



**NeOn: Lifecycle Support for Networked Ontologies**

**Integrated Project (IST-2005-027595)**

**Priority: IST-2004-2.4.7 – “Semantic-based knowledge and content systems”**

---

### **D8.3.2 Revision of ontologies for Semantic Nomenclature: pharmaceutical networked ontologies**

---

**Deliverable Co-ordinator: German Herrero Cárcel**

**Deliverable Co-ordinating Institution: Atos Origin (ATOS)**

**Other Authors: Tomás Pariente (ATOS)**

Document Identifier:	NEON/2009/D8.3.2/v1.1	Date due:	November 30, 2008
Class Deliverable:	NEON EU-IST-2005-027595	Submission date:	September 15, 2009
Project start date:	March 1, 2006	Version:	v1.1
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:

<p><b>Open University (OU) – Coordinator</b>  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</p>	<p><b>Universität Karlsruhe – TH (UKARL)</b>  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</p>
<p><b>Universidad Politécnica de Madrid (UPM)</b>  Campus de Montegancedo  28660 Boadilla del Monte  Spain  Contact person: Asunción Gómez Pérez  E-mail address: asun@fi.upm.es</p>	<p><b>Software AG (SAG)</b>  Uhlandstrasse 12  64297 Darmstadt  Germany  Contact person: Walter Waterfeld  E-mail address:  walter.waterfeld@softwareag.com</p>
<p><b>Intelligent Software Components S.A. (ISOCO)</b>  Calle de Pedro de Valdivia 10  28006 Madrid  Spain  Contact person: Jesús Contreras  E-mail address: jcontreras@isoco.com</p>	<p><b>Institut 'Jožef Stefan' (JSI)</b>  Jamova 39  SI-1000 Ljubljana  Slovenia  Contact person: Marko Grobelnik  E-mail address: marko.grobelnik@ijs.si</p>
<p><b>Institut National de Recherche en Informatique et en Automatique (INRIA)</b>  ZIRST – 655 avenue de l'Europe  Montbonnot Saint Martin  38334 Saint-Ismier, France  Contact person: Jérôme Euzenat  E-mail address: jerome.euzenat@inrialpes.fr</p>	<p><b>University of Sheffield (USFD)</b>  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</p>
<p><b>Universität Koblenz-Landau (UKO-LD)</b>  Universitätsstrasse 1  56070 Koblenz  Germany  Contact person: Steffen Staab  E-mail address: staab@uni-koblenz.de</p>	<p><b>Consiglio Nazionale delle Ricerche (CNR)</b>  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</p>
<p><b>Ontoprise GmbH. (ONTO)</b>  Amalienbadstr. 36  (Raumfabrik 29)  76227 Karlsruhe , Germany  Contact person: Jürgen Angele  E-mail address: angele@ontoprise.de</p>	<p><b>Food and Agriculture Organization of the United Nations (FAO)</b>  Viale delle Terme di Caracalla 1  00100 Rome, Italy  Contact person: Marta Iglesias  E-mail address: marta.iglesias@fao.org</p>
<p><b>Atos Origin S.A. (ATOS)</b>  Calle de Albarracín, 25  28037 Madrid  Spain  Contact person: Tomás Pariente Lobo  E-mail address:  tomas.parietelobo@atosorigin.com</p>	<p><b>Laboratorios KIN, S.A. (KIN)</b>  C/Ciudad de Granada, 123  08018 Barcelona  Spain  Contact person: Antonio López  E-mail address: alopez@kin.es</p>

## 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:

ATOS

## Change Log

Version	Date	Amended by	Changes
0.1	15/10/08	Germán Herrero	Layout and table of contents
0.2	22/11/08	Germán Herrero	Added section 5
0.3	01/12/08	Germán Herrero	Added Sections 2 & 3
0.4	18/12/2008	Germán Herrero	Added section 4 & Conclusions
0.5	21/12/2008	Germán Herrero	Added Executive Summary & Conclusions
0.6	23/12/2008	Germán Herrero	Added URI's of the ontologies and paragraphs review before QA Added Section of References
0.7	09/12/2009	Germán Herrero	Grammar revision and updates after QA
1.1	10/09/2009	Germán Herrero	Added section 5.4 NeOn Toolkit case study experience after 3 <sup>rd</sup> review

## Executive Summary

The Semantic Nomenclature case study is based on the semantic interoperability of pharmaceutical product information between different ontological and non-ontological resources in the Spanish pharmaceutical scenario. This document describes a review of the ontology requirements detected in the case study and how NeOn technologies help us to develop and deploy an ontology network that handles the requirements and address the problems detected in the scope of the case study scenario. For this purpose, this document describes and discusses the ontologies developed for use within the Semantic Nomenclature prototype and introduces an ontology network that specifies and describes the pharmaceutical product information and enables the interoperability between the different resources. Also, is provided a review and analysis of the inventory of knowledge resources and medical terminologies that are useful for the case study.

## Table of Contents

NeOn Consortium	2
Work package participants .....	3
Change Log .....	3
Executive Summary .....	3
Table of Contents .....	4
List of figures .....	5
List of tables .....	6
1. Introduction .....	7
2. Semantic Interoperability in the Semantic Nomenclature Case Study .....	8
2.1 Semantic Interoperability .....	8
2.2 SemanticHEALTH recommendations .....	10
2.3 Review of the intended goals of Semantic Nomenclature case study .....	11
2.4 Review of NeOn technologies .....	12
3. Applying the Methodology to the Semantic Nomenclature Case Study .....	14
3.1 Semantic Nomenclature methodological approach to the ontology development process .....	14
3.2 Methodological approach to the ontology development process .....	15
3.3 Review of Ontology requirements for the Semantic Nomenclature .....	17
3.4 Architecture of the Semantic Nomenclature network of ontologies .....	20
4. Review Inventory and Analysis of Knowledge Resources and Vocabulary	23
4.1 Drug Information: Commercial Drug Databases and National Standards .....	25
4.2 Ontological resources .....	26
4.3 Non Ontological resources .....	29
5. Semantic Nomenclature Network of Ontologies .....	31
5.1 Application Ontologies .....	31
5.2 Domain Ontologies .....	36
5.3 General Ontologies .....	44
5.4 NeOn Toolkit Semantic Nomenclature Experience .....	45
6. Conclusions .....	48
7. References .....	49

## List of figures

Figure 1: NeOn Scenarios for Building Ontology Network [5].....	14
Figure 2: Ontological activities.....	16
Figure 3: Modular approach of the Nomenclature Ontology Network.....	20
Figure 4: Nomenclature Ontology Network.....	22
Figure 5: Correspondence analysis between drug information sources and dimensions of drug information (in four domains pharmaceutical, chemical, biological and clinical) [9].....	23
Figure 6: DrugOnt Schema.....	27
Figure 7: Pharmaceutical product concept in the Reference ontology .....	32
Figure 8: Semantic Nomenclature Reference Ontology relationships .....	33
Figure 9: Digitalis ontology relationships .....	34
Figure 10: Pharmaceutical Product BOTPlus Hierarchy.....	35
Figure 11: BOTPlus ontology relationship .....	36
Figure 12: ATC relationship.....	37
Figure 13: RxNorm ontology relationships.....	38
Figure 14: UMLS Semantic Types relationships.....	40
Figure 15: Drug-Onto Schema relationship .....	41
Figure 16: SPC Ontology relationship .....	44

## List of tables

Table 1: Ontological resources .....	28
Table 2: Non-ontological resources .....	30
Table 3: General ontologies for the Semantic Nomenclature case study .....	45
Table 4: NeOn Toolkit plugins used in the ontology network development .....	47

## 1. Introduction

Starting from the initial vision captured in [1], in this deliverable is reviewed and analyzed the semantic interoperability framework for the Semantic Nomenclature case study within technological and requirements perspective, in the light of improve the ontology network that handles the problems detected in the scenario. Based on the SemanticHEALTH project recommendations, we introduce the semantic interoperability as a declared and desirable feature of health-related information systems. Nowadays, only limited evidence exists that current systems are actually fully interoperable. However, we want to achieve semantic interoperability with our information systems and pharmaceutical terminologies involved in the Semantic Nomenclature scenario.

This Deliverable 8.3.2 discusses and describes a possible solution for the case study scenario, in order to advance semantic interoperability of information systems in the pharmaceutical product description field. Firstly, we introduce the Semantic interoperability concept, based on the SemanticHEALTH descriptions. Moreover, is described how semantic interoperability between our pharmaceutical information systems could be achieved in the case study, based on a review the intended goal of the case study and how an ontology network performed using the NeOn technologies covers the ontology requirements detected in the case study.

Secondly, we describe the Semantic Nomenclature Methodological approach to the ontology development process in its second iteration of the ontology network lifecycle. According to the new requirements of domain experts and the SemanticHEALTH recommendations, we generate a new version of the Semantic Nomenclature ontology network that should cover the lack of communication with the standard terminologies, and in case of the Spanish pharmaceutical sector, should add the disambiguated notion of clinical drug and branded drug. Moreover, this document contains a review of the inventory and analysis of the ontological and non-ontological resources which have been selected for reuse in the context of our ontologies.

Finally, the ontologies that are involved in the case study are described, from a the complete revision of the Reference ontology to new ontologies as the SPC Ontology and how they are intended to be exploited by the users of the Semantic Nomenclature prototype.

## 2. Semantic Interoperability in the Semantic Nomenclature Case Study

As is detailed in the description of the case study, one of the main problems in the pharmaceutical sector is the lack of a unified and common classification system and description of the pharmaceutical products. Several systems classifications of pharmaceutical product information are available in the Spanish pharmaceutical sector, managed by its corresponding entities like professional association (General Spanish Council of Pharmacists), government entities (Ministry of Health, Spanish Agency of Pharmaceutical Product, laboratories...) Despite of the different tools provided, there is a lack of classification of clinical drugs, and this is crucial for hospitals, one of the main actors of the health domain. Hospitals generate their own drug information storing systems, due to the need to classify pharmaceutical products based on clinical aspects like the therapeutically use or dosage, not by the branded name of the product.

The aim of the case study is to provide a solution to the lack of communication between the different sources of information. This goal is known as semantic interoperability, and is needed if computational services have to be able to interpret safely pharmaceutical data that has been integrated from diverse sources. The solution suggested in the Semantic Nomenclature is developing a network of ontologies, which allow define and maintain their own model to each actor of the domain and linking to the rest of the models, providing a better interoperability between the sources.

Furthermore, the ontology network allows enrich the ontology models of each actor, and also, can be connected with new models or ontologies that will appear in the eHealth domain, both national and international. Moreover, a network of ontologies improves the efficiency and quality of the data provided to the pharmaceutical domain.

### 2.1 Semantic Interoperability

Based on a agreement of meaning of interoperability as the ability, to share data whose unambiguously clear, its context is understood, it can be used for whatever purpose and the receiver is not previously known to the sender. Interoperability requires agreement on meanings and labels for those meanings – on ontology and lexicons, which together we label as terminology. The primary goal of ontologies and terminologies for interoperability is to enable the faithful exchange of meaning between machines and between machines and people.

Pharmaceutical professionals require access for all the information about the pharmaceutical products available to dispense in the pharmacies or hospitals. Also, the pharmacists must know the latest decisions about the withdrawals or recommendations of pharmaceutical products, because they have more direct contact with the patients and must apply the latest recommendations or alerts in the dispensation of drugs. So, the information of the pharmaceutical product should be shared and consistent by all actors in the sector. This goal could be carrying out if the information services are able to interpret the data and process it safely, and we can achieve using the semantic interoperability provided by the ontologies and technologies of the semantic web. So, for Semantic interoperability requirements, we need:

1. Enable the safe, meaningful sharing and combining of pharmaceutical data between heterogeneous systems;

2. Enable the consistent use of modern terminology systems and medical knowledge resources;
3. Ensure the necessary data quality and consistency to enable rigorous uses of heterogeneous data.

The project SemanticHEALTH SSA develops a European and global roadmap for research in health-ICT, focusing on semantic interoperability issues of e-Health systems and infrastructures. The roadmap will be based on consensus of the research community, and validated by stakeholders, industry and Member State health authorities. SemanticHEALTH applies the following Interoperability definition [3] :

*Health system interoperability is the ability, facilitated by ICT applications and systems,*

- *to exchange, understand and act on citizens/patient and other health related information and knowledge*
- *among linguistically and culturally disparate clinicians, patients and other actors and organizations*
- *Within and across health system jurisdictions in a collaborative manner.*

From a methodology point of view, SemanticHealth describes that may be different approaches to achieve *semantic interoperability* in the eHealth domain [3] :

1. Everyone adopts *a single, core model*. This – more than likely – becomes a long and tedious, probably even unsuccessful process due to disagreement on key aspects of such a central model (see also standardization process for standards in the health sector).
2. Everyone has its *own model* but follows *interchange standards* (communication, messaging) between the models. An essential prerequisite for this scenario includes bilateral and/or multilateral agreements between the participating parties.
3. Everyone agrees on *common data elements* with systematic unambiguous formats, e.g. data descriptions (data types, terminologies, coding), meta data and information models.
4. Everyone uses a predefined *knowledge representation framework* (classes, attributes, definitions, identification principles) and *inference mechanisms* (inclusions, exceptions, constraints, reasoning etc)
5. Other

In case of the Semantic Nomenclature case study, we define an ontology network where each actor defines its own model as ontology and all ontologies are interconnected and mapped between them to share the information. In [1] we described the first version of the ontologies which compose the network in the case study. In this deliverable is defined a study about terminologies or ontologies that can be added to that already included on the network. The goal of ontologies and terminologies that are included in the network is to make it easier to build systems that successfully exchange meanings in the case study scenario.

## 2.2 SemanticHEALTH recommendations

The ontologies and terminologies are part of the strategy to achieve semantic interoperability in the pharmaceutical domain, and in our case, in the semantic nomenclature scenario. As part of the objective of the SemanticHealth project, they provide some recommendations about the development of ontologies in the Health domain, and how to carry out the semantic interoperability between the different sources. Some recommendations in the research area of terminologies and ontologies are the following[3] [2] :

- » **Ontology modularization:** modularization of large ontologies provides a logic separation of the objects of the model, and it is useful in order to add coherent and consistent extensions of the ontologies. Also, modularization makes possible different views of the model which facilitates the localization of the ontologies. Another key challenge is the re-use of part of ontologies using modularization techniques.
- » **Dynamics (just-in-time):** The rate of change of medicine dictates rapid change. Large standardized systems are usually slow to change and adapt. Systems will not long remain interoperable unless the standards and terminologies evolve quickly. One of the main objectives is provide the mechanisms for reduce the time needed to adapt the ontologies/terminologies to the change from the "alive" objects or information of the domain
- » **Change consequences prediction:** one aspect which is distinctive of the Health domain is the Change. The information systems should learn from the different changes produced in the domain and predict what consequences should appear when one object changes or new variables are added.
- » **Versioning:** the health domain, including the pharmaceutical Managing versioning and tracking identifiers is a major part of any system that includes a controlled vocabulary. Good identifier tracking is one of the major features of successful terminology systems
- » **Collaborative design:** Collaborative open development of terminologies remains a dream rather than a reality in this moment. All actors in the health domain must be more heavily involved. They are the ones who must deploy the systems; they are the ones who must 'interoperate. So, it is needed that the tools for modeling ontologies should add collaborative and social characteristics which involve the end-users in the phase of building ontology models. Argumentation mechanisms are also recommendable for tracking the changes produced in the collaborative design.
- » **Localization & Multilinguality:** Despite major international efforts, most development has been within national or linguistic groups. Multilingual and multicultural systems are of particular importance to Europe (Interoperability across Europe requires interoperability across languages and cultures). Furthermore, on the one hand, clinical medicine and clinical research become ever more international, while practical care delivery at the bedside remains largely national, or more properly, local.
- » **Access Ontology Interface (APIs):** Until recently there was no standard syntax or API, so that every tool was limited to a single reasoner and experimentation with

alternatives was difficult. The emergence of the W3C standard ontology language, OWL, has changed this situation radically. A large and growing community is now developing tools and underlying software to a single standard.

» **Methodology:** The large terminologies in the eHealth cause scaling problems and consistency maintenance difficulties. More methodologies for specify terminologies and ontologies should be available, which help to end-users in the design of the models, providing the adequate solution depending the requirements and reducing the cost of developing new content or generate applications, which should be developed more quickly than currently.

» **Involvement of ontologies in the information systems:** the new ontologies developed based on a re-engineering of the different terminologies available in the Health sector, should be part of the information systems and applications developed around the data because they are the key to enable the semantic interoperability between the different sources of information.

» **Tools for a successful development and maintenance:** The range of tools available to develop terminologies remains extremely limited for authoring, maintaining, and deploying terminologies, let alone for building applications that use terminologies. On other hand, a high number of open and freely available software for developing and maintaining ontologies is appearing, and should help to reduce the complexity of deploying ontologies.

### **2.3 Review of the intended goals of Semantic Nomenclature case study**

The main problem detected in the Spanish pharmaceutical sector is the lack of communication between the stakeholders involved in the description of pharmaceutical products. This lack provokes an upper cost to the different pharmaceutical entities when they try to keep up-date their information systems, and the inconsistency of information provided to the pharmacists and end-users. One of the main goals in the Semantic Nomenclature is the development of the Nomenclature ontology network that will be the key to enable the semantic interoperability between the source information involved in the Spanish pharmaceutical domain, and connect with the terminologies and pharmaceutical product classification adopted by the international community.

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 [1] & [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.

## 2.4 Review of NeOn technologies

Based on the requirements described by the SemanticHealth project to achieve the semantic interoperability in any case study in the Health domain, we review the technologies provided by NeOn that help us to solve the case study proposed in the Semantic Nomenclature:

- » **Ontology modularization:** the NeOn Toolkit (NTK) core provides means to select a part of an ontology model to extract it as a module, maintaining the consistency of the module with the objective to reuse it in other applications or to extend it. This functionality is being researched and providing efficient techniques to modularize ontologies in the scope of the WP1 and WP3. Also, using the NTK we can extend the modules of the ontologies and use operators to combine different parts of the ontologies.
- » **Dynamics (just-in-time):** One of the core objectives of the NTK is the dynamics of the ontology models. The NeOn project is researching and providing a solution to the dynamics of the ontologies and network of ontologies due to nature of the ontologies themselves. Ontologies and the corresponding metadata will evolve and change at a fast rate and NeOn provides advanced methods for change propagation between networked ontologies and metadata. Some Plugins are available in the NTK which cover this requirement, like the Workflow support plugin (tracks the ontology changes and then manipulates these changes according to the role of a user), Change Capturing plugin (capture ontology changes from the NTK editor and log them into Oyster distributed registry and visualize them).
- » **Versioning:** NeOn provides advanced methods for support and management of different versions of a model and advanced services to support the evolution of ontologies
- » **Collaborative design:** Collaborative tools are a major goal of both the NCBO in the US and the NEON project in Europe. In NeOn different tools are available to achieve this goal: C-ODO, Ontology Design Patterns... Moreover, argumentation requirement is covered by the Cicero plugin developed for the NTK, which the main purpose of the Cicero plugin is to keep track of discussions between the developers and users of the ontology. While the actual discussions are held in the Cicero-Wiki on a central server, the toolkit plugin allows for establishing links between elements in the ontology (e.g. classes or properties) and discussions that influenced their design. These discussions are then used by the ontology developers for understanding the design rationale of specific ontology elements.
- » **Localization & Multilinguality:** Nowadays, multilinguality in ontologies is demanded by institutions worldwide with a huge number of resources available in different languages. To solve this problem NeOn proposes the LabelTranslator plugin, a system that automatically localizes ontologies. LabelTranslator takes as input an ontology whose labels are described in a source natural language and obtains the most probable translation into a target natural language of each ontology label.
- » **Access Ontology Interface (APIs):** The NTK core provides interfaces, APIs and methods at infrastructure level to provide access to the different objects of the ontologies in any application or information system.

» Methodology: In the context of WP5 in NeOn a methodology for covering the missing points in developing networked ontologies is being developed and represented in [6]. This methodology not only covers the ontology network specification, it also recommends the best practices in the development of applications based on the ontology network. Also, NeOn provides help and recommendations for the ontology formalization using the ontology design patterns.

» Tools for a successful development and maintenance: The NTK is one of the main results of the NeOn project. The NeOn Toolkit is an extensible Ontology Engineering Environment. It is part of the reference implementation of the NeOn architecture. It contains plugins for ontology management and visualization. The core features include: Basic Editing: Editing Schema, Visualization/Browsing, Import/Export: F-Logic, (subsets of) RDF(S) and OWL; and also a number of commercial plugins extend the toolkit by various functionalities, including: Rule Support (Graphical/Textual editing, debugging), Mediation (Graphical Mapping Editor, life-interpretation of mappings), Database Integration (Database schema import, database-access), Queries (Query-Editor and persistent queries)

### 3. Applying the Methodology to the Semantic Nomenclature Case Study

#### 3.1 Semantic Nomenclature methodological approach to the ontology development process

In the context of WP5 in NeOn a methodology for covering the missing points in developing networked ontologies is being developed and represented in [6]. In this section we explain the main points of this methodology and how it is applied to the Semantic Nomenclature case study in its second iteration.

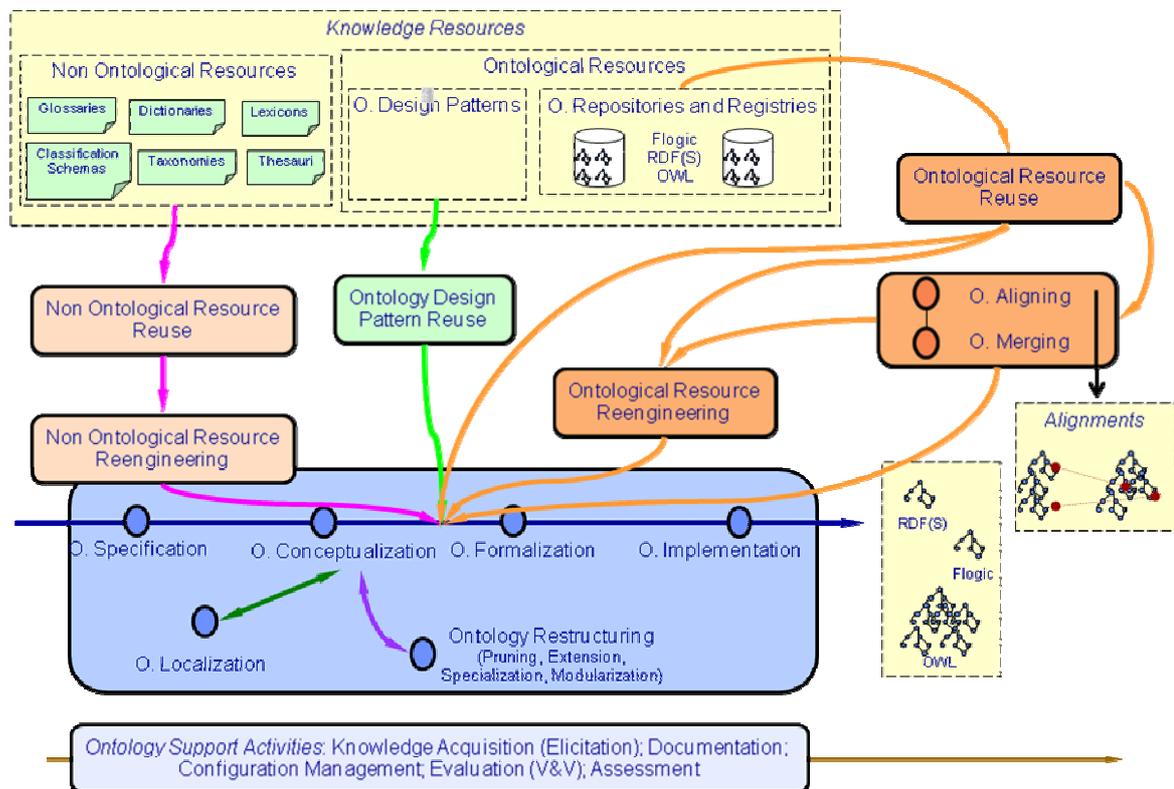


Figure 1: NeOn Scenarios for Building Ontology Network [5]

In [1] were managed and applied the first recommendations extracted from the NeOn methodology in WP5. In this first version of the ontologies which are involved in the case study, some activities of the methodology had been performed in the Semantic Nomenclature but the importance of them varied depending on the type of ontology developed (from scratch, re-use, re-engineering...).

In the following sections we describe how the NeOn methodology has been applied in this new review of the ontology requirements in the case study, and how the lessons learned from the first iteration are applied to evolve the Semantic Nomenclature ontology network.

### 3.2 Methodological approach to the ontology development process

In the first version of the ontologies, we needed to develop some ontologies from scratch (ATC, Reference) and in other cases we needed to re-engineering some ontologies (BOTPlus, Digitalis) or select and re-used other ontologies (time, geographical). So, the workplan followed to develop the main ontologies of the case study included tasks like knowledge acquisition, ontology conceptualization, validation and ontology specialization.

Based on the guidelines and suggestions provided by UPM, the first step was identify the iterative/incremental ontology network lifecycle model as the most appropriate to the case study. The main goal of the ontology network does not change (interoperability between information sources of drug information), but the dynamic nature of the pharmaceutical domain need to produce intermediate results and techniques to include or change models in the ontology network.

For this second iteration of the case study, we do not change the methodological approach to the ontology network development process, despite new ontology requirements are added to the ontologies, the main objective of this second iteration of the ontology network is add new ontologies and improve the existing ontologies. Maybe, this second iteration is closer to the incremental approach “produced & deliver” new ontologies to the case study, but the already available ones are improved (Reference) to enable the interoperability goal in the pharmaceutical domain.

In this second iteration, new ontologies are developed from scratch (SPC), others are included and re-used in the ontology network (MeSH...) and the Reference ontology is improved and aligned to other ontologies to be the core of the ontology network and the application ontology of the Semantic Nomenclature prototype.

#### Ontological activities performed

As is detailed in [7], the goal of the Semantic Nomenclature is focused on the integration of different and heterogeneous pharmaceutical product information repositories. For this purpose, we introduce in [1] an architecture of the ontology network consisted of a reference ontology as the core component of the network, supported by general ontologies (time, location...). Then, we connected via mappings the Reference ontology to medical classification ontologies (ATC) and a set of local ontologies which describe the pharmaceutical product information in its own way of knowledge.

For the second iteration of the ontology network, the objectives are add new resources that enrich the ontology network (connecting to international classification and terminologies) and improve the existing ones. So, the activities performed in the Semantic Nomenclature vary depending on the type of ontology developed. But in this case, activities like Select Standards, Ontology Reuse and Ontology evaluation are the key relevant. Also, activities like ontology environment study, ontology enrichment are important to reach the second iteration. Figure 2 shows the main activities followed by the Semantic Nomenclature case study and included in the incremental lifecycle model.

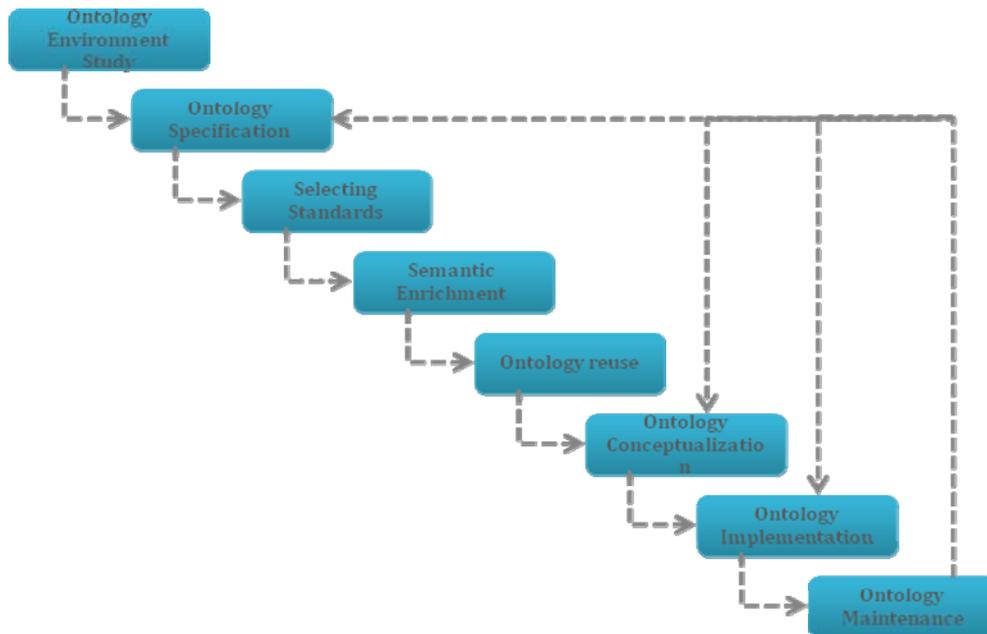


Figure 2: Ontological activities

**Knowledge Acquisition and Ontology Elicitation:** the pharmaceutical domain was specified and studied in [d811]. But for this second iteration, we have added new different interviews with domain experts from Hospitals, which added new requirements to the problems of integration between the different actors when exchange information about drugs, and the difficulties to have a classification of clinical drugs (not branded drugs) in hospitals. From these interviews we identify the problem of ambiguity of how to difference between clinical drugs and branded drugs in Spain.

**Select the standards that cover most of the identified necessities.** Pharmaceutical classification systems and different thesaurus, taxonomies and vocabularies were identified in [d811] and mainly in [1]. After the second review of the NeOn project, for this second iteration, this is an important task on account of connect the Spanish pharmaceutical models with the main terminologies adopted by the international bodies in this domain. So, for this second iteration, medical vocabularies, terminologies and ontologies in the eHealth domain suggested by domain experts like Snomed, MeSH, RxNorm, DM+D are analyzed of how to connect and used in the case study. Also, new standards used for describing pharmaceutical products like the SPC template recommended by the WHO are added to the repository of resources. For this second iteration new local reliable sources of knowledge did not appear, as the government and official pharmaceutical resources used in the previous iteration, so in this deliverable we did not analyze them again. More detailed analysis of the resources is described in [1] and section 4.

**Semantic enrichment of the standards.** The new terminologies, medical vocabularies or documentation 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

**Evaluate Ontology Content.** Based on the new requirements and recommendations described by domain experts, an **Ontology Search** should be performed to find candidate ontologies about general domains as Time, Measures, and Geography to describe these kinds of characteristics in the ontology network. 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.

**Ontology Conceptualization, Ontology Formalization, Ontology Integration and Ontology Evaluation** of the Nomenclature ontology network. Previously, an ontology scheduling is proposed to check the tasks and milestones of the new iteration of the ontology network, including tasks for involve new ontologies (after ontology reuse or ontology reengineering) and tasks to specify new ontologies from scratch or versioned.

**Ontology Implementation** of the Nomenclature ontology network. The language selected to describe the resources is OWL. New ontologies are added from scratch (SPC Ontology) and other ontologies are evolved and versions (Reference Ontology).

**Maintenance Activities.** As part of the second iteration of the Semantic Nomenclature Ontology network, the ontologies evolve, like the Reference Ontology, according to the changes and suggestions given by domain experts from hospitals. Also, the Semantic Nomenclature case study should provide support and maintenance to the ontology network that will be delivered and used by the prototype at the end of the case study. According to the methodology, activities involved to this stage are **Ontology Documentation, Ontology Configuration Management, Ontology Assessment and Ontology Verification & Validation**

Other activity related with the ontology lifecycle and development of the Nomenclature ontology network is the **Ontology Localization**. As is described in section 2, the Multilinguality and localization in the health terminologies is one of the barriers that should be solved due to the nature of multiculturalism of Europe, and in case of Spain, the different dialects used in the regions.

### 3.3 Review of Ontology requirements for the Semantic Nomenclature

The specification activity states why the ontology is being built, what its intended uses are and who the end-users are. For the first iteration, we identified the intended uses of the ontology network in the case study and its users.

The ontology network described in the case study is motivated by scenarios described in [4], which are related to the application that will solve the interoperability between information repositories. Such scenarios describe a set of ontology requirements that we can resume as the ability to enable the interoperability and help in the lack of communication between the resources of the actors.

Also, the analysis of the scenarios and the problems in the current pharmaceutical domain provides who is going to finally use the ontology network developed in the case study. In [1], we identify the following users with intended purposes: Pharmacist (Navigation across the ontology network searching for information), GSCoP technician (Searching for new information drugs, ingredients, etc., and updating the BOTPlus database), Spanish Government (Analyzing drug information or updating content). One of the lessons learned in the last months is that we can add a new intended user: hospital professionals, which need new functionalities (searching clinical drugs, compatibilities...), that could be solved using the ontology network.

Besides the intended uses and users of the ontology network, in [1] we used the competency questions for specifying the ontology requirements and determine the scope of the ontology to be built. For this second iteration, this part of the methodology is not needed, due to previous competency questions covers the main goal of the case

study and the general architecture of the ontology network is not needed to be generated again, only evolved.

So, starting from the initial vision captured in the semantic interoperability description in the earlier section, a set of technological considerations of ontology requirements are reviewed in the light of desirable features of pharmaceutical product information systems. Based on the lessons learned in the first iteration and the recommendations provided by SemanticHEALTH project, the new requirements added from the hospital users, we can review and describe the key and most relevant ontology requirements for the Semantic Nomenclature described:

### **Reusability of modules**

In case of the large terminologies, some concrete parts or modules of the terminologies can be reused in other projects or purposes, and facilitates the localization of the ontology and provides a logic view of part of the objects. NeOn is providing tools and methods to enable the way of reusability of ontologies as modules. In the Semantic Nomenclature and in the eHealth terminologies this requirement is valuable, and in the Semantic Nomenclature could be identified some modules as time, geographical, units that are shared and reused in the reference model and in some large terminologies like MeSH or Snomed, we can extract only the part that is useful for the case study.

### **Mappings management**

The mapping management is one of the most frequent actions in this case study that could occur in the ontology editor role. In this domain, we can find links between drugs and disease, or drugs and active ingredients, or ingredients and therapeutic recommendations, but as mappings, we have different ontology models of the same concepts that should be mapped, mainly between the reference ontology against the application ontologies and the reference ontology against the main health terminologies. So, the NeOn toolkit should provide mechanisms for manual mapping, calculate mappings between objects of two ontologies and a graphical management of the mappings of the stakeholder ontologies (application mappings). The creation of mappings between ontologies in the Semantic Nomenclature case study could be found in [8].

### **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 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 modeled or reengineered, or when ontology should evolve in a new version or only in a new iteration.

### **Re-engineering Semantic Enrichment**

All the information of the pharmaceutical products are modeled and stored in different legacy systems and databases. Moreover, resources like thesaurus, terminologies and classifications or templates of documents for describe the pharmaceutical products 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, tools and methods to re-engineering of this kind of resources are useful to the ontologist to enrich the ontology network of the case study.

### **Multilinguality**

Due to the fact of the multilingualism in Spain (Spanish, Catalanian, Basque...) and the multicultural nature of Europe regions, the multilinguality and localization support is needed in the ontologies in the Health domain. In the Semantic Nomenclature we should provide a reference ontology adaptable to localization in any part. Also, we identify Multilinguality at three different levels: labels for concepts, relations metadata element and ontology content. The multilinguality of the ontologies of the case study should adapt the multilinguality model provided by NeOn. Also, NeOn should provide mechanisms to find and manage links between concepts from ontologies developed in different languages.

### **Ontology Population**

The data information about the pharmaceutical products is stored in several distributed databases. These databases are mainly Access DBs, and in this phase of the case study, is decided to leave the individual records where they are, and only support their integration if is necessarily any population. Some tools in NeOn like R2ODEMapster 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. Also, this technology is based on a description of mappings between relational and ontology elements and we can exploit them using a domain independent processor, both in real-time and design-time.

### **Clinical Drug vs. Branded Drug**

One of the main problems identified in the pharmaceutical sector is the differentiation and ambiguity between clinical drug and branded drug. The differentiation lies in that clinical drug is the drug used and dispensed to patients by doctors in hospitals and health centres, and the branded drug is the drug dispensed to clients in a pharmacy. The best way to visualize the difference is with an example: a patient with a knee illness or injury is admitted in a hospital, the doctor analyzes the problem and dispense a pill of ibuprofen 30mg as treatment to stop the inflammation and pain of the knee; Meanwhile, if we have a person with a sprained knee, he goes to the doctor's office which gives him a treatment based on ibuprofen to stop the inflammation, and in the pharmacy, they dispensed the branded drug Ibuprofen KERN PHARMA EFG effervescent granules 600 mg to the patient, which contains 20 pills of the drug.

The clinical drug is basically a drug depicted by an active ingredient in a concrete dosage and pharmaceutical form, which is dispensed and used in a hospital environment, whereas the branded drug are drugs (group of pastille, pill, tablet, syrup...) which are manufactured by a Laboratory to be sold in a pharmacy to the citizens.

As is described in the case study, one of the main problems in the pharmaceutical sector is the lack of a uniqueness pharmaceutical product classification. Also, in Spain some classifications of branded drugs are provided, which could be used by professionals in hospitals, due they not work with registered trademark of drugs, they need clinical drug classifications. Also, the international terminologies and classifications like ATC or Snomed provides this kind of classifications, so, in the Semantic Nomenclature case study we can cover the gap between clinical drugs and branded drugs classification mapping and linking the models.

### 3.4 Architecture of the Semantic Nomenclature network of ontologies

In the first iteration of the development of the ontologies in the case study was concluded, according with the methodology, that the use of a network of the ontologies seems a good solution to cover the lack of interoperability and communication between the sources in the pharmaceutical scenario. Moreover, in [1] was specified that the adoption of a single and globally semantic model for all actors is too expensive as solution for the semantic integration, due is too difficult to manage and maintain consistency in a large ontology, and are not provided facilities for maintaining contextualized and localized ontologies. Also, was conclusively shown that the Spanish pharmaceutical sector had a lack of a reference ontology or description about all the knowledge around the pharmaceutical products.

For this second iteration, according to the new requirements of domain experts and the SemanticHEALTH recommendations, the Semantic Nomenclature ontology network should cover the lack of communication with the standard terminologies, and in case of the Spanish pharmaceutical sector, should add the disambiguated notion of clinical drug and branded drug. Despite the new requirements and recommendations for the scenario, the use of a network of ontologies and development of reference ontology for enabling the semantic interoperability in the pharmaceutical sector seems to be a good solution.

This second iteration of the Nomenclature Ontology Network is organized in four levels, as the first iteration: the Representation Ontology (OWL), General Ontologies, Domain Ontologies and the Application Ontologies; but now a sublevel is included between Application and Domain level called Mapping level. Figure 3 shows the levels of the Nomenclature Ontology Network based on reusability and usability of the levels.

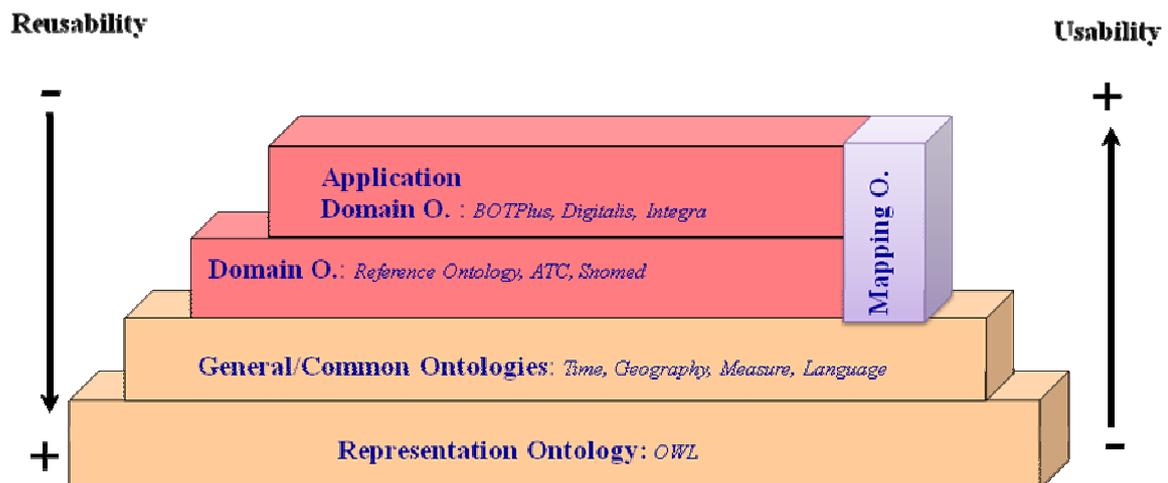


Figure 3: Modular approach of the Nomenclature Ontology Network

As was described in [1], 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 the domain level are located ontologies which define the main notion and concepts of the pharmaceutical domain that are substantial. In this level are included ontologies from the main standard terminologies or vocabularies in the eHealth domain, like Snomed, Mesh, NCI and ontologies which provide a classification of pharmaceutical terms like ATC classification (because of the fact that it is the WHO recommendation and is followed by the pharmaceutical experts in Spain and Europe) or the SPC ontology (Summary of Product Characteristics, a physician's information document used in the European Union).

At application domain are classified the ontologies which represents the knowledge of the real-world resources, after a re-engineering of them. In the Semantic Nomenclature case study, 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. Also, in this second iteration we include here the Semantic Nomenclature Reference Ontology, due its role in this iteration is the main ontology used in the Semantic Nomenclature prototype due this ontology acts as bridge between the domain ontologies and the other application ontologies thanks to the mappings between the ontologies.

As is described before, in this new approach, is added a new box, called Mapping Ontologies because in this iteration the mappings between the application ontologies (mainly the reference ontology) with the domain ontologies are formalized in third ontologies. This approach makes the management of the mappings in the Semantic Nomenclature easier and allows maintaining the mappings as other ontologies that are involved in the ontology network, between the involved source and destiny ontologies in the mapping.

In the following Figure 4 is depicted the new appearance of the Nomenclature Ontology Network in this second iteration in the lifecycle. In this case, more domain ontologies are added to the domain ontologies pool (connecting with international health terminologies); the Semantic Nomenclature 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 and with the most important classification system domain ontologies involved in the Semantic Nomenclature scenario as the ATC or Snomed. As well, in this figure are included these ontologies with the mappings between the main ontologies (small triangles), due they are part of the Semantic Nomenclature knowledge and makes possible the semantic interoperability between the sources. Moreover, as was described in [1] , in this new iterations are added more resources as SPC, RxNorm or DrugOnto that enrich the ontology network.

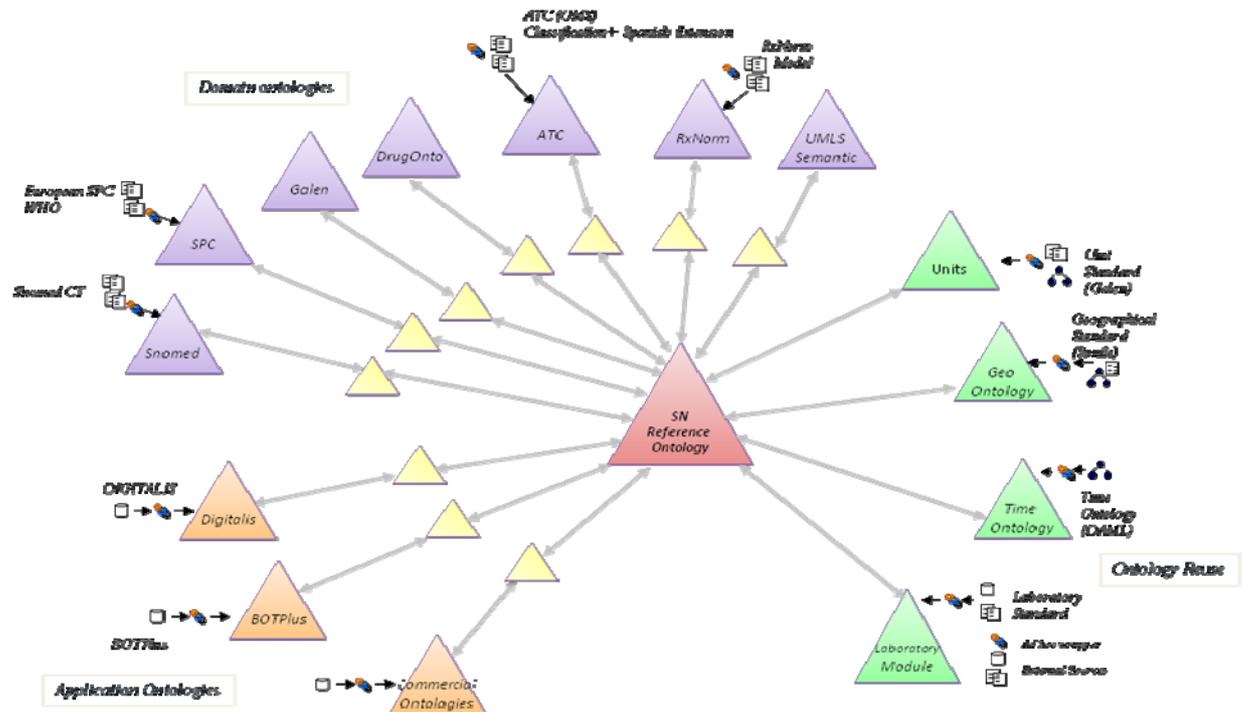


Figure 4: Nomenclature Ontology Network

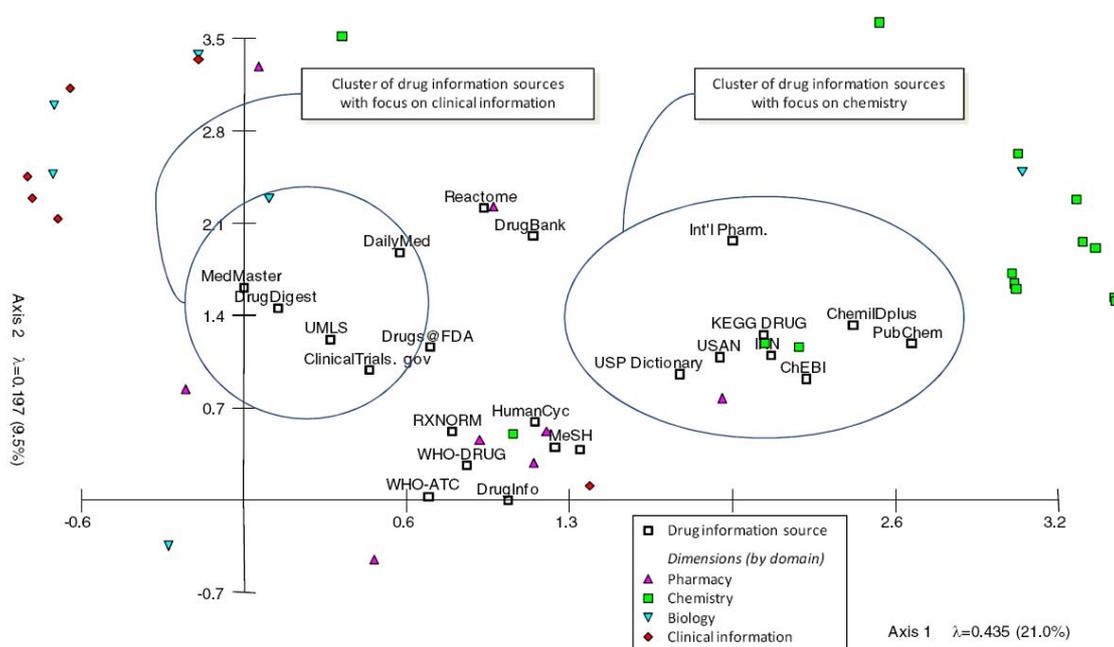
This new version of the Semantic Nomenclature Ontology Network should evolve in more iterations of the lifecycle model, where new resources or ontologies that could appear related with medical vocabularies used in the world are added and integrated. These ontologies may come from the current stakeholders (as ontologies of laboratory products, hospital ontologies) or external ones (ontologies from other countries or similar domains) or new standards or recommendations in the Health domain.

With this new version of the ontology network we maintain the expectations of the first version, so is aligned with the goals extracted from the case study scenarios described in [4] : integration of existing pharmaceutical resources and semi-automatic update of the BOTPlus information. Also, in this case, with the review of the main goal and knowledge described in the Reference Ontology, detailed in section 5, we achieve the new requirements proposed by the domain experts for cover the ambiguity between clinical and branded drugs.

This ontology network makes possible the easy interoperability and integration of the distributed resources for the description of pharmaceutical products. Moreover, the ontology network facilitates the aggregation of drug-related information in a semantic way because the reference ontology (application ontology for the prototype) is connected via mappings, stored as ontologies which are part of the knowledge, with different pharmaceutical ontologies at different levels. This solution makes possible the collection of information for concrete products and maintains the legacy databases updated, because the pharmaceutical product information gathered in the networked ontologies give an added value to the actors. Based on the ontology network, the actors can improve their commercial database and reduce their effort in complementing typical pharmaceutical compendium characteristics by giving flexible, extensible and reliable information about drugs to the users of the Pharmaceutical domain.

## 4. Review Inventory and Analysis of Knowledge Resources and Vocabulary

Multiple sources are available, each providing some elements of information about drugs (usually for a given purpose), but there exists no integrated view or directory that could be used to locate sources appropriate to a given purpose. This scenario is described in [9], where twenty-three sources that provide drug information in the pharmacy, chemistry, biology, and clinical medicine domains are analyzed. Their drug information content could be categorized in different dimensions. They propose this list of dimensions as a framework for characterizing drug information sources. As an evaluation, they show that this framework is useful for comparing drug information sources and selecting sources most relevant to a given use case. Figure 5 shows the resume of the analysis, and in this case, the information sources with focus in pharmacy domain or clinical information are the most interesting for the Semantic Nomenclature.



**Figure 5: Correspondence analysis between drug information sources and dimensions of drug information (in four domains pharmaceutical, chemical, biological and clinical) [9]**

A complete description and explanation of the figure, axis and clusters can be found in the original source [9]. Based on the Figure 5 and in the inventory review depicted in [1] and [7], in this section are reviewed the new resources in the health domain that could be interesting and useful for the case study.

### Glossary in the eHealth domain

The vocabulary used around health terminologies and ontologies is confusing, and different authors use the same words differently. However, as help, we provide the following definitions provided by the SemanticHEALTH project in [2] for the set of terms which could be useful for the readers.

*Controlled Vocabulary* – a list of specified items to be used for some purpose, usually in an information system to reduce ambiguity, misspellings, etc.

*System of identifiers (“codes”)* – Controlled vocabularies, and many lexicons, ontologies, and thesauri, are usually accompanied by systems of identifiers for their units, e.g. typically, identifiers act as the primary unambiguous means of referring to the entities in the system for computational purposes with the text form being used for communication with users. Examples are “Concept Unique Identifiers (CUIs)” from the UMLS, SNOMED Identifiers, etc. In many contexts, identifiers are known as “codes.”

*Lexicon* – A list of linguistic units that may be attached to a controlled vocabulary or ontology, in a specific language or sublanguage, often including linguistic information such as synonyms, preferred terms, parts of speech, inflections and other grammatical material. Example: Term terms and lexical material in UMLS identified by Lexical Unique Identifiers (LUIs)

*Ontology* – A symbolic logical model of some part of the meanings of the notions used in a field, i.e. those things that are universally true or true by definition. The key relationship in an ontology is “subsumption” or “kind-of”. Every instance of a sub-kind must be an instance of the kind, without exception. Typically, ontologies are implemented in logic languages such as OWL or frame systems such as Protégé-Frames. Examples: The GALEN Core Model, the stated form of SNOMED.

*Classification* – an organization of entities into classes for a specific purpose such as international reporting or remuneration. Examples ICD and Diagnosis Related Groups.

*Thesaurus* – a system of terms organized for navigation with the primary relationship being “broader than”/“narrower than”. The “broader than”/“Narrower than” relation is explicitly not limited to subsumption/kind of relation. It is a general form of linguistic hyper/hyponymy aimed at assisting human navigation. However, it is explicitly not intended that it be used as the basis for logical inferences, e.g. in decision support. Examples MeSH, NCI.

*Knowledge Representation System / Background knowledge base* – the common knowledge to be assumed by the system, including both the ontology – what is universally true – and generalizations about what is typically true.

*Terminology* – Any or all of the above in various combinations. Most health terminologies consist, at a minimum, of a controlled vocabulary and a system of identifiers. They may include extended lexicons, ontologies, thesauri or background knowledge base. This definition is deliberately broader and less specific than that in most of the standard references and intended to approximate common usage.

*Coding system* – A terminology with attached identifiers or “codes”.

*Information model or Data model* a model of how information is structured in a given software system, message, or electronic health record. In general, the data structures carry codes for the ontology as their content.

## 4.1 Drug Information: Commercial Drug Databases and National Standards

Nowadays there are more than 100.000 pharmaceutical products that are marketed in Spain. Storing pharmaceutical product information is a challenge due to the huge number of available products. Laboratories may sell similar active ingredients using different trade names, due the variability of dosage, pharmaceutical form, routes of administration or therapeutic use.

There are different approaches in the Health domain for classify the pharmaceutical products used by the different terminologies.

A first strategy is to consider only the active ingredients, and therapeutic and pharmacological features of the pharmaceutical products. This is the case of the Anatomical Therapeutic Chemical Classification System<sup>1</sup> used for the classification of drugs. It is controlled by the WHO Collaborating Centre for Drug Statistics Methodology, and was first published in 1976. 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. This classification was reviewed at [1] .

Also, the WHO maintains a list of International Nonproprietary Names (INN)<sup>2</sup> that facilitates the identification of pharmaceutical substances or active pharmaceutical ingredients. Each INN is a unique name that is globally recognized and is public property. Some countries recommend drug prescription using the INN instead of trade names.

Another approach consists in providing a register for every marketed product and combinations. This is the strategy followed by the WHO Drug Dictionary<sup>3</sup>, which is the world's most comprehensive dictionary of medicinal product information. It is used by pharmaceutical companies, clinical research organizations and drug regulatory authorities for identifying drug names, their active ingredients and therapeutic use, in the course of their drug safety surveillance. Drugs are classified according to ATC codes. Drugs containing the same active ingredient are referred by the INN that acts as a preferred name. The majority of entries refer to prescription-only products, but some are over-the-counter (OTC) or pharmacist-dispensed. Biotech and blood products, diagnostic substances and contrast media are also entered in the dictionary.

In the USA, the National Library of Medicine (NLM) / Unified Medical Language System (UMLS) provides a standard drug vocabulary, RxNorm<sup>4</sup>. RxNorm provides normalized names for clinical drugs and links its names to many of the drug vocabularies commonly used in pharmacy management and drug interaction software

In the UK the DM+D<sup>5</sup> – Drugs, Medications, and Devices – database provides a dictionary containing unique identifiers and associated textual descriptions for medicines and medical devices

---

<sup>1</sup> <http://www.whocc.no/atcddd/>

<sup>2</sup> <http://www.who.int/medicines/services/inn/innquidance/en/index.html>

<sup>3</sup> <http://www.who-umc.org/>

<sup>4</sup> <http://www.nlm.nih.gov/research/umls/rxnorm/index.html>

<sup>5</sup> <http://www.dmd.nhs.uk/>

The third approach is to benefit from vocabularies included in drug knowledge bases. Although not strictly terminological, the provision of drug information, codes, and databases is a major business. This kind of approach is based on providing information models for describing drugs, medical products, devices, etc. However, these systems have major vocabulary components, and their identifiers are widely used in commercial systems that subscribe to them.

Furthermore, the update and liability requirements attached to drug information are stricter. Monthly update cycles are standard. This is the case of the pharmaceutical product classification provided by the Spanish government, which is not based in “clinical” or “pharmacological” classification. Other involved actors in the domain, as the GSCoP or pharmaceutical department in hospitals, should update their information systems, with the comprehensive product lists provided by the government.

## 4.2 Ontological resources

In [1] a detailed inventory of resources related with the Semantic Nomenclature case study was presented. This inventory tries to analyse and consider all the possible resources which are the grounding of the knowledge of the case study. The inventory included ontological and non-ontological resources.

In the following Table 1: Ontological resources, we only review the ontology resources analyzed in the previous deliverable and only specify the new terminologies or ontological resources that are added to the Nomenclature which are discovered in the past months as the case of RxNorm or DrugOnto.

### RxNorm

RxNorm is a standardized nomenclature for clinical drugs, produced by the National Library of Medicine (NLM). In this context, a *clinical drug* is a pharmaceutical product given to (or taken by) a patient with a therapeutic or diagnostic intent. In RxNorm, the name of a clinical drug combines its ingredients, strengths, and form. RxNorm’s standard names for clinical drugs are connected to the varying names of drugs present in many different controlled vocabularies within the Unified Medical Language System (UMLS) Metathesaurus, including those in commercially available drug information sources. These connections are intended to facilitate interoperability among the computerized systems that record or process data dealing with clinical drugs. RxNorm contains the names of prescription and many non-prescription formulations that exist in the United States.[10]

RxNav<sup>6</sup> is a browser for RxNorm, the NLM repository of standard names for clinical drugs. RxNav displays links from clinical drugs, both branded and generic, to their active ingredients, drug components and related brand names.

### DrugOnt Schema

The LSDIS lab’s<sup>7</sup> collaborative research project on Active Semantic Electronic Patient Record with the Athens Heart Center (AHC)<sup>8</sup> consists in an implementation of Active

---

<sup>6</sup> <http://mor.nlm.nih.gov/download/rxnav/>

<sup>7</sup> <http://lsdis.cs.uqa.edu>

Semantic Documents in a healthcare (more specifically cardiology practice) environment. This implementation has so far involved the development of populated ontologies in the healthcare, and the DrugOnt schema, defined in Figure 6, includes concepts such as indications, interactions, formulary, etc. License content equivalent to physician's drug reference was the primary source for populating this ontology.

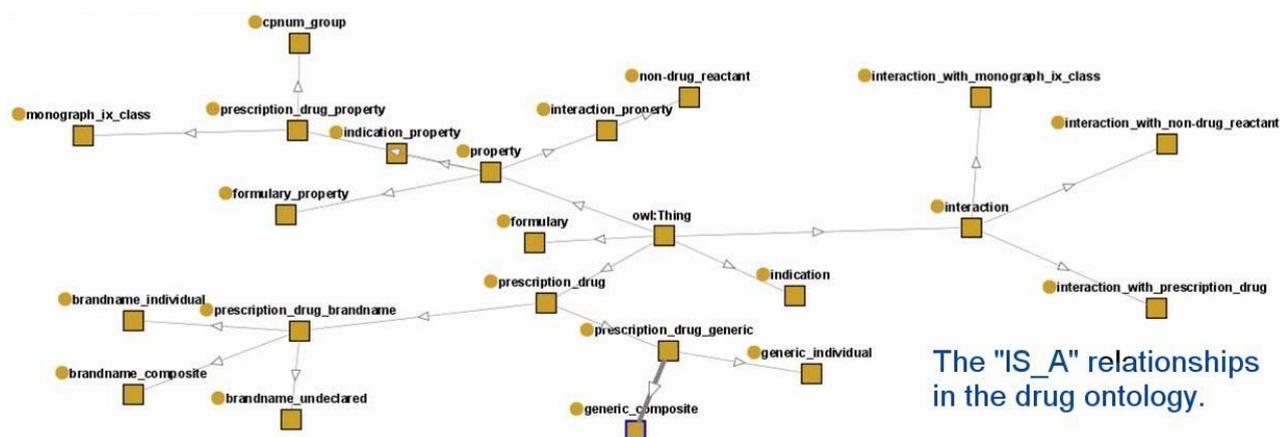


Figure 6: DrugOnt Schema

## Disease Ontology

Disease Ontology is a controlled medical vocabulary developed at the Bioinformatics Core Facility in collaboration with the NuGene Project<sup>9</sup> at the Center for Genetic Medicine<sup>10</sup>. It was designed to facilitate the mapping of diseases and associated conditions to particular medical codes such as ICD9CM, SNOMED and others. Disease Ontology is implemented as a directed acyclic graph (DAG) and utilizes the Unified Medical Language System (UMLS) as its immediate source vocabulary to access medical Ontologies such as ICD9CM. As a graph, the Disease Ontology can be thought of as a subset of UMLS. It fills a niche in the medical ontology world as a lightweight ontology offering context-free concept identifiers designed specifically to facilitate mapping to medical billing codes. Other Ontologies such as SNOMED and MESH lack these features [11].

This ontology is not key in the case study because is not about pharmaceutical product description, but we think that in the future versions of the ontology network, it can include and describe relations between pharmaceutical products, active ingredients or therapeutic use with the diseases described in the ontology.

Table 1 shows the candidate ontology resources that are relevant for the Semantic Nomenclature scenario, which enrich the ontology network and help in the different scenarios depicted in the case study: semantic interoperability, connection with international terminologies...

<sup>8</sup> <http://www.athensheartcenter.com/>

<sup>9</sup> <http://www.nugene.org/>

<sup>10</sup> <http://www.cgm.northwestern.edu/>

Name	Type	Distribution	License	URL	Relevance to WP8
<b>NCI</b>	Thesaurus	OWL Full	Free	<a href="http://www.mindswap.org/2003/CancerOntology/nciOncology.owl">http://www.mindswap.org/2003/CancerOntology/nciOncology.owl</a>	Medical terms, Large ontology
<b>GALEN</b>	Ontology	OWL, XML	Free	<a href="http://www.galen.org/">http://www.galen.org/</a>	Medical terms. Models the active ingredients. Not very important for the case study
<b>UMLSSN</b>	Ontology	OWL	Free	<a href="http://swpatho.ag-nbi.de/owldata/swpatho1/umlssn.owl">http://swpatho.ag-nbi.de/owldata/swpatho1/umlssn.owl</a>	UMLS Semantic Network.
<b>COSMO</b>	Ontology	OWL	Free	<a href="http://colab.cim3.net/file/work/SICoP/ontac/reference/COSMO-ontology/COSMOtopOntology475.owl">http://colab.cim3.net/file/work/SICoP/ontac/reference/COSMO-ontology/COSMOtopOntology475.owl</a>	Top-level ontology
<b>OpenCYC</b>	Ontology	OWL	Free	<a href="http://www.opencyc.org">http://www.opencyc.org</a>	Upper ontology whose domain is all of human consensus reality. Models drugs and substances and contains instances of drugs.
<b>RxNORM</b>	Ontology	OWL	Free	<a href="http://secse.es.atosorigin.com/...">http://secse.es.atosorigin.com/...</a>	Connect with the USA clinical products.
<b>DrugOnt</b>	Ontology	OWL	Free	<a href="http://lstdis.cs.uga.edu/projects/asdoc/DrugOnt_schema.owl">http://lstdis.cs.uga.edu/projects/asdoc/DrugOnt_schema.owl</a>	Improve and enrich the Semantic description of Drugs
<b>MeSH</b>	Ontology	OWL	Free	<a href="http://bike.snu.ac.kr/wp-content/uploads/2008/09/meshonto4.owl">http://bike.snu.ac.kr/wp-content/uploads/2008/09/meshonto4.owl</a>	International Health Terminology. This ontology represents the MeSH taxonomy
<b>OWN</b>	Ontology	OWL	Free	<a href="http://www.loa-cnr.it/ontologies/OWN/OWN.owl">http://www.loa-cnr.it/ontologies/OWN/OWN.owl</a>	Wordnet translation. Not very important
<b>Disease Ontology</b>	Ontology	OWL	Free	<a href="http://purl.org/obo/owl/DOID">http://purl.org/obo/owl/DOID</a>	This ontology is not about drugs, but we can describe relations between drug/disease
<b>ATC classification</b>	Ontology	OWL	Free	<a href="http://secse.es.atosorigin.com/...">http://secse.es.atosorigin.com/...</a>	Anatomical Therapeutic Chemical (ATC) classification system, the drugs are divided into different groups according to the organ or system on which they act and their chemical, pharmacological and therapeutic properties

Table 1: Ontological resources

### 4.3 Non Ontological resources

As in the previous section 4.2, in this section is resumed the non-ontology resources analyzed in the previous deliverable and only is specified the new non-ontological resources (databases, document, web...) that are added in the case study. This new resources were discovered in the past months as the case of SPC template recommended by WHO and the European Commission.

#### SPC Template

The European Union commission provides some rules governing medicinal products. Article 8(3)(j) of Directive 2001/83/EC and Article 6(1) of Regulation (EC) 726/2004 require that in order to obtain a marketing authorization, a Summary of Product Characteristics (SPC).

The SPC is the basis of information for health professionals on how to use the medicinal product safely and effectively. The Package Leaflet (PL) shall be drawn up in accordance with the SPC. The Guideline on excipients in the label and Package leaflet of medicinal products for human use is also applicable to the SPC.

It is not in the remit of the SPC to give general advice on the treatment of particular medical conditions. On the other hand specific aspects of the treatment related to use of the medicinal product or its effects should be mentioned.

The fundamental purpose of the summary of product characteristics is to provide a clear and unambiguous description of the approved conditions of use of a medicinal product in the European Community or Member State(s) concerned, presented in accordance with a single standardized layout.

This guideline provides advice on the principles of presenting information in the SPC. Applicants should maintain the integrity of each section of the document by only including information in each section, which is relevant to the section heading. However, some issues may need to be addressed in more than one section of the SPC and in such situations the individual statements may cross-refer to other sections when these contain relevant additional information.

When a guideline exists for the SPC of a specific therapeutic area (e.g. antibiotics), pharmacological group (e.g. benzodiazepines), or product type (e.g. vaccines), this guideline should be taken into account.

Separate SPCs are required for each pharmaceutical form and strength by the European Commission and certain Member States

The European commission gives the template for the SPC that all laboratories should complete for describe the pharmaceutical products. This document could be input as an ontology learning process in order to obtain a ontology which describes the knowledge expressed by the document. This ontology could enrich the Semantic Nomenclature ontology network at the domain level. The SPC template could be found here<sup>11</sup>

Finally, Table 2 shows the review of the non-ontological resources that are relevant for the Semantic Nomenclature scenario. In this case, the main characteristics of the following resources is that they provide the information data of the pharmaceutical

<sup>11</sup> <http://www.emea.europa.eu/pdfs/human/regaffair/29952707en.pdf>

products, so there are mainly re-engineered to ontologies which are located at the application level of the ontology network.

Name	Type /Distribution	License	URL	Relevance to WP8
<b>BOTPlus</b>	Database/ MS-Access DB	Needs to be subscriber of the GSCoP		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.
<b>Digitalis</b>	Database/ MS-Access DB	Free	<a href="http://www.msc.es/profesionales/farmacia/pdf/Nomenclator_Digitalis_2008_12.zip">http://www.msc.es/profesionales/farmacia/pdf/Nomenclator_Digitalis_2008_12.zip</a>	Digitalis is one of the official 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.
<b>Integra</b>	Database/ MS-Access DB	Free	<a href="http://www.msc.es/profesionales/farmacia/pdf/Integra_2007_12.zip">http://www.msc.es/profesionales/farmacia/pdf/Integra_2007_12.zip</a>	Integra is one of the official Nomenclatures provided by the Spanish government. It contains information about drugs and health material used in Spanish hospitals.
<b>SPC</b>	Document/ PDF-DOC	Free	<a href="http://www.ema.europa.eu/pdfs/human/regaffair/29952707en.pdf">http://www.ema.europa.eu/pdfs/human/regaffair/29952707en.pdf</a>	The SmPC is the basis of information for healthcare professionals on how to use the medicinal product safely and effectively. This guideline will be included in The Rules Governing Medicinal Products in the European Union Volume 2C Notice to Applicants
<b>Vademecum.es</b>	Database accessed via web app / Web app	Free	<a href="http://www.vademecum.es">www.vademecum.es</a>	Upper ontology whose domain is all of human consensus reality. Models drugs and substances and contains instances of drugs.
<b>CedimCat</b>	Database accessed via web app / Web app	Free	<a href="http://www.cedimcat.info/html/ca/dir2451/doc10713.html">http://www.cedimcat.info/html/ca/dir2451/doc10713.html</a>	Cedimcat is a web that provides a lot of information about drugs in Catalonia, the information is provided in Catalan (multilingualism)

**Table 2: Non-ontological resources**

## 5. Semantic Nomenclature Network of Ontologies

### 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).

### 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.

The ontologies are accessible through this URL:

<http://212.170.156.131:10000/ontologies/SNomenclatureontologies.xhtml>

### 5.1 Application Ontologies

#### Semantic Nomenclature Reference Ontology

##### *Motivation*

The Semantic Nomenclature Reference Ontology is the core of the ontology network used in the case study. This ontology has three main goals: act as a bridge between the different application ontologies and domain ontologies; the second goal is implement one the new requirements as the disambiguate between the clinical drug/branded drug; the third functionality is act as the application ontology for the Semantic Nomenclature prototype, because from this ontology, using the mappings and relations can be accessed the rest of the ontology network. This third goal is the main reason for changing the reference ontology from the domain level to the application level.

The Reference Ontology is based on the main recommendations provided by the pharmaceutical product, and also using the semantic model of Snomed as background knowledge, mainly from the Pharmaceutical/Biological product term used in the terminology. Based on the descriptions in the Snomed user manual<sup>12</sup>, we define a hierarchy for distinguishing the pharmaceutical products.

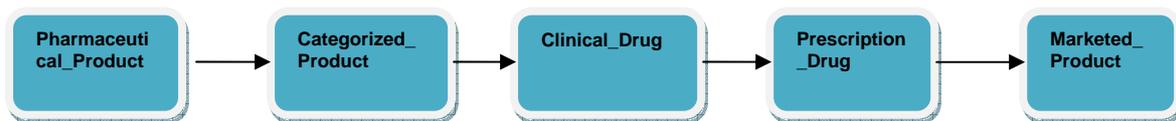
In Snomed [12], the Pharmaceutical/biologic product hierarchy is separated from the Substance hierarchy, in order to clearly distinguish drug products (products) from their chemical constituents (substances). The pharmaceutical product hierarchy contains concepts that represent the multiple levels of granularity required to support a variety of uses cases in the eHealth domain. The levels of drug products represented in the

---

<sup>12</sup>[http://www.ihtsdo.org/fileadmin/user\\_upload/Docs\\_01/SNOMED\\_CT\\_Publications/SNOMED\\_CT\\_User\\_Guide\\_20080731.pdf](http://www.ihtsdo.org/fileadmin/user_upload/Docs_01/SNOMED_CT_Publications/SNOMED_CT_User_Guide_20080731.pdf)

Snomed International Release include at the top level, Virtual Medicinal Product (VMP, product name, strength, and dose form are all represented in the Fully Specified Name i.e. Diazepam 5mg tablet), Virtual Therapeutic Moiety (VTM, include the product name but not formulation, dose or strength in the Fully Specified Name and is could be related to an active ingredient, i.e. Diazepam) and Product Category (*products* related by their functionality mechanism of action or therapeutic use, i.e. *Mineralocorticoid preparation product*). Additionally, US and UK drug extensions have been developed, which represent Actual Medicinal Products (AMPs, i.e. Zoloft 50mg tablet).

Based on this hierarchy provided by Snomed, we define in the Reference Nomenclature a hierarchy that cover the needs of the case study requirements and the distinguish between clinical drugs and branded drugs, as in Snomed. The categorized product concept serves to classify the products according to their therapeutic use. In the next level, the clinical drug concept, is the equivalent for VTM, and could be related with the active ingredient concept. The prescription drug concept is similar to VMP, which could be described by the pharmaceutical form and dosage of the pharmaceutical product, and is useful for the prescription use case in Hospitals. Finally the Branded Drug concept is the concept used for describe the branded drug product that are dispensed in the pharmacies. This last concept is related with the Laboratory, is defined by its national code, price... In the Figure 7 is depicted the hierarchy relation for pharmaceutical products in the reference ontology



**Figure 7: Pharmaceutical product concept in the Reference ontology**

This hierarchy allows us to establish mappings with the domain ontologies and international terminologies (using the categorized product, clinical drug concept or prescription drug concept) and with the application ontologies (branded drug concept). So, this ontology enables the interoperability between the different information sources and terminologies in the case study domain and covers the disambiguation between clinical drugs and branded drugs.

Also, the Reference ontology describes some main concepts as Health Entities or Clinical Findings, Substances or Procedures needed for describe the knowledge of the pharmaceutical domain in the case study scenario.

#### *Brief outline*

This ontology has seven root concepts. Each concept represents a generic part of the pharmaceutical sector. In this way light hierarchies have been defined in order to represent medical product consumers (such as hospitals and pharmacies) or other medical entities (such as Laboratories and Government\_Entity) under the HealthCare Entity concept.

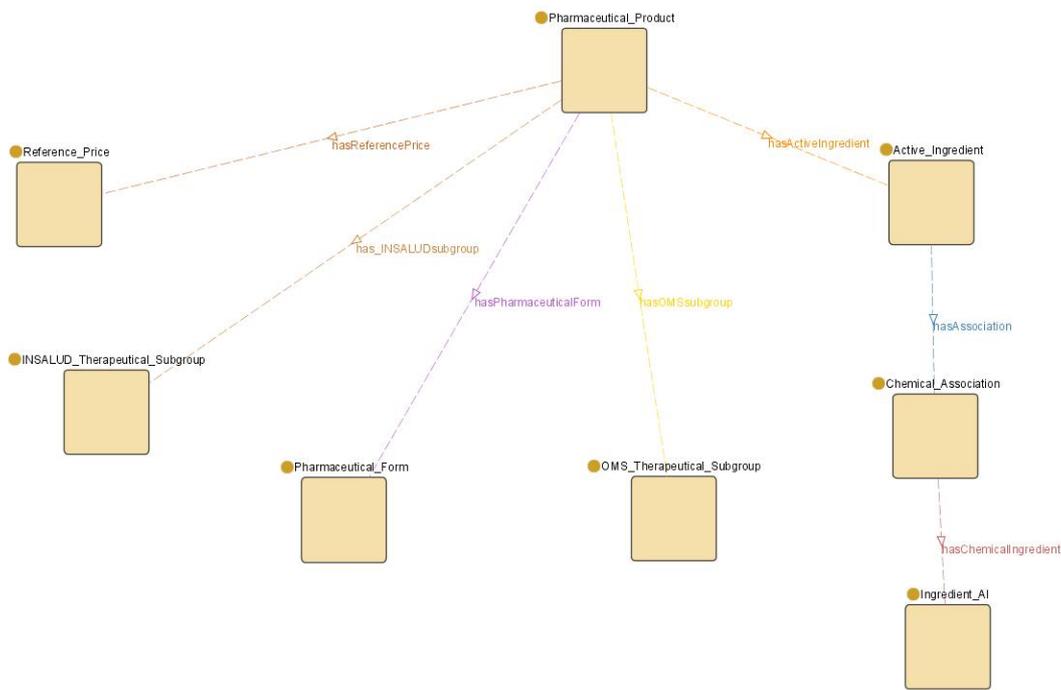
Furthermore, the ontology represents the knowledge necessary for represent and enrich the representation of the pharmaceutical product description, which is the core of the ontology. For this purpose, is formalized a specification of clinical findings, pharmacological events, substances necessities to describe the clinical drugs and procedures in the health domain.

The main hierarchy is Pharmaceutical\_Product, and the underlying Clinical\_Drug, Prescription\_Drug concepts are mapped to the equivalent concepts of the domain



schema model are used to describe with more detail the information around the marketed product and its use.

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).



**Figure 9: Digitalis ontology relationships**

### Ontology Details

URI: <http://212.170.156.131:10000/ontologies/DigitalisOntology.owl>

Concepts: 14, *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*

Object Properties: 7

Datatype Properties: 45

Instances: External Ontology populated using R2O & ODEMapster

### BOTPlus Ontology

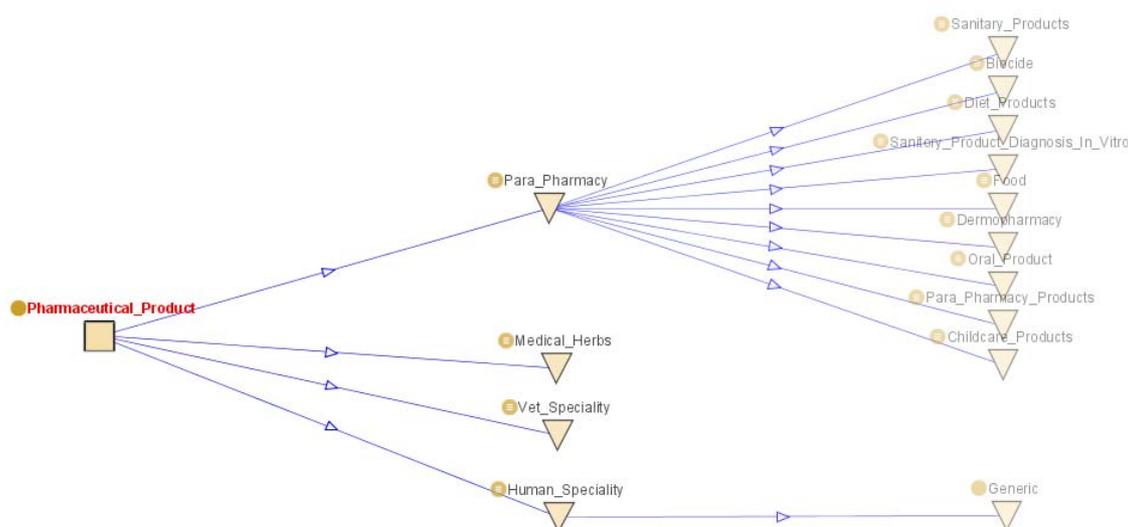
#### Motivation

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. As in the Digitalis ontology, the BOTPlus ontology captures more data than the marketed product information, as information about interactions, pathology, active ingredients. These concepts are related each other conceptualizing the relations represented in the BOTPlus schema model. The BOTPlus schema model is bigger than the Digitalis but we only re-engineered the modules and parts that are

interesting for the case study. In future versions, new iterations of the ontology will include the parts of the schema rule out in this version.

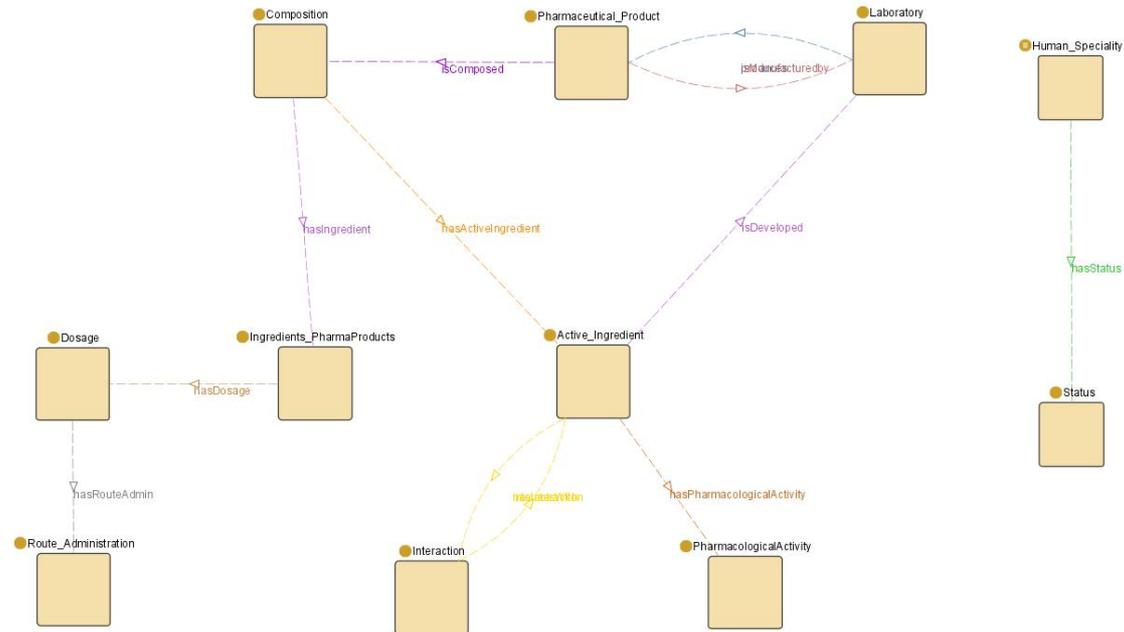
### Brief outline

As was explained in [1], the BOTPlus model provides a classification for marketed products based on a classification code and the purpose of the product, distinguishing between productHuman, Vet\_Speciality, Medical\_Herbs, Dermopharmacy and Parapharmacy. In this last speciality, are also described different types of products that are sold in the pharmacy like Diet\_Products, Food, Childcare\_products... The following Figure 10 shows the hierarchy provided by the BOTPlus model for the marketed pharmaceutical products.



**Figure 10: Pharmaceutical Product BOTPlus Hierarchy**

Moreover, the pharmaceutical products have associated their ATC code if it is disposable or a therapeutic code provided by the Ministry of Health (based on the ATC guidelines) if the pharmaceutical product is a non-chemical pharmaceutical product. As in case of the Digitalis Ontology, 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 and obtaining ontologies with the instances populated with ODEMapster.



**Figure 11: BOTPlus ontology relationship**

As was explained in [1], the 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.

### *Ontology Details*

URI: <http://212.170.156.131:10000/ontologies/BOTPlusOnto.owl>

Concepts: 40

Object Properties: 12

Datatype Properties: 76

Instances: External Ontology populated using R2O mappings & ODEMapster

## **5.2 Domain Ontologies**

### **ATC Ontology**

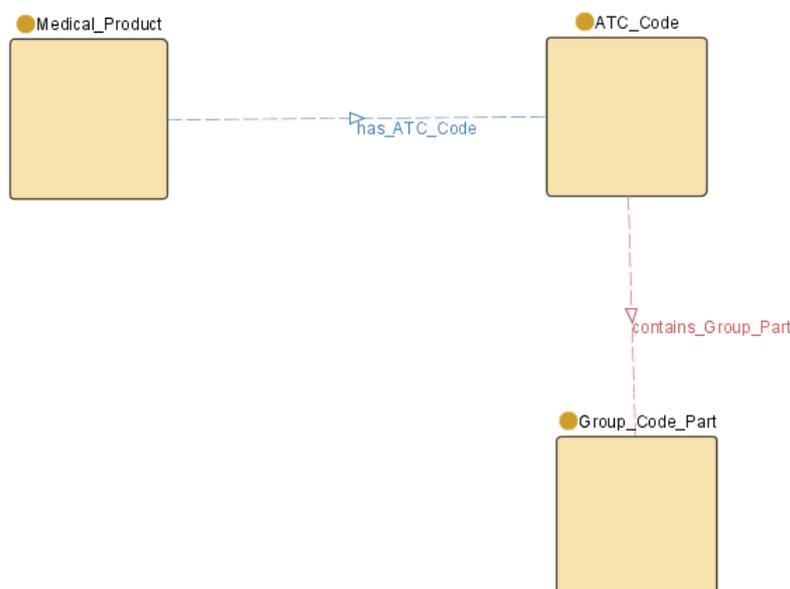
#### *Motivation*

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 modeled the two initial levels of the hierarchy (anatomical, therapeutical) due the extension of the classification. Even a complete example of all levels is modeled in order to demonstrate the mechanism of the classification.

### *Brief outline*

As was explained in [1], the classification of the pharmaceutical products via reasoning 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 12: ATC relationship**

### *Ontology Details*

URI: <http://212.170.156.131:10000/ontologies/ATCOntologyv2.owl>

Concepts: 122

Object Properties: 2

Datatype Properties: 3

Instances: 11399 Using R20 mappings & ODEMapster NeOn plugin

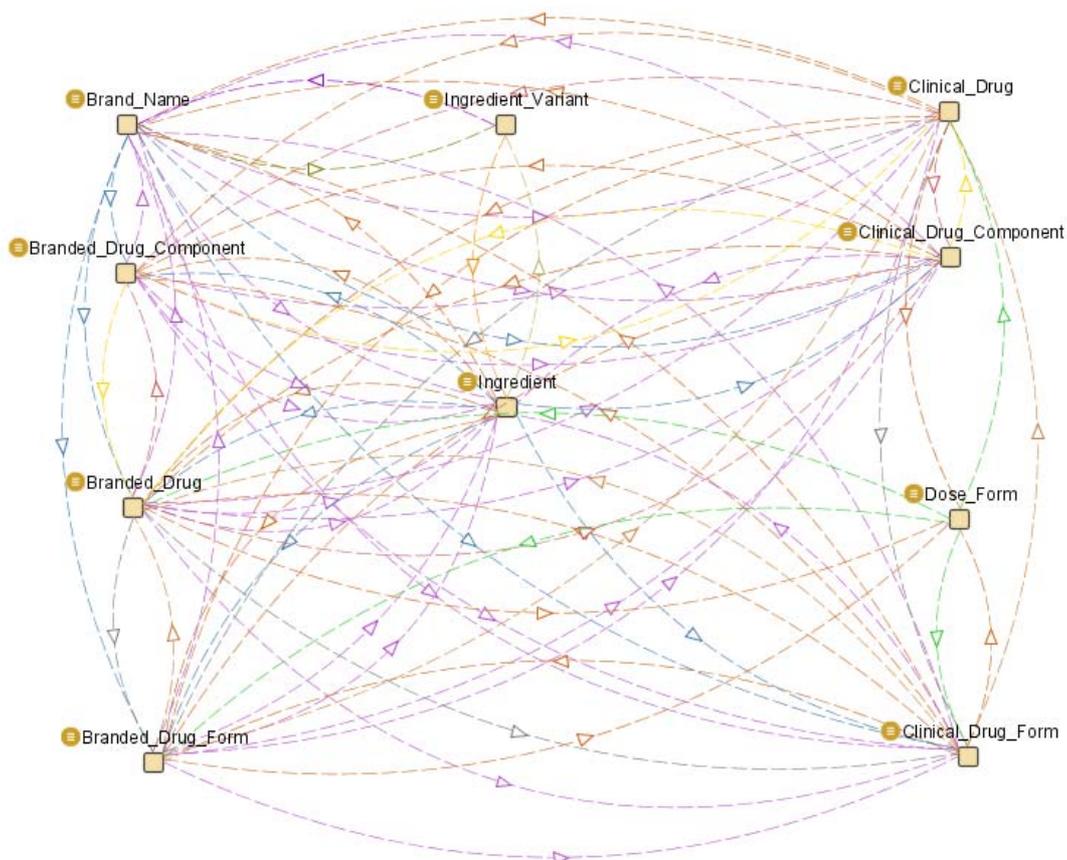
## **RxNorm Ontology**

### *Motivation*

In section 4 we describe RxNorm as the NLM terminology of standard names and codes for clinical drugs. This terminology links clinical drugs, both branded and generic, to their active ingredients, drug components and related brand names. And also, this terminology connects to a dataset comprised by ingredients, brand names, clinical drug components, branded drug components, clinical drugs, branded drugs, clinical drug forms, branded drug forms and dose forms. RxNorm is one of a suite of designated standards for use in U.S. Federal Government systems for the electronic exchange of clinical health information.

### *Brief outline*

The RxNorm terminology, besides the concepts below, describes pharmacologic action, drug-drug interactions, indications and contraindications, adverse reactions, etc. Based on this description and the RxNav model schema<sup>13</sup>, we re-engineered it to the Semantic Nomenclature case study, in order to connect with the reference ontology at a domain level (clinical drugs) and application level (brand products). The development of the ontology was performed using the NeOn Toolkit and include the restrictions detailed in the RxNorm terminology. Thanks to these mappings and the API provided by RxNorm, in the Semantic Nomenclature prototype we can access to the information provided by the U.S. about pharmaceutical products used in Spain.



**Figure 13: RxNorm ontology relationships**

<sup>13</sup> <http://mor.nlm.nih.gov/download/rxnav/>

### *Ontology Details*

URI: <http://212.170.156.131:10000/ontologies/rxnorm.owl>

Concepts: 11

Object Properties: 14

Datatype Properties: 0

### **UMLS Ontology**

#### *Motivation*

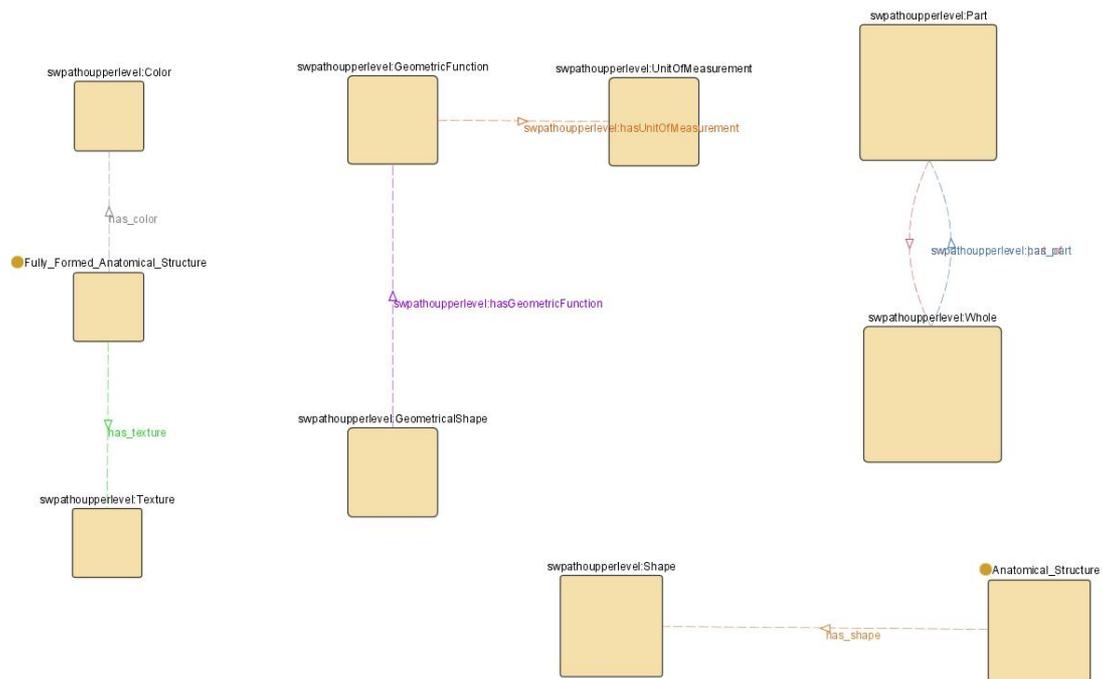
UMLS<sup>14</sup> provides access to multiple knowledge sources in the medical domain (SNOMED included). The purpose of the UMLS Semantic Network is to provide a consistent categorization of all concepts represented in the UMLS Metathesaurus and to provide a set of useful relationships between these concepts. All information about specific concepts is found in the Metathesaurus. The Network provides information about the set of basic semantic types, or categories, which may be assigned to these concepts, and it defines the set of relationships that may hold between the semantic types. A complete description can be found here<sup>14</sup>

#### *Brief outline*

The semantic types are the nodes in the Network, and the relationships between them are the links. There are major groupings of semantic types for organisms, anatomical structures, biologic function, chemicals, events, physical objects, and concepts. The primary link in the Network is the "isa" link. This establishes the hierarchy of types within the Network and is used for deciding on the most specific semantic type available for assignment to a Metathesaurus concept. The relations are stated between high level semantic types in the Network whenever possible and are generally inherited via the "isa" link by all the children of those types. The Semantic Network contains 135 semantic types and 54 relationships.

---

<sup>14</sup> <http://www.nlm.nih.gov/research/umls/meta3.html>



**Figure 14: UMLS Semantic Types relationships**

### Ontology Details

URI: <http://swpatho.ag-nbi.de/owldata/umlssn.owl>

Concepts: 135

Object Properties: 54

Datatype Properties: 1

Instances: 13

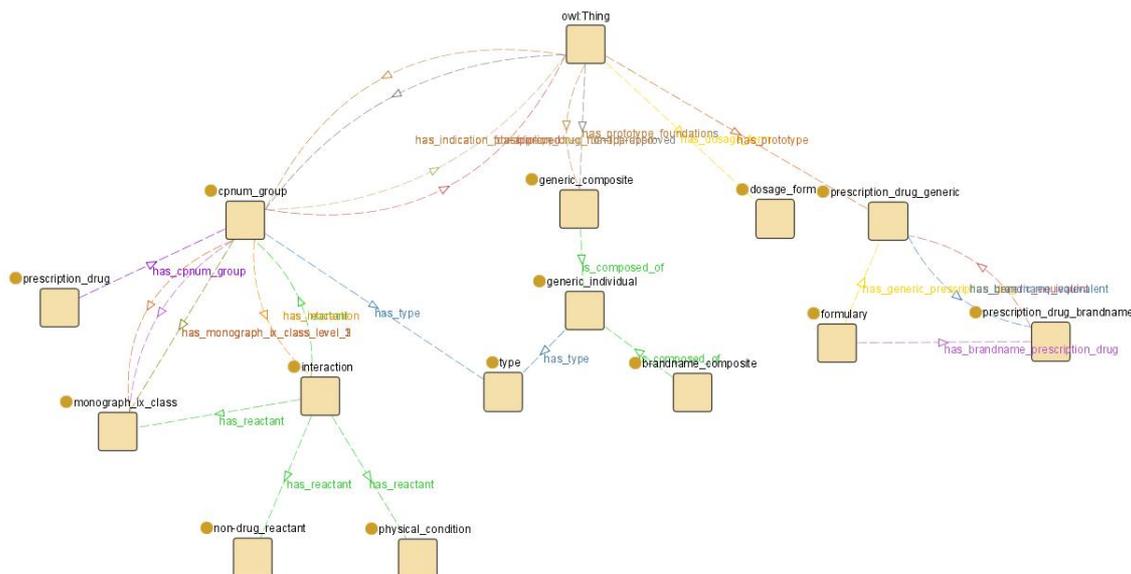
### Drug-Ont Schema Ontology

#### Motivation

The LSDIS lab's collaborative research project on Active Semantic Electronic Patient Record with the Athens Heart Center (AHC) consists in an implementation of Active Semantic Documents in a healthcare (more specifically cardiology practice) environment. This implementation has so far involved the development of populated ontologies in the healthcare, and the DrugOnt schema includes concepts such as indications, interactions, formulary, etc. License content equivalent to physician's drug reference was the primary source for populating this ontology.

#### Brief outline

The ontology describes Prescription\_Drug and its properties and also distinguish between Prescription\_Drug\_Generic and Prescription\_Drug\_Brandname, which is in line with the purpose of the Reference Ontology and contributes with a rich description of this concepts related with interactions, formulary and indications of the drugs.



**Figure 15: Drug-Onto Schema relationship**

### Ontology Details

URI: [http://212.170.156.131:10000/ontologies/DrugOnt\\_schema.owl](http://212.170.156.131:10000/ontologies/DrugOnt_schema.owl)

Concepts: 28

Object Properties: 26

Datatype Properties: 32

Instances: 43

### Mesh Ontology

#### Motivation

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 [13].

#### Brief outline

As was explained in the previous WP8 deliverables, MeSH is not a relevant resource from the case study purpose at the first iterations of the ontology network lifecycle, but in this iteration we will include the hierarchy of MeSH found in this ontology<sup>15</sup>, in order to enrich the Nomenclature ontology network.

<sup>15</sup> <http://bike.snu.ac.kr/wp-content/uploads/2008/09/meshonto4.owl>

### *Ontology Details*

URI: <http://bike.snu.ac.kr/wp-content/uploads/2008/09/meshonto4.owl>

Concepts: 24710 concepts in the MeSH hierarchy

## **Galen Ontology**

### *Motivation*

The GALEN ontology is a result from the OpenGALEN Foundation<sup>16</sup> (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 unit concepts, very useful describing some characteristics of the pharmaceutical products.

### *Brief outline*

The ontology describes in detail a lot of terms used in the health domain for describe procedures, diseases, process, organism... and it is useful and interesting to include the ontology in the ontology network in order to enrich the concepts used for described clinical and branded drug in the application ontologies using the descriptions provided by one of the “standard” ontologies used in the eHealth domain in the world.

### *Ontology Details*

URI: <http://www.co-ode.org/galen/full-galen.owl>

Concepts: 23141

Object Properties: 950

## **NCI Ontology**

### *Motivation*

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

### *Brief outline*

The NCI Thesaurus is built using the Ontylog dialect of description logic (DL). The NCI Thesaurus is designed, first and foremost, to be a *thesaurus* – “a controlled vocabulary arranged in a known order and structured so that the various relationships among terms are displayed clearly and identified by standardized relationship indicators.” NCI Thesaurus has some ontology-like features but NCI Thesaurus is not an ontology and is not designed or intended to be one. Its primary role is that of a bridge for human to human communication across specialties and data resources [14]

---

<sup>16</sup> <http://www.opengalen.org/>

The purpose of the NCI thesaurus in the Semantic Nomenclature ontology network is enrich the description of the concepts related with the pharmaceutical products, procedures, diseases... Also, the goal is connect the ontology network with the most relevant ontologies in the eHealth domain.

#### *Ontology Details*

URI: <http://www.mindswap.org/2003/CancerOntology/nciOncology.owl>

Concepts: over 34.000 concepts classified in 20 taxonomic trees

### **SPC Ontology**

#### *Motivation*

As is detailed in section 4, the Summary of Product Characteristics (SPC) is the basis of information for health professionals on how to use the medicinal product safely and effectively. The Package Leaflet (PL) shall be drawn up in accordance with the SPC. The Guideline on excipients in the label and Package leaflet of medicinal products for human use is also applicable to the SPC.

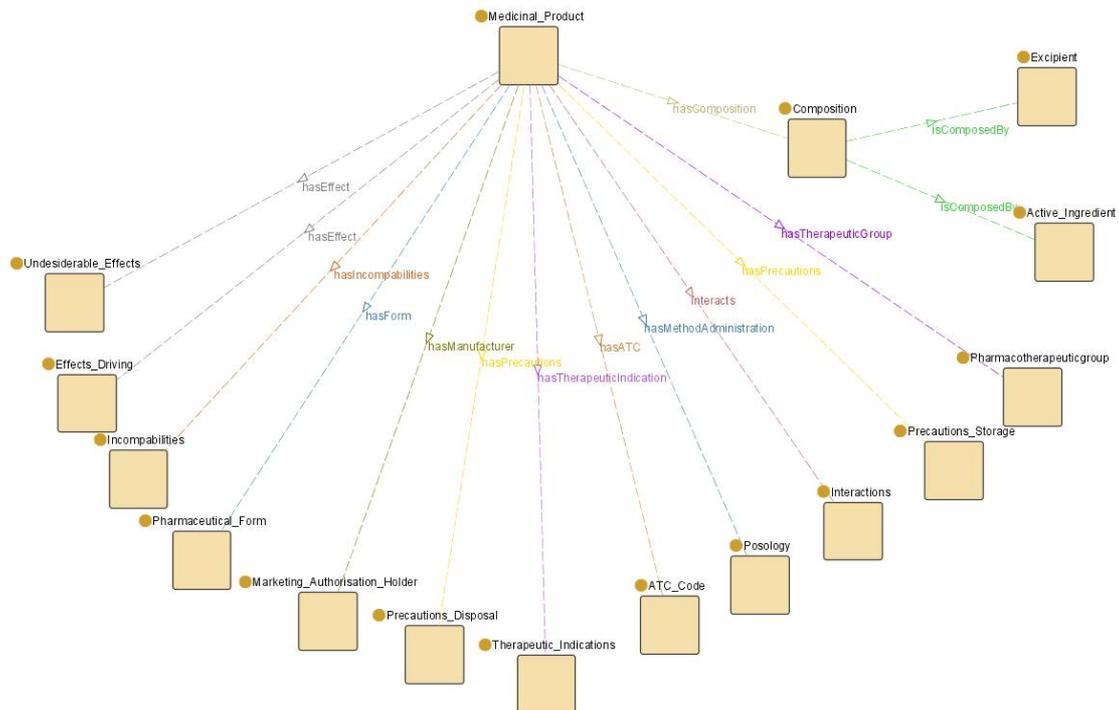
The fundamental purpose of the summary of product characteristics is to provide a clear and unambiguous description of the approved conditions of use of a medicinal product in the European Community or Member State(s) concerned, presented in accordance with a single standardized layout.

The European commission gives the template for the SPC that all laboratories should complete for describe the pharmaceutical products. This resource serves us to provide a new description for the marketed products based on the European Commission guidelines and restrictions for describe pharmaceutical products.

#### *Brief outline*

As a result of a ontology learning from the SPC template, we obtain a new ontology to describe branded drugs, with the purpose of enriching the Semantic Nomenclature ontology network from a domain level perspective.

The core of the ontology is the Medicinal\_Product concept, with is described by the different properties detailed in the SPC template, which are re-engineered to concepts like Composition, Clinical\_Particulars hierarchy, Pharmaceutica\_Particulars or Pharmacological\_Features; or to datatype properties like dateauthorisation, daterenewal... This ontology was performed and formalized using the NeOn toolkit v1.2 with the OWL plugin.



**Figure 16: SPC Ontology relationship**

### *Ontology Details*

URI: <http://212.170.156.131:10000/ontologies#SPCOntology.owl>

Concepts: 31

Object Properties: 12

Datatype Properties: 11

Instances: 4

### **5.3 General Ontologies**

In [1], we separate some competency questions where we identified general terms related to different identified groups as Time, Measure and Location. In [1], we made a selection based on the requirements extracted from the competency questions, identifying temporal properties (dates for the pharmaceutical products), geographical properties (mainly location of laboratories or actors in the domain) or some measure and unit properties of the pharmaceutical products needed in the case study.

For each case, we analyzed the requirements against the different candidate ontologies in order to select and re-use one in the reference ontology according to the temporal properties, geographical properties or units' properties.

As a brief summary, in the following table are depicted the general ontologies selected based on the criteria detailed in [1] according to the case study's requirements for each general domain detected in the case study.

Domain	URL	Details
<b>Time</b>	<a href="http://www.w3.org/TR/owl-time/">http://www.w3.org/TR/owl-time/</a>	DAML Time Ontology (is in OWL format). Covers requirements as Time Points, Time Interval, Absolute and Relative Time, Relations between time intervals, Different temporal granularities...
<b>Geographical / Location</b>	<a href="http://simile.mit.edu/2005/05/ontologies/location">http://simile.mit.edu/2005/05/ontologies/location</a>	Minimalistic Location Ontology (Simile), formalized in RDFS. Covers requirements as Country, State, City, Location, Address...
<b>Units Measure</b>	<a href="http://www.cs.man.ac.uk/~horrocks/nsd07/galen.owl">http://www.cs.man.ac.uk/~horrocks/nsd07/galen.owl</a>	Simplified Galen Ontology (in OWL format). One of the most important ontologies in the Health domain, covers some requirements for units description as distance, weight (milligrams, grams, millilitres, temperature units, dosage...

**Table 3: General ontologies for the Semantic Nomenclature case study**

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.

#### 5.4 NeOn Toolkit Semantic Nomenclature Experience

In this section we provide a summary of the most important features and plugins of the NeOn Toolkit used in the development of the initial network of ontologies of the Semantic Nomenclature. The following table lists the plugins, why they are used in the scope of the case study, a brief assessment on their usability and tries to identify similar features presented in other tools.

<b>OWL Ontology Edition</b>	<p>The NeOn Ontology Navigator shows ontology projects, their corresponding ontologies, and the hierarchies of the current ontology. The Entity Properties panel is the main work area for defining and modifying objects of the selected ontology (accessed through different tabs). This panel should improve the global vision of all characteristics from a selected object (i.e. shows all properties</p>	<p>Other tools like Topbraid or Protégé<sup>17</sup> use form-based or graphical panel editors. In case of Topbraid Composer<sup>18</sup>, graphical edition is not free</p>
-----------------------------	--	--

<sup>17</sup> <http://protege.stanford.edu/>

<sup>18</sup> [http://www.topquadrant.com/products/TB\\_Composer.html](http://www.topquadrant.com/products/TB_Composer.html)

		of a selected concept).	
<b>R20 &amp; ODEMapster</b>	Non-ontological resource reuse Ontology Population	This plugin allows us to create mappings graphically (expressed in R2O language), execute and query the mappings between Digitalis and BOTPlus DB(MySQL databases) and their correspondent ontologies	Protégé provides the Ontobase <sup>19</sup> and DataMaster <sup>20</sup> plugins to turn any relational database into ontology. In this case, the R2O language allows us to define the concrete mappings between our model and the data repository.
<b>Alignment Server</b>	Ontology Alignment	This plugin allows us to semi-automatically match, retrieve, store and manage ontology alignments between the Semantic Nomenclature ontologies	Protégé provides Prompt to define mappings between two ontologies, but to our knowledge it does not detect the mappings automatically. The Alignment server also provides different algorithms for the matching process and means to store the mappings in a server. Topbraid provides a Query relational databases in real time functionality (in the Standard and Maestro licensed version)
<b>Cicero</b>	Ontology Documentation	This plugin allow us to keep track of main design decisions taken by the developers of an ontology	Collaborative Protégé is an extension that supports collaborative ontology editing as well as annotation of both ontology components and ontology changes. This plugin is not supported by Protégé 4
<b>RaDon</b>	Ontology Diagnosis Ontology Evaluation Activity Ontology Repair Ontology Validation Ontology Verification	Radon was used to check the consistency and coherence of the single ontologies of the network and handle the possible mistakes	Protégé allows connecting to reasoner for classification and consistency checking, but not provides means for automatic handle of inconsistencies in single or networked ontologies.
<b>Reasoner (Pellet or Kaon2)</b>	Reasoning	We configured the DIG interface to connect with Pellet reasoner in order to use with the repair and classification methods	Protégé allows configuring connection to reasoners. The Maestro and Standard version of Topbraid allow work with different reasoners and configure inference options
<b>SPARQL</b>	Ontology Search	The SPARQL Plugin allows us to do queries (using SPARQL syntax) against the OWL Nomenclature ontologies	SPARQL is supported in Protégé 3.x but not yet in Protégé 4. TopBraid supports create and execute SPARQL queries
<b>Watson</b>	Ontology Enrichment Ontology Reuse Ontology Statement Reuse	Watson was used in the prototype design as part of ontology repository.	There are other ontology repositories as Jena or Sesame in the market for this purpose, although Watson does offer different functionalities and has other goals apart from being a repository.
<b>OWLDoc</b>	Ontology Documentation	After the first version of the ontologies were finished, we used this plugin to generate HTML files that provide the documentation about the ontology and all its resources	Protégé has another plugin with same functionality to generate static HTML pages for publishing ontology documentation

<sup>19</sup> <http://protegewiki.stanford.edu/index.php/OntoBase>

<sup>20</sup> <http://protegewiki.stanford.edu/index.php/DataMaster>

**Table 4: NeOn Toolkit plugins used in the ontology network development**

Based on the training experience with this set of plugins and the use case scenario, we extracted some conclusions about how to improve the current plugin functionalities or other functionalities that should be covered by NeOn:

- In this version of the ontology network we found a functional specification of the NeOn methodology, but still a lack of tooling support for the different methodological steps in the NeOn Toolkit. We had to follow the methodology guidelines using the documentation provided in different deliverables and go back and forth to the different plugins of the NeOn Toolkit needed to apply the guidelines. This was fine in the scope of the case study, but sometimes hard to put into practice. At the moment of the edition of this deliverable, NeOn is working on providing a methodological tooling support. It is worth mentioning that for the final steps of the project the case study is using the Gontt plugin. This plugin supports the case study in carrying out the scheduling activity for the ontology network development. It provides methodological guidelines for each process and activity and information about the existing NeOn plug-ins for each process and activity; On the other hand, we plan to use the Kalima plugin which leverages C-ODO Light-based descriptions in order to support end-users in managing ontology lifecycles based on design principles (Kalima plug-in will equip the NeOn Toolkit with a design-oriented, rather than language-oriented user interface).
- OWL 2 Support: an extension and revision of OWL that is currently being developed within the W3C which will be more robust and solve some of the expressivity limitations, syntax issues and other problems detected. Also it is especially important for the use case the shift to the OWL Manchester API in the NeOn Toolkit and plugins, because our ontologies are OWL and it gives more flexibility and fits better with our initial requirements.
- Regarding the use case ontology development we also found a lack of more easily tools to ontology re-engineering of non-ontological resources. Most of the data found in the pharmaceutical scenario are stored in non-ontological resources like databases, documents (pdf, doc...). This information is data (instances) and we need to bring the information from the documents to ontological resources (RDF, OWL...).
- A graphical tool for manual OWL mapping between concepts of two ontologies loaded in the workspace will be useful
- More support to other databases engines in R2O & ODEMapster will be also useful
- SWRL support in the NeOn Toolkit, although not essential for the case study, would be also a very interesting feature.

## 6. Conclusions

In this deliverable, the second iteration of the ontology network that has been developed in the context of the Semantic Nomenclature case study in NeOn is described. For this purpose, we have reviewed and analyzed the work done in the previous deliverable to improve the problems addressed by the case study, and add the new requirements provided by domain experts detected in the case study scenario. The main requirement detected by domain experts is the need of disambiguate between clinical drugs (used in prescription) and marketed drugs (sold in pharmacies) in the Spanish scenario. Moreover, this new iteration of the ontology network was based on the NeOn methodology guidelines, which has been applied to the management of the knowledge lifecycle

Based on the SemanticHEALTH semantic interoperability recommendations for the eHealth domain and the new iteration of the ontology network lifecycle, the Semantic Nomenclature ontology network is reviewed in order to enable the interoperability between the different resources identified in the case study. For this purpose, the Reference ontology is reviewed and covers the ambiguity between clinical and marketed drugs, and makes possible the easy interoperability and integration of the distributed resources for the description of pharmaceutical products, and facilitates the aggregation of drug-related information in a semantic way due the reference ontology is connected via mappings with different pharmaceutical ontologies at different levels. Also, is provided a brief guideline about how the new version of the ontology network will be deployed and used by the Semantic Nomenclature prototype.

Moreover, we have provided a review of the inventory of the pre-existing resources (ontological and non-ontological resources) surveyed in order to accomplish the case study requirements and then refined it into the definitive set of resources that we have finally reused in the case study giving the motivation and reason to involve them.

The Ontologies involved in the Semantic Nomenclature ontology network are published on a Web-Site and in the NeOn ACollab repository.

## 7. References

- [1] Gómez-Pérez J.M, Pariente T, Buil C, Herrero G, Baena A, Candini J. *Ontologies for the Pharmaceutical case studies*. NeOn Deliverable D8.3.1, 2007.
- [2] Rector A. *Barriers, approaches and research priorities for integrating biomedical ontologies*. SemanticHEALTH SSA project Deliverable D6.1, 2008.
- [3] SemanticHEALTH partners. *Semantic Interoperability Deployment and Research Roadmap*. SemanticHEALTH SSA project Deliverable D7.1, 2008
- [4] 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.
- [5] Suárez-Figueroa M.C, Aguado de Cea G, Buil C, Dellschaft K, Fernandez M, García A, Gomez-Pérez A, Herrero G, Montiel E, Sabou M, Villazon B, Yufei Z. *NeOn Methodology for Building Contextualized Ontology Networks*. NeOn Deliverable D.5.4.1, 2008
- [6] Mari Carmen Suárez-Figueroa et al. *NeOn Development Process and Ontology Life Cycle*. NeOn Deliverable D.5.3.1, 2007
- [7] Gómez-Pérez J.M, Pariente T, Daviaud C, Herrero G. *Analysis of the pharma domain and requirements*. NeOn Deliverable D8.1.1, 2006.
- [8] Kohler S, Euzenat J, Herrero G et al. *Evaluation of mappings*. NeOn Deliverable D341 to be published in 2009
- [9] Sharp M, Bodenreider O, Wacholder N. *A Framework for Characterizing Drug Information Sources*. AMIA Symposium 2008. U.S. National Library of Medicine, Bethesda.
- [10] *RxNorm Release Documentation File Full Release*. U.S. National Library of Medicine, 2008
- [11] *Disease ontology description*. Found at <http://diseaseontology.sourceforge.net> 2008
- [12] *SNOMED Clinical Terms User Guide*, ITHSDO organization, 2008
- [13] *Medical Subject Headings MeSH Factsheet*, U.S. National Library of Medicine NLM, 2008
- [14] *NCI Thesaurus Semantics, National Cancer Thesaurus*, U.S. National Institutes of Health, found at <ftp://ftp1.nci.nih.gov/pub/cacore/EVS/ThesaurusSemantics/NCI%20Thesaurus%20Semantics.pdf>