

Towards the use of networked ontologies for dealing with knowledge-intensive domains: a pharmaceutical case study

Germán Herrero Cárcel

Tomás Pariente Lobo

¹ATOS Origin SAE – Atos Research and Innovation
Albarracín 25, 20037 Madrid, Spain
tomas.parientalobo@atosorigin.com
german.herrero@atosorigin.com

Abstract

Knowledge intensive sectors, such as the pharmaceutical, have typically to face the problem of dealing with heterogeneous and vast amounts of information. In these scenarios integration, discovery and an easy access to knowledge are the most important factors. The use of semantics to classify meaningfully the information and to bridge the gap between the different representations that different stakeholders have is widely accepted. The problem arises when the ontologies used to model the domain become too large and unmanageable. The current status of the technology does not allow to easily working with this type of ontologies.

In this chapter we propose the use of networked ontologies to solve these problems for the particular case scenario of the nomenclature of products in the pharmaceutical sector in Spain. Instead of using a single ontology, the idea is to break the model in several meaningful pieces and bind them together using a networked ontology model for representing and managing relations between multiple ontologies.

The Semantic Nomenclature is a case study that is currently under development in the EC funded FP6 project NeOn [1]. Among the main objectives of the case study, are helping in the systematization of the creation, maintenance and keeping up-to-date drug-related information, and to allow an easy integration of new drug resources. In order to do that, the case study tackles the engineering of a drug reference ontology, the provision of easy mechanisms for discovery, model and mapping of drug resources in a collaborative way, and the ability to reason on the context of user and ontologies to ease the mapping and retrieving processes.

1 Current Situation

One of the most important issues in the pharmaceutical sector regarding the description of medicines is that of having a common and unified nomenclature. Steps in that direction have been taken by different international bodies and organizations. A number of classifying systems, thesauri, taxonomical classifications and even ontologies have arisen in the last years. However, the current scenario is that there is no unified way of naming and work with drug-related information. In the next paragraphs a quick overview of the most used international classification schemas and nomenclatures are depicted.

The Anatomical Therapeutic Chemical Classification System (**ATC classification**) is one of the most widely used classification of drugs [2]. It is controlled by the WHO¹ 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).

¹ WHO: World Health Organization <http://www.who.int>

SNOMED (Systematized Nomenclature of Medicine) is a systematically organised computer processable collection of medical terminology covering most areas of clinical information such as diseases, findings, procedures, microorganisms and pharmaceuticals. The design of this Nomenclature is based on Description Logics. SNOMED CT is one of a suite of designated data standards for use in U.S. Federal Government systems for the electronic exchange of clinical health information [3].

Medical Subject Headings (MeSH) is a huge controlled vocabulary (or metadata system) for the purpose of indexing journal articles and books in the life sciences including drugs and pharmaceutical preparations. Created and updated by the United States National Library of Medicine (NLM), it is used by the MEDLINE article database and by NLM's catalog of book holdings. MeSH contains around 23.000 subject headings which are arranged in a hierarchy and could be viewed as a thesaurus [4].

Also, the globalisation in the marketing of drugs and sharing information among pharmaceutical professionals, the competent authorities and laboratories has contributed to the creation of terminologies as MedDRA. This terminology is used mainly in pharmacovigilance, due to MedDRA has a high number of terms for coding diseases, symptoms, diagnosis...

There are more thesaurus, taxonomical classifications, or medical languages as the Unified Medical Language System (UMLS) [5]. UMLS is a controlled compendium of many vocabularies, not only about pharmaceutical products, such as LOINC, RxNORM, HL7, NCI and other, which also provides a mapping structure between them.

Some medical ontologies such as Galen², OpenCyC³ or the NCI thesaurus⁴ define the concept of pharmaceutical product and attempt to classify different categories of drug-related information. But the current status is that there is a limited set of ontologies focused specifically on drugs, particularly in the description of pharmaceutical products.

Meanwhile, the Spanish governmental organizations are working in improving their Nomenclature providing their information about drugs in the Spanish market following mostly the ATC classification. Figure 1: Overview of the Spanish Pharmaceutical sector main stakeholders shows the main actors of the pharmaceutical sector in Spain, depicting the main life-cycle relationships.

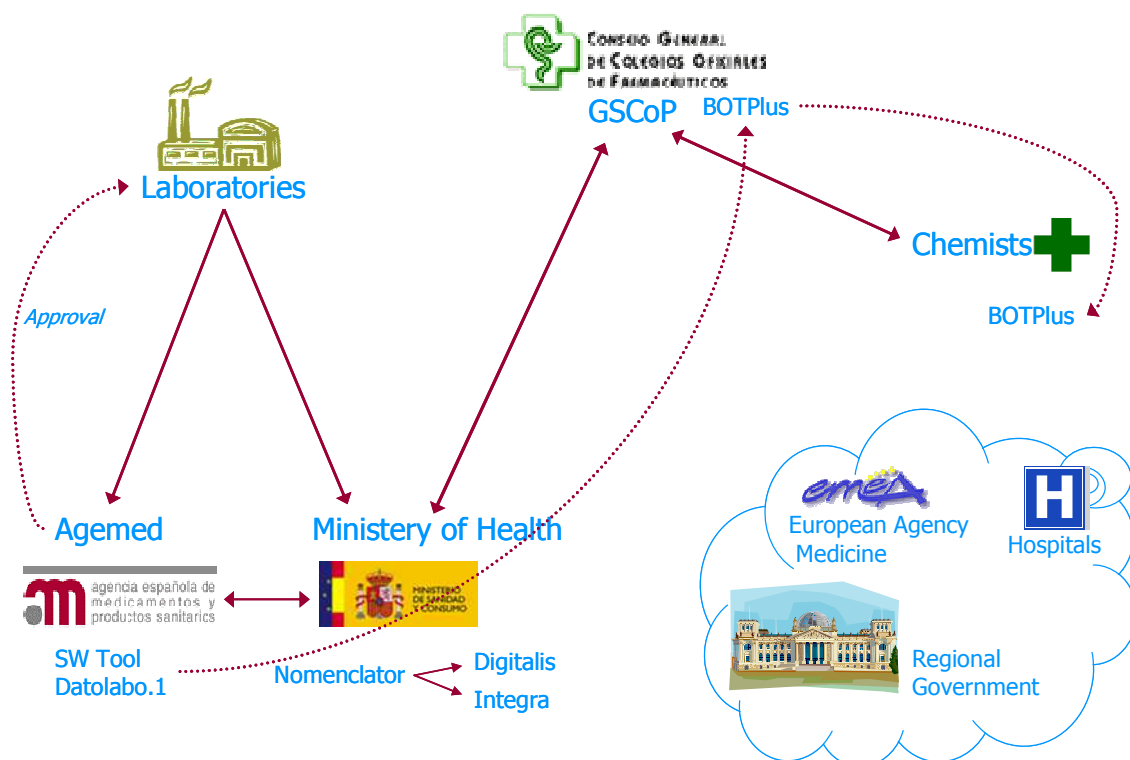


Figure 1: Overview of the Spanish Pharmaceutical sector main stakeholders

² <http://www.opengalen.org/index.html>

³ <http://www.opencyc.org/>

⁴ <http://www.cancer.gov>

The main actors represented in the previous figure are:

- The Ministry of Health edits and provides two official databases (*Digitalis*, *Integra*) with information about pharmaceutical products in Spain every month. *Digitalis* is the nomenclature officially used in the invoicing of prescriptions and contains data such as the identification of the pharmaceutical product, prices, composition of the medicine, etc. *Integra* stores information about pharmaceutical products consumed (and assimilated) in the hospital field.
- The Spanish General Agency of Drugs and Sanitary Products (Agemed⁵) provides a SW tool (*Datolabo*) with the information about all pharmaceutical specialities for human-use authorized or in process to be licensed. Reports and alerts about the products.
- Some other Regional governmental bodies have responsibilities for managing the drug market in their region delegated by the central government.
- The European Agency of Medicine is in charge of setting European-wide rules and recommendations in the pharmaceutical sector. EudraPharm is intended to be a source of information on all medicinal products for human or veterinary use that have been authorised in the European Union (EU).
- Among the responsibilities of the General Spanish Council of Pharmacists (GSCoP), are the representation, coordination and enabling co-operation among the pharmacy profession in Spain. GSCoP provides to its members the Health Information Data Base (*BOTPlus*) for having access to harmonised, updated information on medicines, patients, diseases, treatments, etc. It includes around 140.000 pharmaceutical products.
- Laboratories and wholesalers provide legacy systems and resources with descriptions about their own products.

The number available of drugs today in the Spanish pharmaceutical domain is enormous. The vast amount of information related with these drugs is difficult to gather, combine and comprehend. There are several lists of information about drugs available to the professionals, but the interoperability and easy cross-checking is far from being achieved. Using a standardized outline format would be desirable, but is not the current situation.

On the other hand, the information must be kept up-to-date. Thus the inclusion, modification or withdrawal of drugs is subject to a workflow or life-cycle that involves most of the stakeholders and it is performed according to the Spanish and European regulations.

In November 2004, the Spanish Ministry of Health published a strategic plan⁶ where some actions to promote the sharing of information about medicines between the professionals of the sector was proposed. One of the main goals of this plan was to reduce the costs derived from the promotion and searching of information, because these costs have a direct influence on the price of the medicines. A new nomenclature (widely called “vademecum” in Spain) with more advanced features would be of great help to achieve this objective.

In summary, there is a clear lack of systematization for the creation, maintenance and update of the drug-related information. Moreover, there is no reasoning involved in management of this information, and the interoperability is far from being achieved.

1.1 Problem Statement

The past decade has seen the rapid development of ontologies applied to the medical sector. There are some efforts devoted to develop ontologies in the pharmaceutical domain, but these ontologies are far from being mature and have not reached the Spanish sector at all.

One of the tasks of the W3C is to establish recommendations for generic technologies on the World Wide Web that operate between the underlying transport layer and the applications specific to individual industrial sectors. We can read from [7]: *For classifying symptoms, diagnoses, treatments and drugs, medicine abounds with indexing systems and controlled vocabularies. The Resource Description Framework (RDF) language from W3C is an XML application that supports the representation of*

⁵ AGEMED - Agencia Española de Medicamentos y Productos Sanitarios

⁶ <http://www.anisalud.com/ficheros/Plan.pdf>

controlled vocabularies, thesauri and term lists where the relations between the terms are weakly defined. Built on top of this is the OWL language to represent ontologies. An ontology is a conceptual model of a domain of discourse, it is not just a thesaurus or a term list. The relations between items in an ontology are defined much more strictly than in a simple term list or thesaurus. This enables more precise reasoning to weed out erroneously defined terms, and to dynamically generate new concepts and their definitions that can be derived from those already included. In the medical domain thesauri such as UMLS and ontologies such as UK Drug Ontology can be represented in these languages and then interact with the XML representations of other documents.

However, as stated before, there is a clear lack on ontologies tackling the description of pharmaceutical products. Besides, the number of drugs available today is enormous and the Spanish pharmaceutical sector is a very complex, heavily regulated sector. According to the GSCoP, this sector is even more complex and opaque than it seems. There are several stakeholders interacting with each other, but the communication and sharing of drug-related information is very limited. In fact, in the opinion of the GSCoP, a high confidentiality leads to a lack of communication and brings difficulties in the exchange of information between the different stakeholders.. As one of the activities promoted by the Spanish Ministry of Health, as it is stated in the law, is the collaboration and exchange of information about pharmaceutical products between the stakeholders, helping to solve this issue is critical.

The complexity of the sector has a reflection in the technological problems they face. Each stakeholder maintains the information in their own format. When an actor requires information from others, in the best case scenario the information is provided partially and in a different format. Consequently the data has to be reformatted and completed by digging on different sources. That means that not all information is obtained in an electronic format. For instance the AGEMED sends monthly a list of the new pharmaceutical products in paper to the GSCoP, containing a brief description of each product (around eight attributes). With this information, the GSCoP have to contact other actors to obtain complementary information. Depending on the stakeholder, the contact is made using different means of communication (fax, telephone, e-mail, etc.). Finally, once the information is retrieved, it takes 2-3 hours of a technician's time to update the pharmaceutical database. In some cases it is obvious that the information is obtained depending on the technician's abilities and experience in persuading laboratory and government workers to facilitate the required information.

This scenario suggests that there is a place for a platform that would allow different actors with diverse, complex and heterogeneous interests to develop a collaborative environment, in which the information and the processes, enriched with underlying semantics, will provide an improvement in the quality of service presented to the different stakeholders.

2 Proposed Solution

The need to share diverse knowledge and information with other applications already built has given rise up to a growing interest in research on ontologies. But, what is an ontology? *While* a variety of definitions of the term ontology have been suggested, *we* will use the definition first suggested by Studer et al [8] “... *an ontology is a formal specification of a shared conceptualization...*” Through the representation of domain-specific knowledge, ontologies provide a way of sharing and reusing knowledge among people and heterogeneous applications. So, ontologies are one of the core resources in any Semantic Web application enabling semantic interoperability and integration of data and processes.

In recent years, studies published by Gartner Group [9] ranks ontologies as third in their list of the top 10 technologies. According to this study, we are now entering a phase of knowledge system development, in which ontologies are produced in larger numbers and exhibit greater complexity.

High-intensive knowledge sectors, like the pharmaceutical, are the objective of the ontologies in order to build a new generation of semantic web applications which can deal easiness with large volumes of data and large and reusable resources which are complex to deal with them with other technologies.

This new generation of applications will thus reflect the fact that new ontologies are embedded in a network of already existing ontologies and that ontologies and metadata have to be kept-up-to-date with the changing application environments, sector and users' needs.

Based on this scenario, semantic integration by means of one global semantic model is probably unmanageable. In contrast, a solution for the semantic integration based on a network of contextualized ontologies provides more facilities for maintaining a set of ontologies locally consistent and it is clearly easier to manage.

However, in order to work with several interlinked ontologies a set of new tools, methodologies and techniques are needed. The NeOn project addresses this challenge, offering a set of methodologies and technologies for realizing this new generation of semantic applications, from foundational research for contextualised networked ontology environments to application-level demonstration of collaborative ontology creation, population, maintenance and access. Consequently, the Semantic Nomenclature case study will use NeOn as foundations.

2.1 Objectives

As is described before, there is no common model or significant development on ontologies of drugs in the Pharmaceutical sector in Spain. One of objectives of this scenario is cover this lack, where our proposal is to create, from existing sources and schemas, a common reference model of the Semantic Nomenclature valid for all the stakeholders. In simpler words, Semantic Nomenclature consists of the creation of a reference model for a Semantic Nomenclature that complements typical pharmaceutical compendium characteristics by giving flexible, extensible and reliable information about drugs to the users of the Pharmaceutical domain.

2.2 Overview

NeOn provides the following definition: *Network of Ontologies is a collection of ontologies related together via a variety of different relationships such as mapping, modularization, version, and dependency relationships. We call the elements of this collection Networked Ontologies* [10]. And define four main ontology assumptions: *Dynamic* (ontologies will evolve), *Networking* (ontologies are interconnected via mappings or by means of reuse), *Shared* (ontologies are shared by people and applications), and *Contextualized* (ontologies are dependent of the context in which are built or are used) [11].

The reference ontology model of the Semantic Nomenclature will be based in this definition. In this sense, the reference ontology model is developed based on the schemas of the main sources of information about drugs in Spain: Integra, Digitalis, BOTPlus and in the ATC Classification. Moreover this reference model will be mapped to external sources of the same or similar domain, such as Snomed, connecting the reference model with other international medical vocabularies.

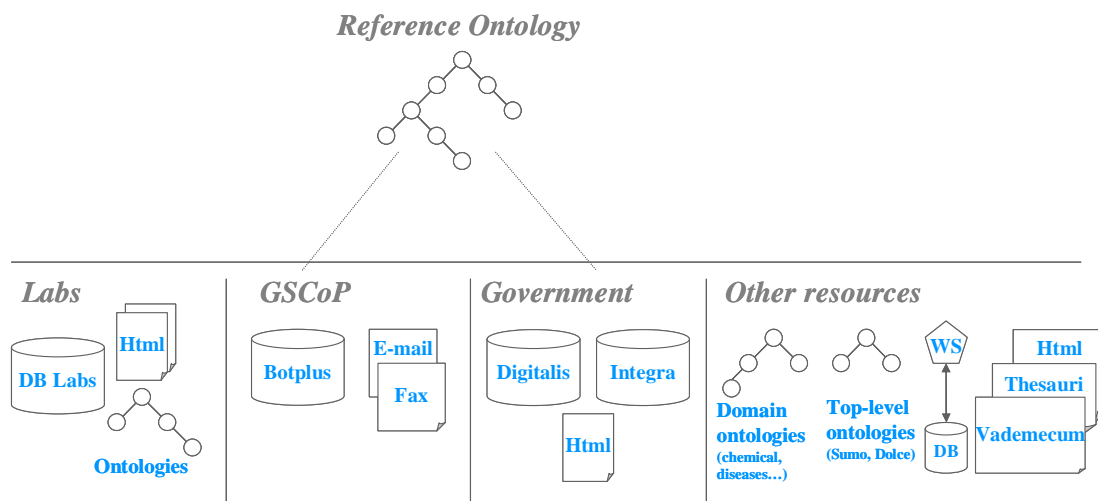


Figure 2: Reference ontology model and drug data sources

Based on this solution, the case study provides a collaborative scenario to ease the communication between the stakeholders, including an application that allows users to query and process all this information in order to get and make use of all this knowledge and infer new one.

The first version of the reference ontology, plus a set of core ontologies (Digitalis, BOTPlus, Integra and ATC), was developed in 2007. In 2008 the ontology network will evolve and will be mapped with other ontologies, such as Snomed, MedDRA, EudraPharm, etc.

The aim of the Semantic Nomenclature application is to improve the medical product information management within the Spanish pharmaceutical domain which is presently decentralized in several data bases and repositories. The application uses the pharmaceutical ontology network, which contains the Pharmaceutical Reference Ontology, which centralizes the key medical product information and links the rest of ontologies each others. This ontology network is especially useful for allowing a more complex information retrieval as well as resolving medical product identification and different classification alignments. In spring of 2008, the first version of the Semantic Nomenclature prototype application for the end users will be provided.

3 Solution Details

3.1 Ontology development

3.1.1 Methodology followed

The realization of real-life applications in the Semantic Web requires the ability to deal with heterogeneous ontologies fragmented and distributed over multiple autonomous nodes.

Next generation semantic applications will be characterized by several ontologies linking each other and networking and sharing knowledge. The aim of NeOn project is to create the first ever service-oriented, open infrastructure, and associated methodology, to support the development life-cycle of such a new generation of semantic applications that will rely in a network of contextualized ontologies.

As is described before, a *Network of Ontologies* is a collection of ontologies related together via a variety of different relationships such as mapping, modularization, version, and dependency relationships. So with the help of networking ontologies and methodologies, technologies and tools provided by NeOn, we aim to fulfil the requirements posed by the pharmaceutical sector. We will be able to carry out the development of a new generation semantic applications able to achieve a better management, integration and exploitation of the large quantity of knowledge required by this particular sector and also provide new functionalities and services to the business.

Using networked ontologies instead of a single, centralized one, will allow to the different stakeholders of the pharmaceutical sector to maintain their current models (formalised as ontologies), while the NeOn infrastructure and modular framework takes care of their interlinking and mapping.

Currently there is not a methodology to work with networked ontologies. NeOn will provide methodology to support the collaborative construction and dynamic evolution of networked ontologies in distributed environments [12]. In the mean time this case study is following the Methontology guidelines.

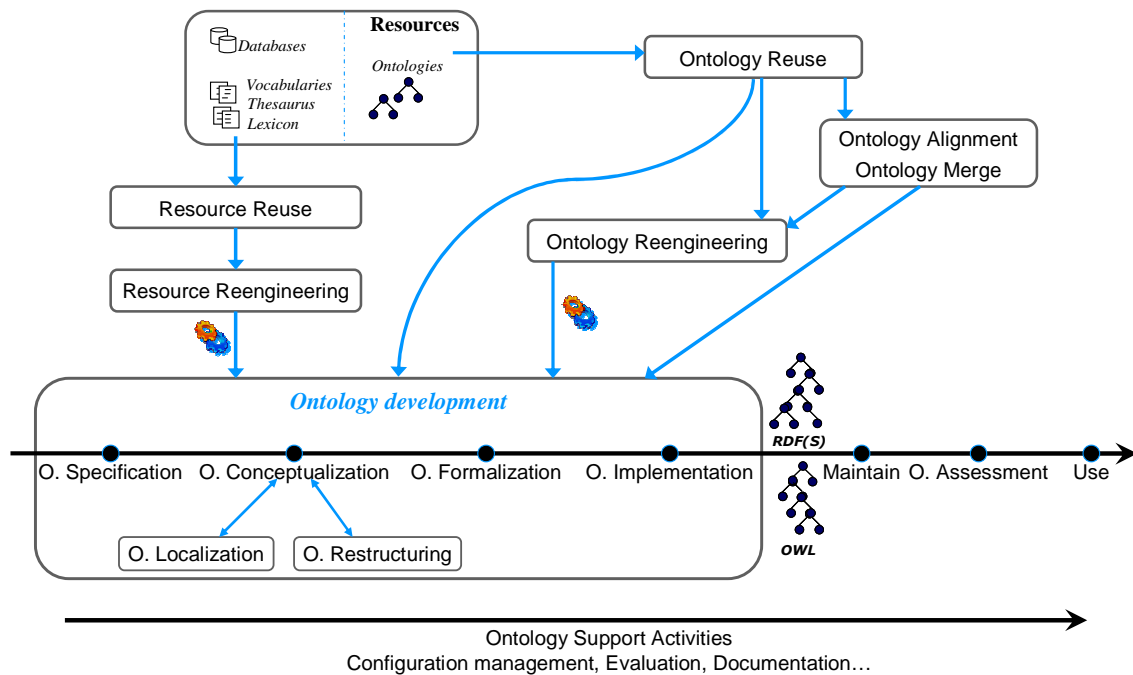


Figure 3: NeOn methodology

It was decided that the best method to adopt for this investigation was to use the recommendations and suggestions coming from NeOn to develop networked ontologies in a distributed scenario. This methodology is an extension of the Methontology [13] more dedicated to networked ontologies and the new generation of semantic web applications. In broad terms, the proposed methodology describes different activities and tasks identified in the ontology lifecycle development and each simple scenario is an instance of the needed activities needed in each specific scenario.

According to the methodology, first, the ontology engineer must specify the needs that the ontology has to satisfy in the new application, their intended uses and which are the users of the ontology. For this preliminary task, the methodology suggests to specify different competency questions in order to help in solve the first doubts, and identify the objects of the discourse of the domain. In the Semantic Nomenclature scenario these objects of discourse are medical products, active ingredients, substances, laboratories, pharmaceutical forms, administration method, etc. After this specification, the ontology expert should identify terms that belong to common ontologies like terms related with time, geographical information, etc.

In parallel, the methodology recommends a review of the inventory of resources needed in the scenario, such as standards and classifications, thesaurus, vocabularies, taxonomies or ontologies that could be reused. The goal of this study is the reusability of resources. The ontology engineer searches, evaluates and selects the most appropriate resources. The nature of the resource will lead to include the ontology in the networked ontologies, or perform ontology reengineering tasks.

Finally, the ontology expert should carry out the conceptualization of the objects of the discourse of the domain, considering tasks such as of ontology localization and modularization, pruning, etc. Then the ontology is formalized and implemented and ready to be used and maintained in the new semantic application.

Around this ontology lifecycle, the methodology suggests different ontology support activities that should be taken into account, like configuration management, documentation, elicitation, evaluation, etc.

3.1.2 Towards a Semantic Nomenclature reference ontology

It has conclusively been shown that the Spanish pharmaceutical sector lacks of a ontology that encompasses the description of pharmaceutical products. According to the characteristics of the scenario and the recommendations of the methodology, the use of a network of ontologies for representing the reference ontology in the pharmaceutical sector seems to be a good solution.

This reference ontology model should be a compilation of the main terms and objects related to drugs, the general aspects of them and classify this pharmaceutical terms according to the ATC classification, since it's the WHO recommendation and is followed by the pharmaceutical experts in Spain (Europe).

Also, this reference ontology model is connected with the ontology models of the main databases which contain the information about the pharmaceutical products available in the Spanish market as Digitalis or BOTPlus. In the end, the reference ontology could be linked with the main medical vocabularies used in the world and should facilitate the integration of new resources or ontologies that will appear in the evolved scenario.

3.1.3 The Semantic Nomenclature network of ontologies

For the purpose of developing a network of ontologies for the Semantic Nomenclature scenario, we followed the modular approach suggested by the NeOn methodology. 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.

The ontology network lifecycle model decided was the iterative/incremental model. This decision was taken up based first on the decision tree provided by the NeOn methodology suggested in [12], but also from the identification of requirements of the Semantic Nomenclature Ontology network, and on past experiences developing ontologies.

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

Following the methodology described before, first of all, several competency questions were defined. Specific competency questions related with the Pharmacist are for instance: What is the drug commercial name? What is its Spanish national code? Which one is the drug therapeutical WHO group?, etc.

From the competency questions, we extracted the terminology that is formally represented in the ontology by means of concepts, attributes and relations. Terms (also known as predicates) and the objects in the universe of discourse (instances) are identified.

From this terminology, we checked that some of the terms (dates related with pharmaceutical product, dosage of the drugs, location of the laboratory manufacturers, etc), were related to common domains such as time, geographical or measure. Accordingly, we reviewed existing ontologies on these domains, and evaluated them. The result was mapped to the reference ontology.

The next steps was studying standards and resources, evaluate them and starts defining the domain ontologies of the model. Starting from the basis of the glossary and terminology extracted, the pharmaceutical reference model was modeled, based on the main schemas and the ATC classification.

In parallel, the selected standards and pharmaceutical classification systems, thesaurus, taxonomies and vocabularies were reengineered in order to provide a semantic enrichment before reusing them in the model. These resources are the main databases in Spain (Digitalis, Integra, BOTPlus). The ATC classification and Snomed vocabulary are considered as the most important de facto standards that connect and the Nomenclature network ontology.

Finally, the main source of knowledge and data in this scenario, the pharmaceutical databases, were reengineered and modelled as ontologies. In this phase, Digitalis, Integra and BOTPlus are modelled as ontologies and populated their models from the databases. Concepts from these models are networked with the Nomenclature ontology network, integrating their information.

Other resources suggested by pharmaceutical professionals were the online vademecum (vademecum.es), and medical vocabularies. Thesaurus and taxonomies (UMLS, MedDRA) from international bodies related with pharmacy were analyzed trying to find how the pharmaceutical product are described in order to find connections between them and the our models.

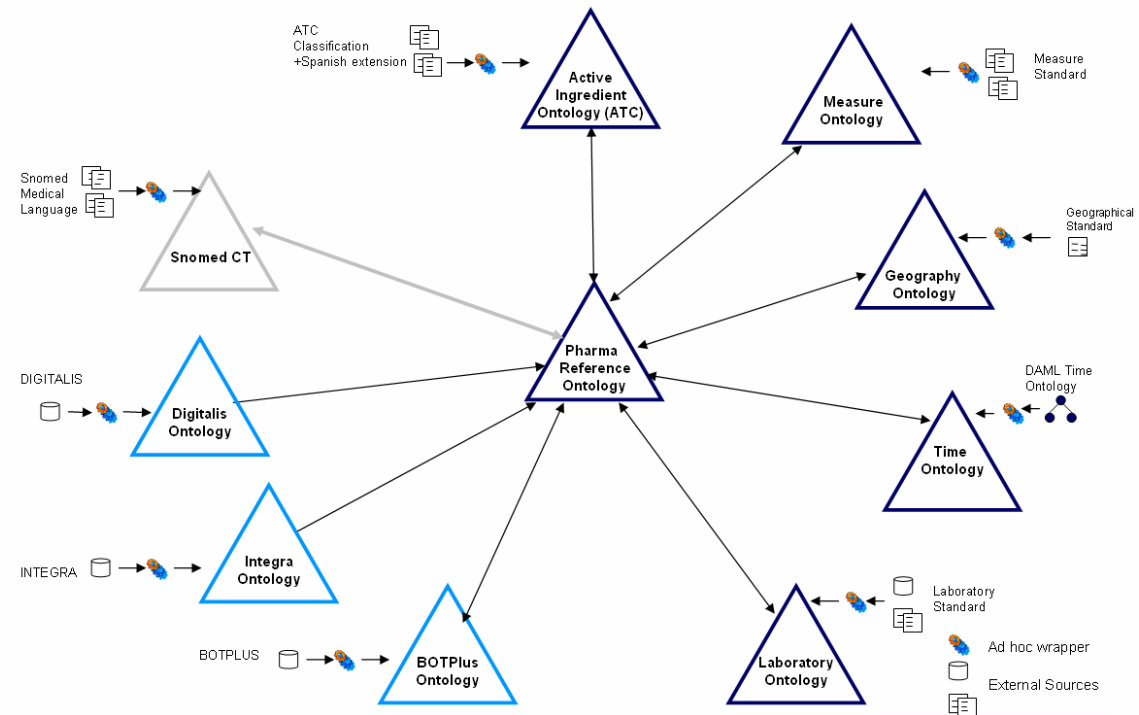


Figure 4: The reference ontology as a network of ontologies

Figure 4 shows the Semantic Nomenclature Ontology Network. The pharmaceutical reference ontology model is enriched with the general ontologies (Time, Location, Units...), and connected via mappings with the ontology models as Digitalis or BOTPlus. In addition, the pharmaceutical reference ontology is related with the most important classification system ontologies as the ATC or Snomed (in grey meaning that is due for a second stage of the network). Finally, in new iterations of the lifecycle model, the pharmaceutical reference ontology should integrate new resources or ontologies related with the main medical vocabularies used in the world.

This ontology network outlined in this chapter is aligned with the goals and expectations extracted from the pharmaceutical scenarios depicted before: integration of existing pharmaceutical resources. The ontology network architecture facilitates the aggregation of drug-related information, because the reference ontology is mapped and related with a set pharmaceutical ontologies at different levels. This network will be useful as well as foundation of one of the Semantic Nomenclature application objectives: The update of the BOTPlus database. This is a perfect scenario to show the possible business impact of networked ontologies. The pharmaceutical product information gathered in the networked ontologies is an added value to bodies such as the GSCoP in order to improve their commercial database reducing their updating process effort and complementing typical pharmaceutical compendium characteristics by giving flexible, extensible and reliable information about drugs to the users of the Pharmaceutical domain

3.1.4 Architecture of the Semantic Nomenclature network of ontologies

The Nomenclature Ontology Network is organized in three levels: the Pharmaceutical domain ontologies, the Application domain ontologies and the General ontologies. Figure 5 shows the levels of the Nomenclature Ontology Network based on reusability and usability of the levels.

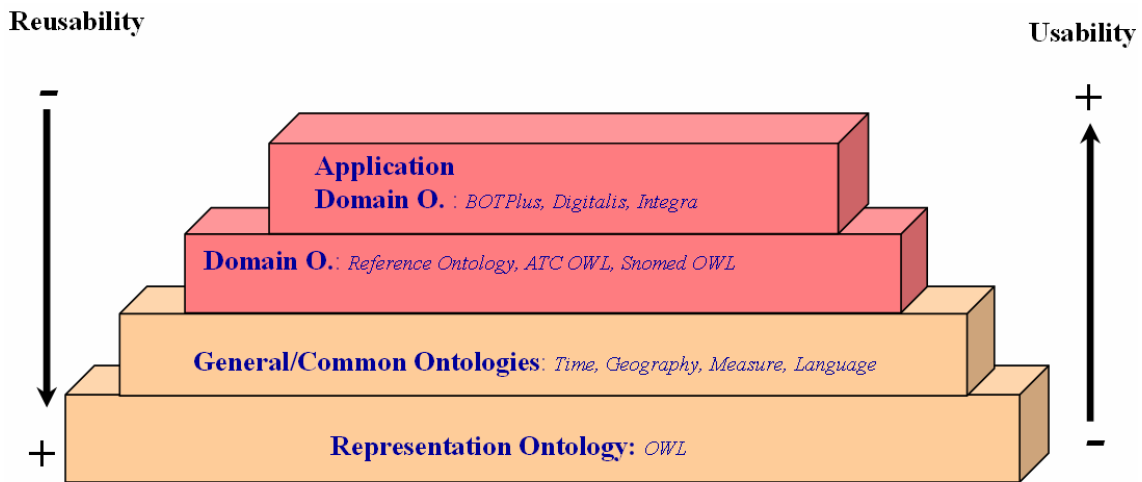


Figure 5: Modular approach for ontology construction

The representation ontology language (OWL) is at the base of the pyramid showed in the previous figure. On top of it, the reused vocabulary is depicted.

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

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

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

3.2 Architecture of the case study

3.2.1 The Semantic Nomenclature and the NeOn architecture

The Semantic Nomenclature application is based on the NeOn architecture and the functionalities that NeOn provides. The general architecture of NeOn is ordered into three layers. The layering is done according to the level of abstraction of the data and the process flow between the components.

In the lowest layer, called Infrastructure services, are classified the basic services like repository service, reasoner...needed by any semantic application. In the middle layer are included all the ontology engineering components developed in NeOn. These components realized their tasks over the infrastructure services, and they are differentiated between tightly coupled components (editors, mapping, visualization algorithms, etc), and loosely coupled services, such as annotation or collaboration support.

The GUI components are placed at the top level of the NeOn architecture. In this layer this components are user front-ends for the engineering components and for the infrastructure services.

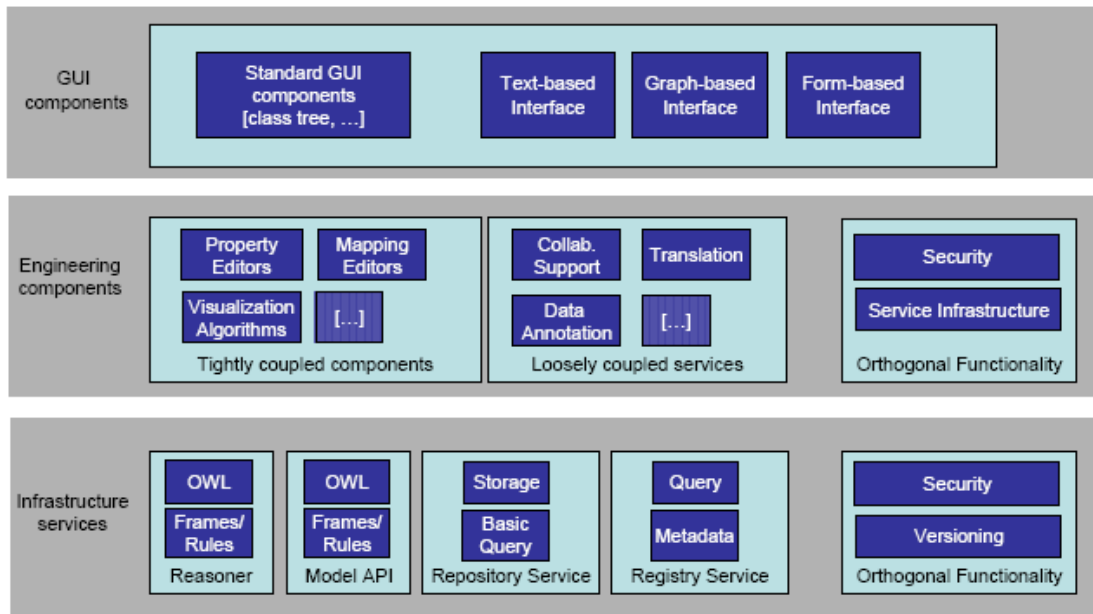


Figure 6: NeOn (Toolkit) Architecture [14]

NeOn provides a number of services which allow managing and supporting the entire ontology lifecycle, from design to exploitation. The main properties of the NeOn architecture are its adaptability and extensibility. These two properties are crucial for the use and deployment of the NeOn infrastructure in the context of the pharmaceutical scenario, where NeOn infrastructure will be customized to support specific user requirements.

Figure 7 describes the Semantic Nomenclature architecture, where different blocks with different functionality can be distinguished. These blocks are related to the three layers differentiated in NeOn architecture.

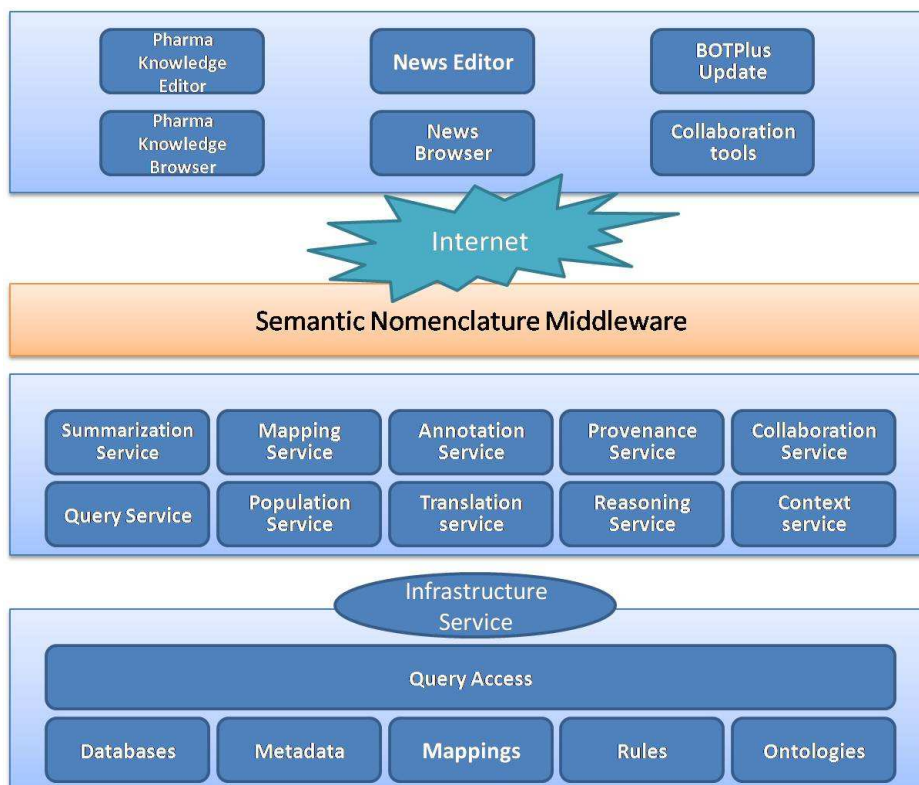


Figure 7: Semantic Nomenclature architecture

At the infrastructure block, the Semantic Nomenclature is supported by the services provided by NeOn, such as the model API, a repository service, a reasoner, etc. Above this block, the application contains the engineering components needed for the purpose of the network of ontologies and the semantic application. In our case, we identified services for ontology experts, such as the mapping service, annotation service, etc. Also services for the end-user (pharmacist) such as the query service, population service, etc., were identified.

The Semantic Nomenclature middleware is located between the GUI and the engineering layers. This block manages the communication between the visualization and engineering components blocks. The visualization components interact and link to the different services across the Semantic Nomenclature middleware and provide to the users a graphical view of these services. With these services, the users will interact with the different resources stored in the infrastructure services and can create/modify them. This block is the one which manage all the workflow of the different scenarios identified.

The visualization block contains the user interfaces that interact and show to the end user (ontology experts & pharmacists) the services to edit ontologies and visualize, browse and validate the pharmaceutical knowledge.

3.2.2 The architecture for Ontology Engineers

Based on the problems detected on the pharmaceutical scenario and the solution proposed, two sub-scenarios can be differentiated according to the nature of the tasks to be developed and the users involved. On the one side, ontology engineers need an architecture that allows them to develop and support the entire network ontology lifecycle and on the other side a semantic web application which allow pharmacists browse, query and reason over the Nomenclature Network ontology and keep update their databases has to be provided.

In the first scenario, the ontology engineer need all the components required to support building and maintaining the network of ontologies. Derived from the scenario, different functionalities as ontology edition, mapping edition, annotations, population and contextualization of ontologies are needed to develop the scenario. These functionalities are provided and covered by the NeOn toolkit. Also, NeOn Toolkit provides different visualization components are identified in the scenario to facilitate these tasks to the ontology engineers and show the results. Moreover, the infrastructure services provide a repository and access services to the engineering components for the different resources of the Nomenclature networked ontologies.

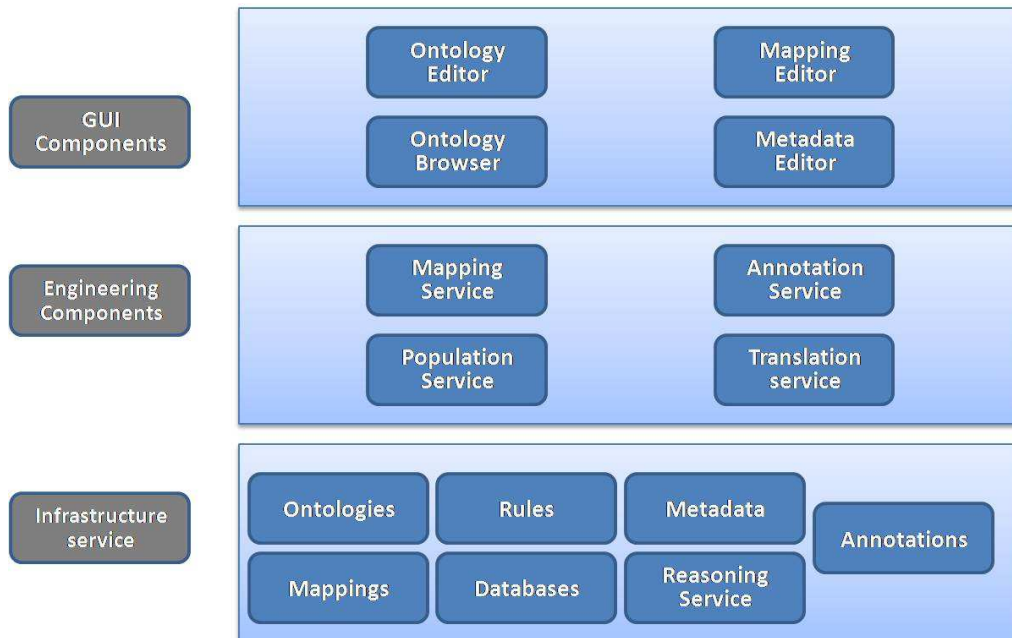


Figure 8: Semantic Nomenclature Ontology Engineers architecture

3.2.3 The architecture for end-users

Services needed by domain experts are related with searching, querying, navigating and updating information about drugs. To fulfil this scenario the architecture should provide different functionalities than the previous one. End-users are not ontology experts, so they need a user interface that should provide a higher level of abstraction of the underlying ontologies

The end-user will interact and navigate with the Semantic Nomenclature network ontology in order to search and select information related with the Spanish pharmaceutical medicines. The engineering components provide functionalities such as summarizing the content and rating the provenance of the sources, querying the network of ontologies about the latest changes in the pharmaceutical domain, reasoning services, and population and updating of databases.

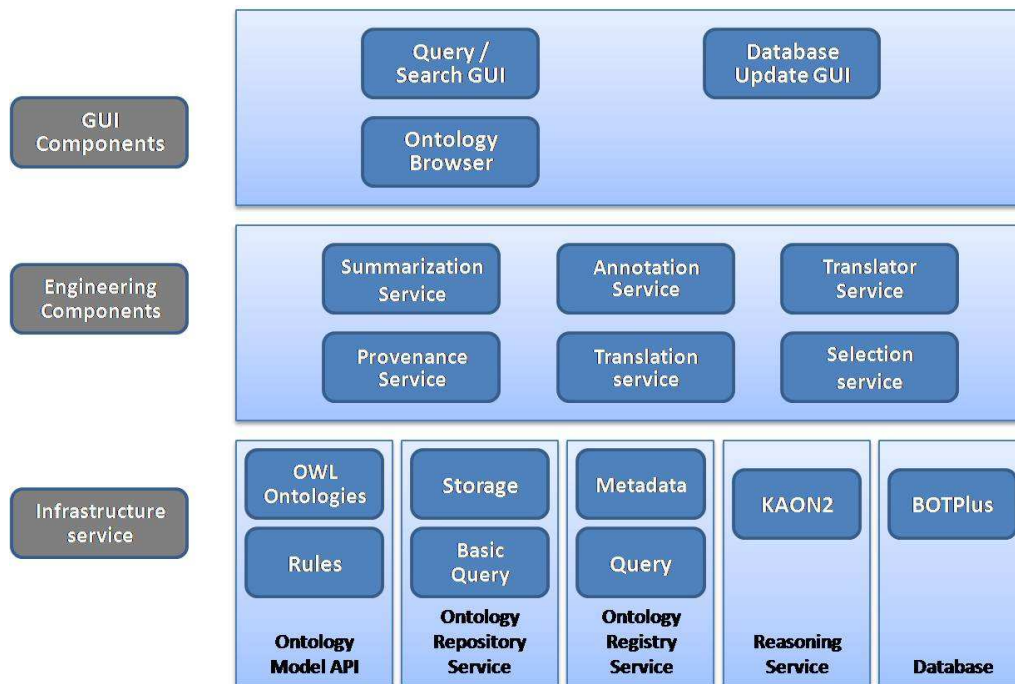


Figure 9: Semantic Nomenclature End-users architecture

3.3 Implementation

3.3.1 Usage and adaptation of the NeOn toolkit to the Semantic Nomenclature

The different services, components and functionalities identified in the three layers of the architecture are covered by different tools and services of the NeOn toolkit, mainly at ontology engineering components and infrastructure services.

In the first sub-scenario, one of the most important services is a GUI component for creating and editing the Nomenclature network ontology. This GUI is provided by the NeOn toolkit, which evolved from the latest versions of Ontostudio⁷, allowing ontology engineering in OWL format. Moreover, the NeOn toolkit provides other tool for model ontologies called Ontogen, a system based on machine learning and text mining techniques [15]. Ontogen suggests concepts and relations extracted from a large and mass quantity of documents provided by the user at the beginning of the construction process.

Other important issue in the scenario is the mapping functionality. This functionality is crucial for the ontology engineer in the development and lifecycle of the networked model. The NeOn toolkit provides a Alignment Server [16] for discovering and suggesting mappings based in their context between two ontologies in an automatic way. Also, NeOn toolkit provides a mapping metamodel and graphical means for draw mappings manually between different models.

⁷ http://ontoedit.com/content/index_eng.html

Other important aspect of the solution is the connection between the main databases and the ontology models. For this purpose, NeOn toolkit and the UPM provide a technology called R2O & ODEMapster [17]. This technology is based on the declarative description of mappings between relational and ontology elements and in an exploitation of mappings by a domain independent processor. R2O & ODEMapster is a useful framework to upgrade relational legacy data (databases) to the ontologies.

For ontology annotation, NeOn provides two tools for this purpose. Magpie is an annotation tool that uses the ontology infrastructure to semantically mark-up web documents on-the-fly [18]. Magpie tool aims to identify and filter out the concepts-of-interest from any webpage it is given. The other annotation tool called GATE provides an easy to use interface for indicating which pieces of text denote which of your concepts of your ontologies, doing the annotation manually or automatically [19]. In the pharmaceutical scenario annotation is clearly identified and used when some source of knowledge are identified as reports or documents from the government about alerts, news or recommendations about the drugs and their content should be annotated into one ontology model of the Nomenclature network ontology.

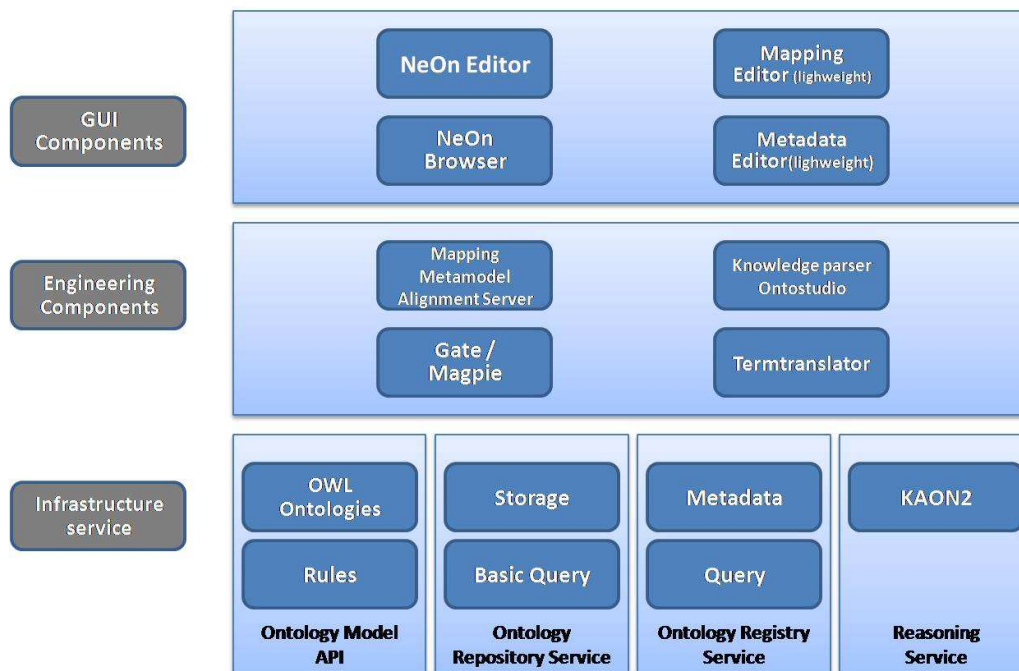


Figure 10: Instantiation of NeOn architecture for Semantic Nomenclature (I)

Other functionalities identified as the translation service or multiguage of the model are covered by the multiguage model provided by NeOn and the TermTranslator tool. Summarization, provenance and trust service are supported by the NeOn toolkit in order to help ontology engineers and end-users in their analysis of the purpose of the ontologies and data in the application.

The Infrastructure services block will be in charge of storing all the data the system needs. It stores ontologies, metadata, annotations, mappings, queries, rules that the users have designed and the most relevant databases in the Spanish pharmaceutical domain. For the end-users, NeOn provides Query service, a registry to access all the elements and Reasoning service at infrastructure level. The Query service is quite important scenario. In the Semantic Nomenclature application we provide searching facilities to the users, in order to find the information in a faster and more accurate way. This Query service attack the infrastructure services and resources in order to find this information to the user, and should provide facilities to build queries in natural language or similar. This natural language query functionality is beyond the repository service of the basic NeOn infrastructure, and it is not one of the main priorities of the case study. ORAKEL, an engineering component from University of Karlsruhe tackles this functionality [20].

The reasoning service is a crucial service in the case study. Using this service, the Semantic Nomenclature can check the Semantic Nomenclature system when a new pharmaceutical resource is added, or can search about new resources provided by the networked ontologies in order to add it to the system or find new information and pharmaceutical data across the different information sources in order to update the BOTPlus database. NeOn includes KAON2 as reasoning service [21]. KAON2 is an infrastructure for managing OWL-DL, SWRL, and F-Logic ontologies. The main features of the

infrastructure are an API for programmatic management of OWL-DL, SWRL and F-Logic ontologies, a stand-alone server providing access to ontologies in a distributed way, an inference engine for answering queries expressed in SPARQL syntax, and a module for extracting ontology instance from relational databases. The main relevance for the Semantic Nomenclature scenario is the KAON2 ontology management, inference service (in order to search new drug information across the ontologies) and the module for extract ontology instances from the main pharmaceutical databases in Spain (BOTPlus, Digitalis, Integra).

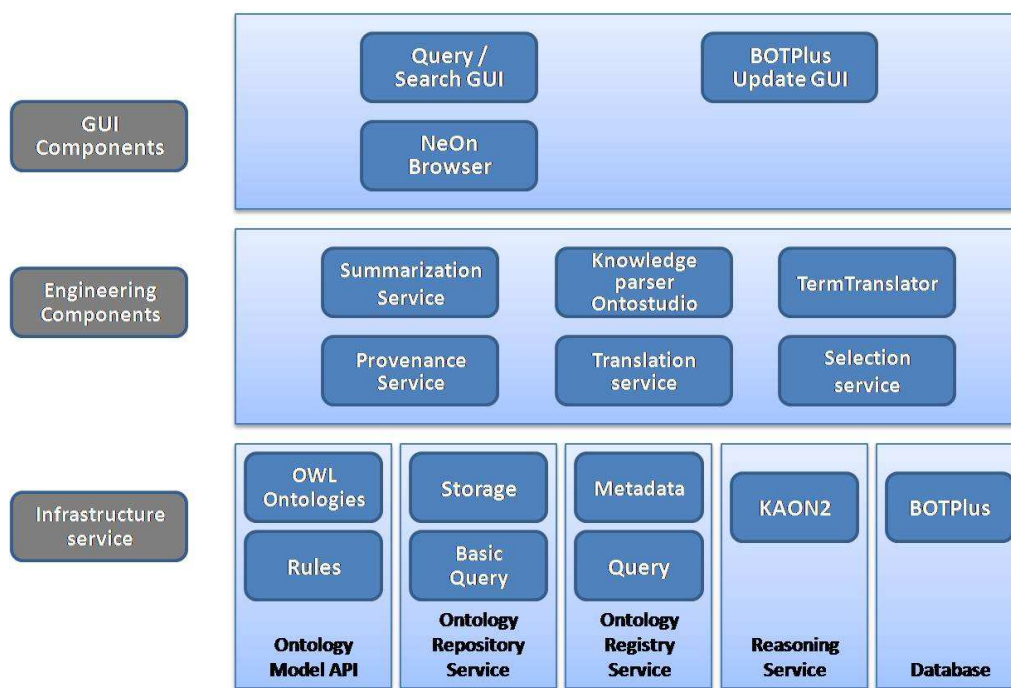


Figure 11: Instantiation of NeOn architecture for Semantic Nomenclature (II)

The communication between the GUI components and the services will be based on HTTP and Web services for the loosely coupled services like the “Pharmaceutical Browser” and a tightly coupled client. The application will be web-based.

3.3.2 Business case implementation of the case study

There are four main application functionalities and services that the Semantic Nomenclature will provide to end-users: drug knowledge alert system, news service, collaboration tools and update service..

One relevant functionality is the Drug knowledge alert system. The Semantic Nomenclature application will select relevant drug information and pharmaceutical alerts to inform to the pharmacists, taking into account their preferences and experience. As a result of that, the pharmacists could access to the latest information provided by the main information sources (government, laboratories and professional entities).

The News Service provides to the users additional information and news about pharmaceutical extracted from web pages, documents, and other resources,. related with the pharmaceutical professional and suggested by the end-users.

Other relevant and innovative functionality is the collaborative scenario through a social network among professionals for a better communication between the pharmaceutical stakeholders. The collaborative tools could provide to the pharmaceutical community a new way of sharing knowledge. A Semantic Nomenclature wiki could be very useful for the end-users, because this tool could provide accurate content and the pharmacists could use it in their everyday work. Furthermore, the end-users have the chance to edit the content, validate it, exchange and share experiences and subscribe to their topics of interest. Moreover, this collaborative tools and the information provided by the pharmacists could be used to analyzed and processed by the Semantic Nomenclature in order to annotate this information to the correspondent elements in the ontology network. These annotations could be offered as added-value to the knowledge of the Nomenclature network.

The Update service is one of the core facilities provided to particular actors to update their legacy systems with the latest information. The main pharmaceutical actors and end-users of the information can take advantage of the Nomenclature network of ontologies in order to gather information about the pharmaceutical products from different sources it and incorporate into their legacy systems. This information is checked and provided with a provenance rating and the end-user can validate it before update their legacy. We aim to implement an example of this service during the NeOn project life-span.

4 Alternatives

As it was explained before, the vast amount of information related with drugs in the Spanish pharmaceutical sector is difficult to gather, combine and comprehend, but the interoperability and easy cross-checking is far from being achieved. Besides, there are not several alternatives for solving the lack of communication between the main actors in the pharmaceutical business scenario.

As is described in the beginning of the chapter, in recent years the Spanish Ministry of Health published a strategic plan [22] where some actions to promote the sharing of information about medicines between the professionals in order to reduce the costs derived from the promotion and searching of information. A new nomenclature with more advanced features would be of great help to achieve this objective.

The main competitors in the Spanish case study are the current applications already available, as BOTPlus or vademecum on-line. These resources are the typical compendium of characteristics of the pharmaceutical products, but these applications not cover the lack of communication and integration of the information. In both cases, the tools are not free of charge, although some information and searches are publicly available in throughout the web.

At semantic web technology level there are not real competitors in Spain or Europe. In the USA market, Snomed and UMLS are the most important vocabularies, largely used by the official entities in their applications. But the penetration on the European market and mainly in Spain is not significant.

5 Cost and Benefits

The Semantic Nomenclature focuses on a real domain with information heterogeneity problems that have direct economical impact on the pharmaceutical sector. The main result of the Semantic Nomenclature application is the Nomenclature network ontology. The potential of the networked ontologies would serve as a proof of concept for the pharmaceutical sector.

However, all the previously mentioned solution and technologies suffer some costs. Although the case study software will be delivered as open source, in order to develop a fully-fledged application, there will be several steps with associated costs.

Although much of the NeOn toolkit is free of licensing costs, there will be a number of engineering components and services that will be exceptions. At the present time, NeOn is a project in progress, so these issues will be further clarified during the duration of the project.

Also, just as the pharmaceutical sector and market evolve, the prototype should be customized for the real client needs. That would mean for instance review the requirements of the pharmacists, reengineer the network of ontologies, add new ontologies or create new software modules.

Another problem with this approach is that is needed some time for personnel training for ontology engineers, dedicated personnel to the management of the resources (finding new resources, rating information, reengineering the ontologies in case the systems evolves, etc.) and specially for end-users which are not related with the semantic web technologies.

In contrast, the Semantic Nomenclature provides some different benefits for the pharmaceutical scenario. The two main concepts of the solution provided for the scenario are integration of diverse systems together and interoperability of the information used by the different actors of the pharmaceutical sector. Networked ontologies is the better way of dealing with information coming from multiple resources and allow to the actors develop and maintain their own heterogeneous models, mapping one to other, integrate and exchange information in a dynamic view. Moreover, this solution reduces costs for all stakeholders, against trying to solve their problems of integration with one of each of the other stakeholders separately.

Other benefit is that the Nomenclature Reference ontology will serve the purpose of bridging the gap between different classifications and nomenclatures. Instead of oblige everybody to follow the same vocabulary (with has proven to be difficult), the Semantic Nomenclature bridges the gap between vocabularies encompassing current and future resources.

Another important finding is the better and new semantic-based search facilities. Against other traditional types of searching mechanisms, the idea of "personalized search" actually requires semantic capabilities and provides facilities to build queries in natural language or similar.

Interestingly, the most striking result to emerge from this business scenario is that the Semantic Nomenclature prototype is a clear example of how to improve the process of updating a private stakeholder (GSCoP). This organization could keep their legacy system because it is described abstractly in an ontology, connecting it with other models, enable NeOn infrastructure and Semantic Nomenclature business prototype as a bridge between models and integration solution.

6 Risk Assessment

In this business scenario and solution, different types of risks need to be taken into account, including technological risks, resources risks and market and exploitation risks.

At technological level, NeOn should reach a wide technical community in order to enrich its platform and to be accepted as a good ontology development toolkit. Technology changes and evolves requiring redesign of NeOn architecture/tools adapting the NeOn toolkit to them.

To take a stance against the market and exploitation risks, the Semantic Nomenclature must be presented in pharmaceutical events to create awareness to the solution and review the needs and requirements of the end-users in order to adapt them into the prototype. In this line, finding a client to test the solution is of great interest, although the risk is not as high as the previous ones. Also, we must take into account the GSCoP opinion about the difficulties sharing information and in the communication between the main actors in the pharmaceutical sector.

In this SWOT Analysis the strengths and weaknesses of the current Spanish pharmaceutical scenario in relation with the Semantic Nomenclature are depicted. This SWOT analysis has been conducted with respect to the foreseen NeOn technology innovations.

	POSITIVE	NEGATIVE
INTERNAL	<p>Strengths</p> <ul style="list-style-type: none"> • No solution currently offers software that allows semantic interoperability of different pharmaceutical vocabularies. • BOTPlus is being developed and maintained by ATOS. This issue give us a good understanding of this tool, apart from the possibility of future interaction. 	<p>Weaknesses</p> <ul style="list-style-type: none"> • No real client available. • The work to be done is too much for the time of the project. Simplification of the ontologies. • There are other non-semantic tools (BOTPlus, vademecum on-line) that offers reasonable functionality in the Spanish market. • The Nomenclature model has a slowly evolution, low frequency of changes.

EXTERNAL	<p>Opportunities</p> <ul style="list-style-type: none"> • The need of sharing information among stakeholders is clear and is fostered by the government action plans. • There is a clear need of improving the interoperability between drugs vocabulary. • A Semantic Nomenclature could be a very useful tool for the set of actors of this sector. • There is a good quantity of resources and legacy systems about pharmaceutical area, but in a few years are going to be more and more complex and specific in the pharmaceutical and health environment. • No semantic competitors in the Spanish market for this technology. 	<p>Threats</p> <ul style="list-style-type: none"> • Semantic technology is not very well known by the stakeholders. • The complexity of the sector may derive in a difficult implementation in real stakeholders. • There is a governmental initiative for developing an integrated Nomenclature in Spain. It is not based on semantic technologies. • Failure of NeOn in delivering the promising technology • Not good Semantic Nomenclature reference ontology development • Lack of awareness: <ul style="list-style-type: none"> ○ Failure in dissemination and adoption of the NeOn technology. ○ Failure in the dissemination of the Semantic Nomenclature prototypes among the pharmaceutical stakeholders in Spain.
----------	--	--

Table 1: Semantic Nomenclature SWOT Analysis

7 Future Trends

According to Tim Berners-Lee – the inventor of the World Wide Web (WWW) – data integration is the Web’s next leap forward: “Progress towards better data integration will happen using the same basic technology that has made the World Wide Web so successful in the first place: *the link*. Nevertheless, the Semantic Web will derive its power through linking data instead of documents as the current Web does.”

The current global business environment, which is evolving swiftly, brings to the different companies research and innovate new habits managing information in order to be more competitive.

According to these two premises, the use of semantics to classify meaningfully the information and to bridge the gap between the different representations that different stakeholders have is widely accepted. The Semantic Web provides tools and technologies (RDF, OWL, AJAX...) in order to facilitate the integration of distributed and heterogeneous sources.

The current situation in the lifecycle of semantic web applications, technologies and ontologies is anchored and need mature and dive into a new step more robust, where new methodologies and tools are supported and ready to be used by the community for the new challenges as integration tools. In this aspect, the NeOn project takes advantage and provides support in the lifecycle of the new generation of semantic web applications.

Nowadays, each company has their own legacy systems, databases, formats... and just as is described before, companies should integrate and share their knowledge with others in order to exploit their data. In order to facilitate these tasks, the Semantic Web community should provide mechanisms to structure some relevant unstructured information of companies (emails, reports, HTML...) or semi-structured (databases) that the companies manage in their work. Also, it will be useful provide to the end-users the possibility of reasoning over the aggregated information.

Other steps of the Semantic Web technology is provide augmentation of the information on the web. With this feature, semantics provides more inference and automation on the decisions or steps done by computers inside business workflows. This will provide reduce the off-line human decisions and lead to faster overall completion of workflow. This results in saving money and greater productivity, and this is why "semantics" is important to business [24].

The solution presented in this chapter for the Spanish pharmaceutical sector, people from industry observes that Networked Ontologies are useful for dealing with heterogeneous and shared knowledge. Networked Ontologies adds a level of intelligence into the knowledge of the companies and provides to the end-users a new way of sharing and mapping knowledge with people in related activities.

Market predictions for knowledge technologies from companies like Gartner are based on market pull, and assume the existence of mature technology. Nowadays, is depicted that the state of Semantic Web applications are not mature and at the early adopter stage focused on the knowledge management area. The challenge faced in the coming years is to "cross the chasm" to the early majority of the mainstream market, and extend the area to more business to business situations.

8 Conclusions

In this chapter is introduced the Semantic Nomenclature case study of the NeOn project as a solution to cover the profound lack of systematization for creating, maintaining and updating drug-related information in the Spanish pharmaceutical sector. The main contributions of the case study could be identified in different sides. The pharmaceutical scenario is a good test bed for the networked ontologies and their benefits against one large ontology.

In other hand, the main contribution of the Semantic Nomenclature is the development of a set of ontologies about the main drug information repositories and resources in the pharmaceutical Spanish sector, networked between them and also provides a reference ontology. Currently, there is no common model or significant development on ontologies of drugs in the Pharmaceutical sector in Spain. This reference model for the Semantic Nomenclature complements typical pharmaceutical compendium characteristics by giving flexible, extensible and reliable information about drugs to the users of the Pharmaceutical domain.

Semantic Nomenclature and NeOn technologies will contribute in several ways to solving the mentioned business problems in the Spanish pharmaceutical scenario. Through networked ontologies we hope to facilitate the consensus forming process required by such a large-scale initiative. With the technology provided, only partial consensus needs to be achieved; the rest would be automatically managed by the ontological infrastructure. Second, using automatic annotation and mapping techniques, we significantly reduce the cost of joining the initiative. We aim at zero cost for inclusion: any organization may join and maintain its actual information interchange format, while the technology takes care of mapping its internal format to the agreed-upon ontology network. And also, we will provide more added-value services as a news service and a better exploitation of the information stored in a distributed way in several repositories.

The main benefits expected when these large ontologies are created and networked are made their knowledge more robust and redundant, reasoning in order to obtain new aggregated knowledge, provide more functionalities and more faster information to the customer.

The direction of the future work is to continue to develop the networked ontology model for the pharmaceutical scenario in Spain. A middleware to interact with the services provided by the NeOn toolkit will be developed, and services to the pharmaceutical sector will be delivered on top of it.

References

[1] www.neon-project.org

[2] Wikipedia, the free encyclopedia. Anatomical Therapeutic Chemical Classification System http://en.wikipedia.org/wiki/Anatomical_Therapeutic_Chemical_Classification_System

- [3] College of American Pathologists, 2002-2007. SNOMED CT User Guide – January 2007 Release. Retrieved August 7, 2007, from http://www.ihtsdo.org/fileadmin/user_upload/Docs_01/Technical_Docs/snomed_ct_user_guide.pdf
- [4] MeSH factsheet. Retrieved August 7, 2007 from <http://www.nlm.nih.gov/pubs/factsheets/mesh.html>
- [5] UMLS factsheet. Retrieved August 7, 2007 from <http://www.nlm.nih.gov/pubs/factsheets/umls.html>
- [6] Gómez-Pérez J.M, Pariente T, Daviaud C, Herrero G., 2006. Analysis of the pharma domain and requirements. NeOn Deliverable D8.1.1.
- [7] W3C and the Medical Sector. Retrieved August 7, 2007 from http://www.w3c.rl.ac.uk/QH/WP5/handouts/health_w3c.html
- [8] Studer, R., Benjamins R., Fensel D., 1998. Knowledge engineering: principles and methods. Data Knowledge Engineering, vol 25, pag 161-197. Retrieved August 7, 2007 from <http://citeseer.ist.psu.edu/225099.html>
- [9] Jim Jacobs Alexander Linden , 2002. Semantic Web Technologies Take Middleware to Next Level Retrieved August 7, 2007 from http://www.gartner.com/DisplayDocument?doc_cd=109295
- [10] Haase P, Rudolph S, Wang Y, Brockmans S, 2006. Networked Ontology Model. NeOn Deliverable D1.1.1
- [11] Sabou M. et al, 2006. NeOn Requirements and Vision Deliverable
- [12] Gomez-Perez A, Suarez M.C, to be appeared in 2008. NeOn methodology for building contextualized ontology networks. NeOn Deliverable D5.4.1
- [13] Fernandez M, Gomez-Perez A, Juristo N, 1997. METHONTOLOGY: From Ontological Art to Ontological Engineering. Workshop on Ontological Engineering. Spring Symposium Series. AAAI97 Stanford
- [14] Waterfeld W, Weiten M. 2007. Design of NeOn architecture. NeOn Deliverable D6.2.1
- [15] Fortuna B., Grobelnik M., Mladenic D. Semi-automatic Data-driven Ontology Construction System Proceedings of the 9th International multi-conference Information Society IS-2006, Ljubljana, Slovenia
- [16] Euzenat J, 2007. Matching ontologies for context. NeOn deliverable D3.3.1
- [17] Barrasa J., Corcho O., Gomez-Perez A. R2O, an extensible and semantically database-to-ontology mapping language.
- [18] Domingue J., Dzbor M., Motta E., 2004. Magpie: Supporting Browsing and Navigation on the Semantic Web.
- [19] Kenter T., Maynard D, 2005. Using Gate as an Annotation tool.
- [20] Cimiano P., Haase P., Heizmann J., Mantel M. ORAKEL: A portable Natural Language Interface to Knowledge Bases, March 2007
- [21] KAON 2 – Ontology Management for the Semantic Web <http://kaon2.semanticweb.org/>
- [22] Ministerio de Sanidad, 2004. Plan estratégico de Política Farmacéutica para el Sistema Nacional de Salud Español. Retrieved August 7, 2007 from <http://www.anisalud.com/ficheros/Plan.pdf>
- [23] Berners-Lee T. *Welcome to the Semantic Web*. The Economist, num. 21, The World in 2007, page 146.
- [24] Roth B, 2007. Semantic Web the next big thing? Retrieved August 7, 2007 from <http://www.web2journal.com/read/407403.htm>

Additional Reading

Authors Biographicity

Tomás Pariente Lobo has a degree of Bachelor in Telecommunication Engineering by the UPM (Spain). His technical expertise is mainly in semantic technologies, e-Government, Business Process Mangement (BPM) and knowledge management. He is expert in Java J2EE. This expertise started in 1987 when he joined the company Indra Sistemas, where he has taken part in multiple commercial and R&D projects. Since June 2006 he works as project manager and technical consultant for EU-based projects in semantic technologies in ATOS Origin. He is involved in several working groups in this technology. He worked on EU projects such as Ontologging, SmartGov, OntoGov, INFRAWEBBS and is currently involved in the NeOn, LUISA and SOA4ALL projects.

Germán Herrero Cárcel was born in 1980 in Valencia (Spain). He has a degree in Computer & Software Engineering from the University of Valencia. His technical expertise is in Semantic technologies, with strong knowledge in ontology development. He is expert in Java programming. He worked as assistant Scholar in the Robotics Institute from the University of Valencia involved in Semantic Web projects. Then, he worked as a software programmer in the computer department of a financial entity. Currently he is working for Atos Origin in the Semantic & Service Engineering Unit of the ARI department. He is involved on two of the most relevant Semantic Web projects financed by European Union, called NeOn and TAO.