500MG 20 CAPSULAS	IBUPROFENO	SERVIER S.A.	INYECTABLE GENERAL
GRADALIN COB12 2MG 12 CAPSULAS	DICLOFENACO	AVENTIS PHARMA, S.A.	INYECTABLE PERFUSIÓN
CODURETAS 20 GRAGEAS	BUTIBUFENO	INOFARMA	HEMODIÁLISIS
REGULATEN 400MG 56 COMPRIMIDOS	PENICILAMINA	PHARMACIA IBERIA, S.A.	COMPRIMIDOS

RECUBIERTOS			
GRIPESOL-S 15 COMPRIMIDOS	NIFLUMICO ACIDO	BAYER DIAGNOSTICOS S.A.	COMPRIMIDOS LIBERACIÓN RETARDADA
RASAL 500MG 50 CAPSULAS	MABUPROFENO	MENARINI,S.A.	COMPRIMIDOS RECUBIERTOS
GROVIXIM 30 CAPSULAS	GALAMINA	IFARMAX S.A.	COMPRIMIDOS EFERVESCENTES
TETRA HUBBER 8 GRAGEAS	TETRAZEPAM	BEECHAM, S.A.	COMPRIMIDOS MASTICABLES
PLENACICLINA '250' 8 CAPSULAS	TILUDRONICO ACIDO	MURILLO BENEDICTO	CÁPSULAS
OMEPRAZOL NORMON 20MG 28 CAPSULAS EFG	PROCAINA	NORMON S.A.	SOLUCIÓN/SUSPENSIÓN ORAL
TEUTIS 150ML SOLUCION	KETAMINA	CUSI S.A.	POLVO/GRANULADO ORAL
THROMBOCID 20G POMADA	CLOTIAZEPAM	OFTALMISO, S.L.	POLVO/GRANULADO EFERVESCENTE
AMOXIC/CLAVUL SANDOZ 500/125MG 12 COMPRIM REC EFG	TRIAZOLAM	NESTLE A.E.P.A.	GEL/PASTA/LÍQUIDO BUCAL
TIONER 20 COMPRIMIDOS	DOXEPINA	PFIZER, S.A.	INHALACIÓN PULMONAR
STATICUM 5MG 30 COMPRIMIDOS	OXITRIPTAN	PELLETIER	PRODUCTO DIETOTERÁPICO
STATROL 5ML COLIRIO ESTERIL	ANFETAMINA	PEDEMONTE	LÍQUIDO USO TÓPICO
STAXIDIN 30G POMADA	PIRACETAM	INKEYSA S.A.	SÓLIDO USO TÓPICO
GENPROL 40MG 28 COMPRIMIDOS	FIPEXIDA	PENTAFARM S.A.	APÓSITO
STOMOSAN 70ML SOLUCION	METADONA	PONS	SUPOSITORIO
DIAZEPAN LEO 2MG 100 COMPRIMIDOS	ACAMPROSATO	REIG JOFRE S.A.	PRODUCTO USO NASAL
VASPIT 0.75% 60G CREMA	PILOCARPINA	CRISOL S.A.	PRODUCTO USO BUCAL TÓPICO
DERGOM 30G POMADA	BETANECOL	ROCHE FARMA, S.A.	LÍQUIDO OFTÁLMICO
DEMETRIN 10MG 20 TABLETAS	RETINOL	S.A.L.V.A.T., S.A.	SEMISÓLIDO OFTÁLMICO
DEFLUINA 20 COMPRIMIDOS	FENOTEROL	CASA SANTIVERI S.A.	IMPLANTE OFTÁLMICO
TERAZOSINA MABO 5MG 30 COMPRIMIDOS EFG	MESNA	NOVARTIS FARMACEUTICA S.A.	RADIOFÁRMACOS
GELOCATIL INFANTIL 100MG/ML 30ML	DIHIDROCODEINA	SCHERING ESPAÑA S.A.	PREPARADOS URETRALES

SOLUCION ORAL			
OPTIRAY 300 ULTRAJECT 636MG/ML 50ML JE PRE	CLOBUTINOL	BEIERSDORF, S.A.	CONSERVACIÓN ÓRGANOS
PRED - FORTE 1% COLIRIO 5 ML	ZIPEPROL	TORLAN S.A.	POLVO/GRANULADO GASTRORRESISTENTE
FLUIMUCIL 100MG/5ML 200ML JARABE	ASTEMIZOL	FARMAPROS	INHALACIÓN ENDOTRAQUEOPULMONAR
TIONER 100MG 12 SUPOSITORIOS	CLEMIZOL	INSTITUTO GRIFOLS S.A.	HEMOFILTRACIÓN
SONDA VESICAL NELATON	DOXAPRAM	PHARMASOL S.A.	POLVO/GRANULADO LIBERACIÓN PROLONGADA
APOSITO ESTERIL	GENTAMICINA	VINSI	SEMISÓLIDO RECTAL
ALGODON ARROLLADO MEZCLA	DICLOFENACO	VIÑAS S.A.	SÓLIDO INTRAUTERINO
ESIMIL 20 COMPRIMIDOS	ACETAZOLAMIDA	KENFARMA S.A.	COMPRIMIDOS SUBLINGUALES
ZOLBEN 500MG 30 COMPRIMIDOS	GLICERINA	NOVO ESPAÑA, S.A.	COMPRIMIDOS DISPERSABLES
SYMMETREL 100MG 60 GRAGEAS	AZELASTINA	SMITHKLINE, S.A.	PRODUCTO PARA EL BAÑO
GAMASOL 90% 30ML AEROSOL	NISTATINA	BIOTERAX S.A.	COMPRIMIDOS BUCODISPERSABLES/LIOTABS
ALCORTEN 30G CREMA	TRETINOINA	INTERPHARMA S.A.	PULVERIZACIÓN ÓTICA

Table 17: Nomenclature objects in the universe of the discourse