SECTOR DIALOGUES
EUROPEAN UNION BRAZIL

ONTOLOGICAL ENGINEERING FOR PUBLIC AND PRIVATE PROCUREMENT

ONTOLOGICAL ENGINEERING FOR PUBLIC AND PRIVATE PROCUREMENT

MINISTRY OF GOVERNMENT SECRETARIAT

MINISTER OF STATE
GEDDEL VIEIRA LIMA
SECRETARY OF MICRO AND SMALL ENTERPRISES
JOSÉ RICARDO DA VEIGA
DIRECTOR OF MARKETS AND INNOVATION
ALEXANDRE MONTEIRO E SILVA
TECHNICAL SUPPORT TEAM
FABIO MEDEIROS DE SOUZA (GENERAL-COORDINATOR OF MARKETS ACCESS)
CARLOS VELOSO DE MELO JR. (FOREIGN TRADE ANALYST)
JANIO MOREIRA DA COSTA (INFORMATION TECHNOLOGY ANALYST)

EUROPEAN UNION DELEGATION TO BRAZIL

HEAD OF THE EUROPEAN UNION DELEGATION
JOÃO GOMES CRAVINHO
MINISTER COUNSELLOR - HEAD OF DEVELOPMENT AND COOPERATION SECTION
THIERRY DUDERMEL
COOPERATION ATTACHÉ – EU-BRAZIL
SECTOR DIALOGUES SUPPORT FACILITY COORDINATOR
ASIER SANTILLAN LUZURIAGA
IMPLEMENTING CONSORTIUM
CESO DEVELOPMENT CONSULTANTS/FIIAPP/INA/CEPS

MINISTRY OF PLANNING, DEVELOPMENT AND MANAGEMENT

MINISTER OF STATE
DYOGO OLIVEIRA
SECRETARY OF MANAGEMENT
GLEISSON CARDOSO RUBIN
PROJECT NATIONAL DIRECTOR
MARCELO MENDES BARBOSA

CONTACTS

PROJECT COORDINATION UNIT
EUROPEAN UNION-BRAZIL SECTOR DIALOGUES SUPPORT FACILITY
SECRETARIAT OF PUBLIC MANAGEMENT
MINISTRY OF PLANNING, DEVELOPMENT AND MANAGEMENT
TELEPHONE: + 55 61 2020.4645/4168/4785
DIALOGOS.SETORIAIS@PLANEJAMENTO.GOV.BR
WWW.SECTORDIALOGUES.ORG

ONTOLOGICAL ENGINEERING FOR PUBLIC AND PRIVATE PROCUREMENT

Authors:
Edilson Ferneda
eferneda@pos.ucb.br
Fernando William Cruz
fwcruz@unb.br

Brasília • March • 2016

LIST OF FIGURES AND CHARTS

Summary

Executive Summary

In the last years, the Secretariat for Micro and Small Enterprises (SMPE) has made efforts in several fronts in order to improve the relationships between business partners in Brazil and abroad, particularly those benefiting Brazilian micro and small enterprises. Among these initiatives are the following: *(i)* simplify the access to finance, credit and banking services; *(ii)* simplify the access to technology and innovation; *(iii)* simplify foreign trade procedures; *(iv)* simplify the access to public procurement; *(v)* simplify the access to the certification of products and processes, and; *(vi)* the development and implementation of an industrial catalogue of products. In this context, SMPE is currently developing a Electronic Business Market, which, when fully implemented, will be a "meeting point between the Brazilian company and its business partners", so that the entrepreneur – particularly the one involved with micro and small enterprises – will have simplified and safe access to public services. The Electronic Business Market is part of a number of initiatives aiming at reducing transaction costs of affecting both business to business and business to government relationships.

Marketable products (goods and services), arranged and combined in different ways, are essential supplies and the intermediate or final result of the multitude of existing chains of production, supply chains or value chains. These, have normally being represented by a kind of knowledge organization system known as taxonomy, as in the case of the National Classification of Economic Activities (CNAE) and the Products Lists (PRODLIST). Although these taxonomies are good enough to describe items and relationships inside the particular, and rather limited chain of production to which they were conceived, they are not the best conceptual solution to describe the complex, evolving and dynamic reality of actual business relations, at least not at the level inherent to an effective, user-friendly Electronic Business Market. In fact, such taxonomies are generally easier to develop in terms of the conceptual architecture of information and the computational science required to support them, but, in contrast, they preserve a certain structural rigidity, making them difficult to be used in contexts different of that originally aimed at).

This document proposes a new approach to describe the chains of production of relevance to micro and small enterprises without discarding the traditional taxonomic approach already adopted by the Brazilian Government at the Federal level. In this case, the presented suggestion involves an approach which provides the required flexibility to meet the following requirements presented by SMPE:

- Products must be defined on an non-ambiguous ways, according to their essential attributes, but their functionalities must be taken in account, as is the case in the traditional classification schemes such as CNAE, CPV and other taxonomies;
- It must be possible to insert a product into different positions of any chain of production, supply chain or value chain;
- The new classification scheme must allow a detailed and complete understanding of the composition of a product (i.e., if a certain product is supply in the production of another product – as in the case of grapes relatively to wine, or if the final product is integrated by the juxtaposition of other products – as in the case of an automobile;
- The new classification scheme must be compatible with taxonomic classification currently in use in terms of product characteristics and attributes;
- The classification scheme must allow the maximum interoperability between MPEs [Micro and Small Enterprises] in Brazil and with commercial partners and customers from other countries;
- In other to avoid problems resulting from linguistic differences between business partners, an appropriate terminology shall be used to describe products.

Given the complexity and diversity of the links that can be established between products and producers involved in a multitude of chains of production, the development of a knowledge organization system (KOS) to support the description and characterization (catalogue) of products in a way useful to business is no trivial matter. In this context, ontologies emerge as the most convenient KOS because they properly create a common conceptual basis, which is an essential requisite to achieve a high degree of interoperability between independent catalogue systems.

Thus, considering the question of the most convenient KOS to industrial and business purposes, instead of adopting a strict taxonomic structure as it has been traditionally done, the proposal herein presented is the construction of an ontological conceptual model of products as a supply and/or result of any chain of production, supply chain or value chain, called Essential Concepts Database (– acronym in Portuguese), for the classification and descriptions of goods and services in the Electronic Business Market context. This proposal was made based on information collected in the related bibliography which suggests that the conceptual modeling supported by a specific type of ontology, called foundational ontology, seems to cover most of the presented demands because they promote clearer and outlined description and classifications.

Although it is so far unknown for many people, this approach has been adopted in several fields of the chemical, manufacturing and natural resources production industries because it represents with more with accuracy the relationships required for the description of each phase of or process in a chain of production such as those related to a product which is part (*part-of)* or supply of another product, in addition to allowing the conception of a product under different viewpoints allowing to insert it into different positions of the production chain.

BCE proposal contemplates the product configuration or composition and is able to absorb other relevant definitions for the electronic commerce, such as the offer of products and other functionalities increasing the interoperability between commercial partners and incorporates the following characteristics: *(i)* a product can be linked to more than one segment; *(ii)* a product can be the raw material of another product; *(iii)* a product can be composed by other products; *(iv)* a product can be named in different ways; *(v)* a product must have the traditional coding (CNAE, CPV etc) linked to it.

In order to meet the needs and requirements of the electronic commerce throughout a multitude of industries, the conceptual modeling supported by ontologies needs to be done considering, among other things, the product creation dynamism, the variability present in the chain of production, supply chain or value chain and the endless range of details about each of them. For this reason, the following needs are identified:

- A representation language appropriate for the development of this conceptual model. For this purpose, the use an ontological level language is recommended;
- An expert team definition with domain on philosophical concepts and on the category systems used in foundational ontologies in order to guarantee the growth and adaptation of such specification in order to make it the most complete one in the shortest possible time;
- A work methodology to reach a consistent conceptual model even in face of the large amount of products to be specified. In order to reduce the time to conceive the initial models, it is suggested to verify the experiences of other research groups which have succeeded in creating ontologies from the conversion process of one or more existing taxonomies. The results obtained by these groups can be sketches and delineations intended to help with the decision process of the experts who will be conceiving BCE.
- In addition, it is suggested that the products modeling is conducted in steps by choosing the production chain portions to be served for the creation of a first version of the model. From this first version, then, through successive refinements it is possible to enrich the elements allowing the detailed configuration of products and services. With this respect, an intensive work is foreseen for the conceptual modeling group to conduct activities such as:
- Extension of the basic ontology of classifiers, configuration and offer of goods and services in order to contemplate the other typical concepts of the production chain, supply chain or value chain which remained out of this pilot project.
- Construction of new domain ontologies based on conceptual modeling (or exploitation of any existing ontology) in order to contemplate functionalities which were not conducted in the pilot project. A typical example is the linguistic ontologies issue for use in the goods and services annotation schemes (tag attribution).
- Maintenance of the existing functionalities and development of new functionalities in order to contemplate new items included in the conceptual modeling of the project.

Under the point of view of the development of application which makes use of BCE, especially those directed towards Semantic Web, it is important to consider that the ontologists team is in line with the Web environment developers for product description and offer. This strategy is intended to guarantee that a work and improvement detection culture is formed so that the project expansion can follow without major problems during the subsequent phases.

It is still worthy emphasizing the need of a wide disclosure of the project, once reaching the Electronic Business Market depends on the engagement of a number of related sectors of the society. SMPE is also responsible for paying attention to the need to provide a structure in the form or curatorship for the continuous evaluation and adjustments of the system to be offered. In other words, care is required in the conduction of the work because the mere construction of BCE (and its duly validation) without the involvement and awareness of the peers (managers and developers who deal with the electronic commerce in Brazil) affects the interoperability requirement intended for the markets.

Introduction

The main mission of the Public Administration is to provide help to the State in its essential task of meeting the society's demands. The evolution of the Information and Communication Technologies (ICT) and the globalization process lead the Public Administration to adjust itself to these continuous changes, demanding attention by the public power for the State efficiency combined with a growing offer of public services, with higher quality, social control and guarantee of the individual rights (Ferneda et al., 2011). In this context, the electronic government emerges as a catalyst or unprecedented interactions between the citizens and public and private organizations. The Internet accessibility, low cost and rapid capacity to disseminate a large volume of information, associated to its interactivity, decentralization and flexibility capacities, provided, since the beginning of this millennium, the emergence of social organization opportunities under three essential principles (Smith and Smythe, 2004, Brousseau, 2003, Alonso et al., 2011): self-organization; democracy and customization.

In 2005, the United Nations evaluated that ICTs revolution had the potential to transform the government as a whole, affecting its work in four main areas (United Nations, 2005): *(i) processes and internal relationships*, both through the automation of routine tasks, especially the records maintenance data recovery, and by enabling increased communication and collaboration levels between the member of an organization, regardless of the physical location (virtual network organization); *(ii) relationships between the government and the users,* via the electronic service provision; *(iii) relationships between the government and the citizens,* through the feasibility of a number of digital democracy forms, including virtual communities which facilitate the exchange of information and political mobilization; and *(iv) relationships between the government and companies,* by means of electronic business opportunities in areas such as the purchase of goods and services. Ten years later, ICTs presented a number of advances, particularly with the consolidation of the so called Web 2.0, and its assimilation by society, with the dissemination of the social networks, particularly by the government, through electronic government initiatives. Thus, to this relationship, the *intermediation between the society's sectors* could be added, and, particularly, between the companies.

In line with this new reality, the Secretary for Micro and Small Enterprises (SMPE), of the Brazilian Ministry of Industry and Commerce, has made efforts in several fronts in order to improve the relationships between trade partners in Brazil and abroad. Among these initiatives are the following: *(i)* facilitation for addressing credit applications to the financial institutions, *(ii)* the conduction of investment in innovation solutions, *(iii)* facilitation of the foreign trade operations, *(iv)* the viability of sales channels to the Government itself, *(v)* the certification of products and processes and, particularly, *(vi)* the products listing, focus of this document. In this context, a project is ongoing, called Electronic Business Market[1], which proposes to be a "meeting point between the Brazilian company and its business partners", so that the entrepreneur – particularly the one involved with micro and small enterprises – has simplified and safe access to public services and with focus in the feasibility of a context to increase the commercial operations in and out of the Internet. With this respect, SMPE has followed the discussions of the international community involved with product classification and listing systems, particularly those promoted by the European Committee for Standardization, which deal with standards for product identification, business processes and product classification, among others related to the electronic commerce.

Concerning the product classification, some consolidated taxonomies which have become a reference at international level are of note, such as the European standards *Classification of Products by Activity* (CPA), *Central Product Classification* (CPC), *Common Procurement Vocabulary* (CPV), eCl@ss and *Global Product Classification* (GPC). At national level, SMPE also considers the systems currently being used by the Government, such as the National classification of Economic Activities (CNAE) and the Products Lists (PRODLIST).

This report shows a discussion and presents suggestions towards the viability not only of the classification of products in the context of the Brazilian production chains, but their description and offer in the context of the Electronic Business Market, considering the compliance with the following requirements:

- The products must be defined on an non-ambiguous ways, according to their essential attributes, but their functionalities must not be disconsidered, as in the traditional classification schemes such as CNAE, CPV and other taxonomies;

- The products can be inserted into different positions of the production chain, supply chain or value chain;

- The classification scheme must allow the explicitness of the composition lists

1. https://www.empresasimples.gov.br/informacoes-sobre-a-praca-eletronica-de-negocios.

(i.e., a product is a raw material of another product, or a product is formed by the juxtaposition of other products);

- The new classification scheme must not prevent the product from being visualized according to the current traditional taxonomies;
- The classification scheme must allow the maximum interoperability between MPEs in Brazil and with commercial partners and customers from other countries;
- Products description is supported by appropriate terminology in order to solve the linguistic differences between the parties involved in the negotiation.

This document presents a review proposal of the Brazilian production chains representation design in order to meet the micro and small entrepreneurs' demands, however, without disconsidering the traditional taxonomic forms already adopted by the Government. In addition, the following are presented: *(i)* a methodology for the construction of a knowledge organization system on goods and services and *(ii)* an alternative for the incorporation of the proposed design associated to the offer of products in the Electronic Business Market's Web portal.

The document was divided into two parts, being the first related to a theoretical reference so as to provide the reader with subsides to understand the proposal, and a second part on which considerations are made and presentation of an approach suggestion for the product classification problem of the Brazilian production chains.

Part I – Theoretical Reference

The classification of the production chain products can be conducted by using knowledge organization systems and, at the same time, consider the use of the Semantic Web principles. Below, the theoretical and conceptual elements supporting the proposal described in Part II of this report are presented. For this purpose, this part presents a general and introduction view of some practicable alternatives of knowledge organization systems, namely taxonomies, thesauri and ontologies. Subsequently, the conceptual modeling definition as a privileged approach for the conception of a knowledge organization system directed to the product listing is presented. Finally, it presents a history on the electronic commerce based on the evolution of the traditional Web to the Semantic Web.

1. Alternatives for knowledge organization

Products (goods and services) can be seen as concepts which, for being understood, need to be classified or listed (Best, 1995; Sternberg, 2000), and contextualized by the use of qualifiers or attributes. However, it is noted that there may exist different forms of structural relationship between the attributes of a product and that the understanding of what a product is depends on the structure of its attributes. Such structures can be understood as a Knowledge Organization System. Therefore, Knowledge Organization Systems refer to systems directed towards the structuring of knowledge units, or concepts, and can be seen as tools intended to model reality aspects of the understanding and communication effects. Some options of such systems are shown in Figure 1.1.

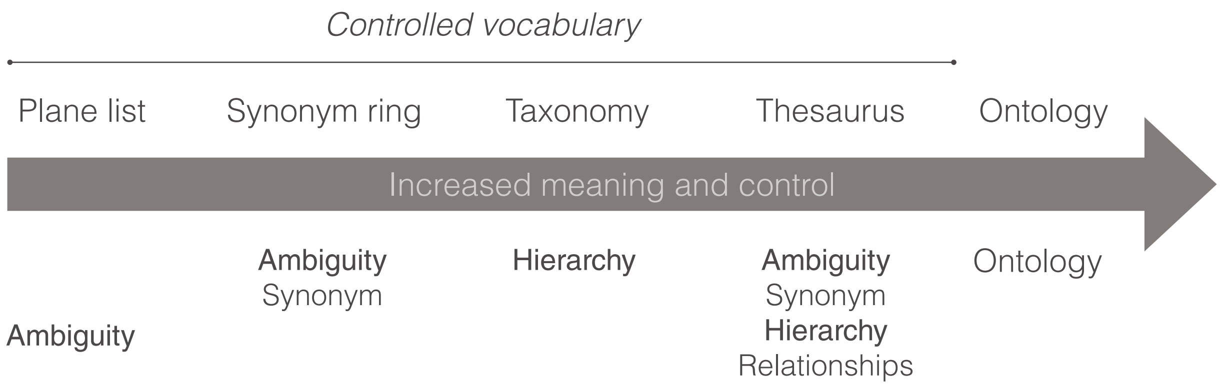


Figure 1.1: Some of the main knowledge organization systems
Source: Adapted from Zeng and Salaba (2005)

In the left part of this figure, the plane lists refer to attributes which, in practice, are terms which do not have any structural relationship between each other and the synonym rings, which establish that the attributes can be synonymous between each other (the same meaning). A little more towards the right of this figure are the taxonomies, the thesauri and the ontologies, which promote an increasingly complex relationship between the attributes, according to the order on which they are presented.

Another finding concerning the knowledge organization systems refers to the qualifier's description form. If the qualifiers are described in a natural language (text in Portuguese language, for example) the implications (acquisition of knew knowledge) can be naturally performed by human beings. On the other hand, if the description of such qualifiers is conducted on a formal notation (the formal logics, for example), the implications (knowledge acquisition) can be performed by machines. Here, three of these systems must be highlighted: taxonomies, thesauri and ontologies which are described below:

1.1 Taxonomies

A taxonomy is essential a classification tool "through similarities and differences between the object characteristics in a certain domain, on which the objects and phenomena are divided into classes, in their turn subdivided into subclasses, theses into sub-subclasses and so on" (Carlan and Medeiros, 2011), as in Porphyrian Tree presented in Figure 1.2 (Sowa, 2000). In it, each element is composed by successive dichotomies, proceeding from the general to the particular, from the highest extension to the highest understanding, from the genus to the *infima species,* i.e., until reaching an individual concept.

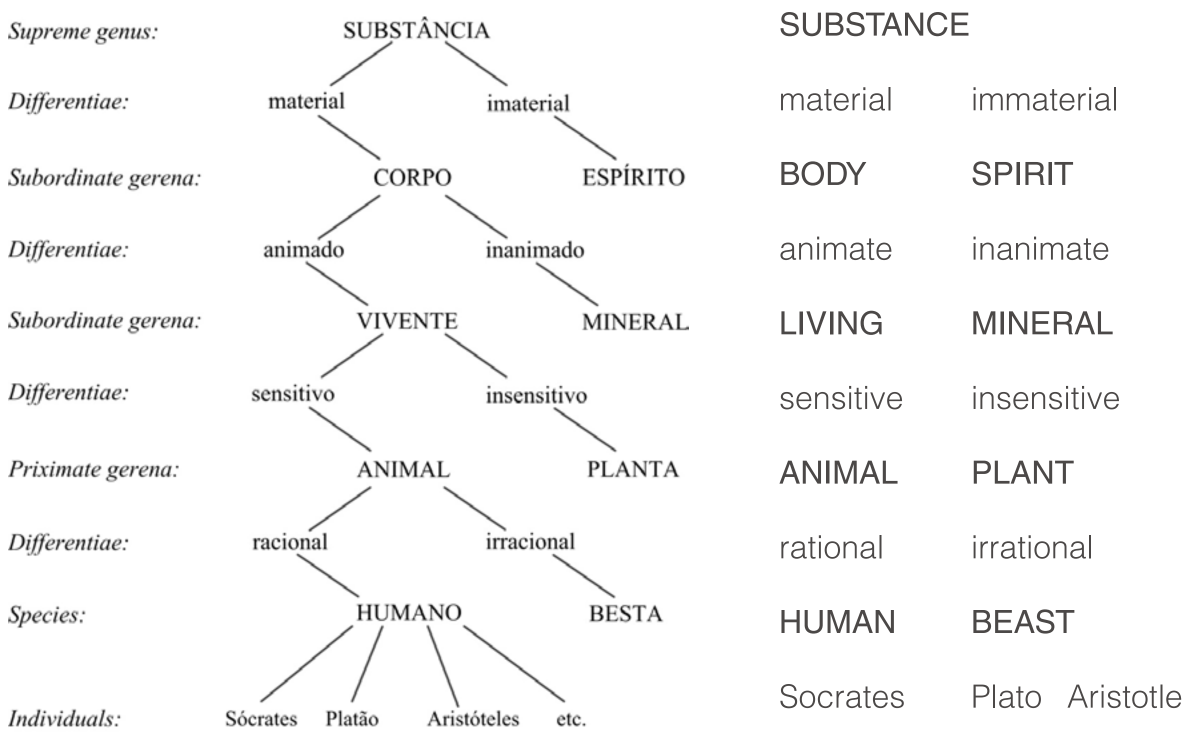


Figure 1.2: Porphyrian Tree
Source: Adapted from Peter of Spain (1239) by Sowa (2000)

With the advent of Informatics, the idea of the specializations tree built-in in Porphyrian tree has been transformed an used to describe objects, aiming at recovering information, support to browsing and organization of Web pages contents. An example of taxonomy is shown in Figure 1.3, on which concepts of a certain level as sons and, therefore, a specialization of the parent concept, of higher level.

Ranganathan (1967) noticed some application limitation of the Porphyrian tree partitions and, as a result, he proposed a non-dichotomous variable know as Banyan tree, shown in Figure 1.4, in order to represent the knowledge organization in the Library Science. In this figure, the branches connected to the ground establish a metaphor to indicate that the concepts may and must be related to different branches, without the exclusivity relationship typical from the Porphyrian tree's conceptual relationships.

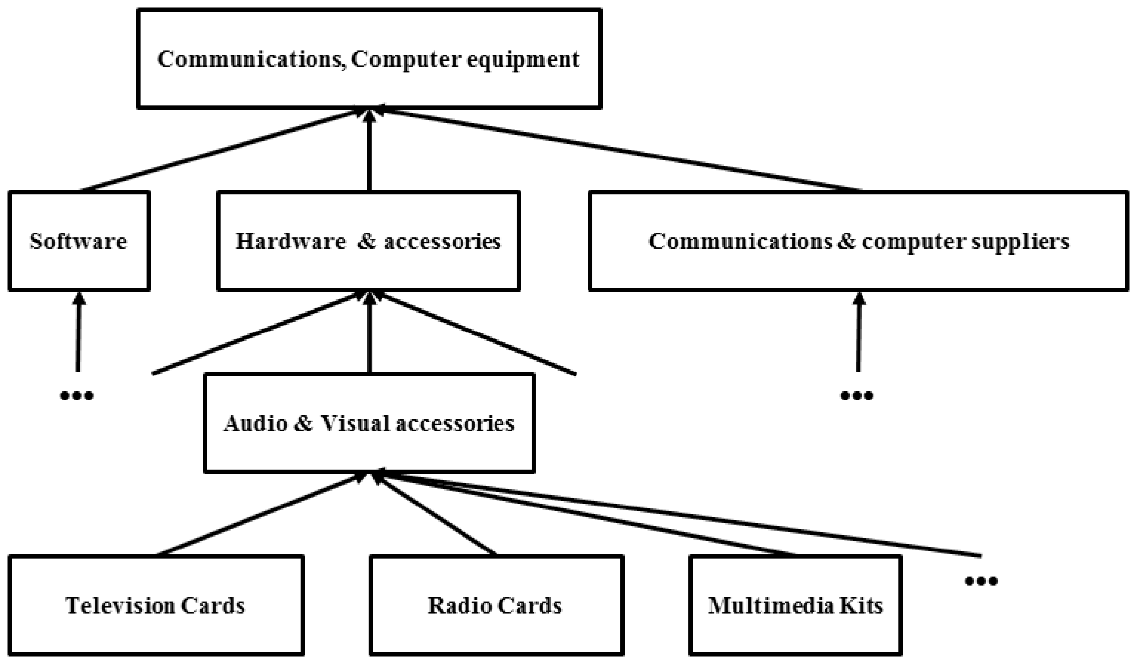


Figure 1.3: Example of taxonomy
Source: Adapted from Gómez-Pérez et al. (2004)

Figure 1.4: Banyan tree
Source: http://www.merriam-webster.com/art/dict/banyan.htm

The taxonomies have currently been used as knowledge organization systems not only in their dichotomous version, but also in the version proposed by Ranganathan. With this respect, they are seen as a special type of ontology (*lightweight ontology*) which will be discussed later on, once they have only specialization relationships between the concepts. In addition, the taxonomies are not static and are normally adapted to include new contents and knowledge related to the defined interest domain.

Due to their characteristics, the taxonomies have been used in several fields of knowledge and, especially, by the governments for the classification of companies, production areas, collection sectors and other applications. Specifically in the context of this work, the taxonomies deserve special attention, given the similarity between the production chain designs which normally arrange themselves in a hierarchical tree. For this reason and due to the ease of use, the taxonomies can be considered as one of the most used systems in the product classification.

1.2 Thesauri

A thesaurus is a list of terms indicating their classification according to the ideas they represent. It reflects the logical structure of an area of knowledge, whether specific or general. It can be seen as a strongly constituted semantic structure based on the lexical units which compose its vocabulary.

To all and any field of knowledge corresponds a set of concepts proper to it and it is the conceptual system of an area which constitutes a basic parameter for supporting what is called thesaurus. The concepts are represented by terms with respect to their descriptors. The structure of the concepts and its several relationships must consider the context on which the term appears. (Lima et al., 2015)

A thesaurus is a controlled vocabulary which incorporates semantics by means of the relationships between the terms.[2] These relationships can be classified into hierarchical, equivalence and associative. The hierarchical relationships express the positioning of a concept with respect to wider or more specific terms. Some relationship examples:

2. The relationship types can be, for example (IBICT, 1993): NE - Scope Note; TA = Associated Term; TE - Specific Term; TEG - Specific Generic Term; TEP - Specific Partitive Term; TG - General Term; TGG - General Generic Term; TGP - General Partitive Term; TR - Related Term; TT - Top Term / Broader Term; UP - Used For; USE - Use.

- G*enus/Species:* between a class or category and its members of species
  Birds
  TEG Parrots
  Parrots
  TGG Birds

- *Whole/part:* between a part of the system and the system
  Cardiovascular system
  TEP Heart
  Heart
  TGP Cardiovascular system

- *Of Instance:* between a general concept, class of things or events, and an individual class instance, which is frequently represented by a proper name
  Mountainous regions
  TEI Alps
  ...
  Alps
  TGI Mountainous regions

The equivalence relationships provide substitutive indexation and search terms. It occurs between preferred and non-preferred terms. Example:

Theory of equations (preferred term)
UP Algebraic equations (non-preferred terms)
Where,
Algebraic equations USE Theory of equations

In their turn, the associative relationships are those occurring due to the strong association between the concepts. They occur between concepts which are not hierarchically related, but semantically related. Some examples of associative relationships are represented below:

Boats
TR Vessels
...
Vessels
TR Boats

Other examples of associative relationships are: Subject / Study object (forestry / forests), Process / Agent (temperature control / thermostat), Action / Action product (weaving / clothes), Action / Target (harvest / cultivation), Material / Properties (poison / toxicity), Artifact / Parts (optical instruments / lenses), Cause / Effect (death / loss), Object / Counter-agent (plant / herbicide), Concept / Measurement unit (electric current / ampere), Compound name / Origin name (reptile fossil / reptile), Organisms or substance / its derivatives (mules / donkeys).

Another aspect of the thesauri is related to the coverage degree of the terms. The scope of the term is restrict to the meaning within the controlled vocabulary domain. On the other hand, the homograph terms require qualifiers to clarify their sense. For example, *(i)* a compound term can be used instead of a simple term with a qualifier if it occurs in natural language: religious tolerance instead of Religion (Tolerance), *(ii)* a qualifier must be added to each homograph: Mango (Fruit) and Sleeve (Clothing) or *(iii)* when a term is used a multidisciplinary search system: Abortion (Medicine) Abortion (Crime) and *(iv)* for terms with ambiguous sense: Mercury (Planet) and Mercury (Metal).

International standards, such as ISO 25964-1:2011[3] and ANSI/NISO Z39.19-2005 (R2010) (ANSI/NISO, 2010), address on construction, formatting and maintenance of mono-and multilingual thesauri. With these, it is possible to structure thesauri in their theoretical bases (concept, term, categories and aspect) and technical-operational (planning, term collection, terminological control, establishment of relationships between concepts and disclosure and publication means).

Finally, it is important to note that the maintenance of thesauri is influenced by the evolution of the adopted description language. In other words, the inclusion of new terms in the dictionary of the language adopted in the thesaurus necessarily requires changes to the already ready descriptions. More details about standards, tools and forms of use of thesauri can be found in Morville and Rosenfeld (2007).

3. http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=53657

1.3 Ontologies

According to Robredo (2005, p. 321), an ontology is "the result of the formulation, as strict and complete as possible, of a conceptual scheme on a domain". For Noy and McGuinness (2001), the importance of the ontologies is justified for: *(i)* allowing to share the common knowledge in information structures between distinct communities or between *software* agents; *(ii)* allowing the reuse of the knowledge; *(iii)* being able to perform inference in a knowledge domain. *(iv)* separating the domain knowledge from the operational knowledge; and *(v)* performing the analysis of the structure knowledge having as a result answers that are more relevant than the traditional methods and techniques.

In Computer Science, ontology concept has been discussed for more than two decades (Neches et al., 1991; Guarino and Giaretta, 1995; Albertazzi, 1996; Uschold and Gruninger, 1996; Guarino, 1997; Guarino, 1998; Embley et al., 1998; Chandrasekaran et al., 1999; Wache et al., 2001; Guizzardi and Wagner, 2005; Guizzardi, 2005; Guizzardi, 2007). These discussions have been promoting the development of a number of works related to methodologies, tools, languages and applications. In this context, ontology is used as an artifact to represent certain aspects of reality in the form of concepts and in the relationships between such concepts. In addition, a full ontology also contemplates a set of axioms which help to reduce ambiguities in the description of this reality.

Guarino (1998) classifies ontologies into *(i) foundational ontologies,* which describe more general concepts for having a category system which does not depend on a specific domain, *(ii) domain ontologies,* which describe a vocabulary related to a specific domain, *(iii) task ontologies,* which describe an activity or device from components, processes or functions, and *(iv) application ontologies,* which describe concepts which depend both on a specific domain and a specific task, and are generally a specialization of both. Below are more details about the two first ones.

1.3.1 DOMAIN ONTOLOGIES

The literature on knowledge organization systems highlight the importance of domain ontologies. These, on their turn, are reusable vocabularies of the concepts of a domain and its relationships, of the activities occurring in such domain, and elementary theories and principles governing such domain. They share many structural similarities, regardless of the language on which they are expressed. Most of these ontologies describe:

- *Individuals:* instances or objects. Individuals in a given ontology may include concrete objects as well as abstract individuals. An ontology does not need to include individuals, but one of the main purposes of an ontology is to classify individuals, even if these are not an explicit part of ontology.
- *Classes*: sets, collections, concepts, programming classes, types of objects, or types of things.
- *Attributes:* aspects, properties, resources, characteristics or parameters that the objects (and classes) may have.
- *Relationships:* forms on which classes and individuals relate between each other.

The example in Figure 1.5 shows properties (binary relations) between the individuals. The property *hasBrother* connects the individual *Mateus* to individual *Gilda* and *hasBrother* connects *Pedro* to *Mateus* and *Gilda*. The classes Person, Pets and City.

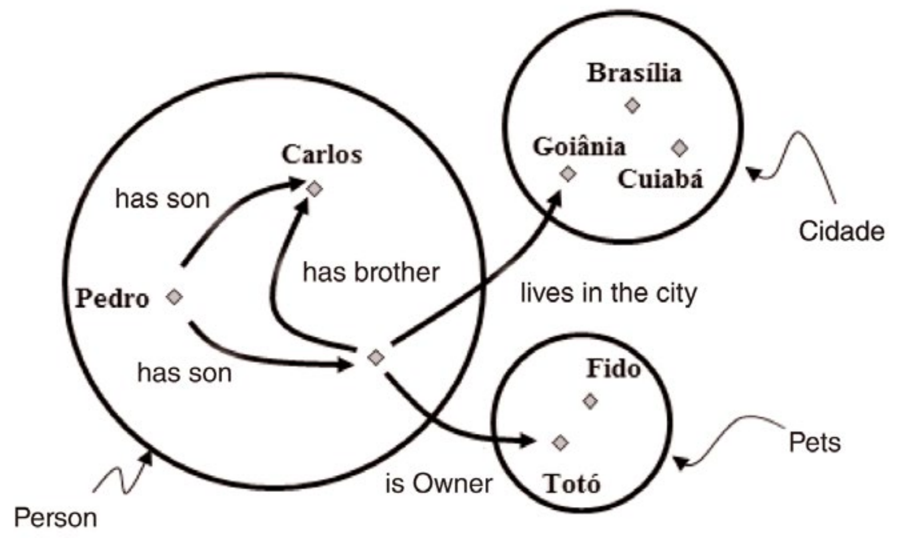


Figure 1.5: Example of class representation containing individuals
Source: Adapted from Horridge (2011)

The properties may also be inverted[4]. The properties can be limited to a single value (functional properties) such as *hasID*. They cans also be transitive, such as *hasAncestor* (if *hasAncestor* b and b *hasAncestor* c, then *hasAncestor* c) or symmetrical, such as *hasBrother* (if *hasBrother* b, then b *hasBrother* a). Other components which may compose a domain ontology are (Pinto et al., 1999):

- **Function terms:** complex structures formed from certain relationships which may be used instead of an individual term in a declaration.
- **Restrictions:** descriptions formally declared of what must be true for some declaration to be accept as input.
- **Rules:** declarations in the form of a if-then clause (precedent-consequent) which describe the logical inferences which can be extracted from a declaration in a particular form.
- **Axioms:** assertive (including the rules) in a logical form, which, together, compose the general theory described by ontology in its application domain. Here, "axioms" also include the axiomatic declarations-derived theory.
- **Events:** a change to the attributes or relationships.
- **Vocabulary:** words or group of words which can be found in the application domain. It may include, among others: *(i)* controlled vocabulary schemes (the use of pre-defined authorized terms, which were pre-selected by the designer of the vocabulary of by the user community, in contrast with natural language vocabularies, where there is no vocabulary restriction, is mandatory); *(ii)* key-words; *(iii)* proximity area (a syntax which indicates where the desired knowledge is in a context according to a specific semantic); *(iv)* dictionary; *(v)* homographs, *(vi)* synonyms and polysemy, *(vii)* morphemes (prefixes, suffixes and radicals). This component reduces the ambiguity inherent to the normal human language, where to the same concept different names can be given and guarantee the coherence.

In the context of the product description and electronic commerce, the taxonomies have been used (Mizoguchi et al., 1995; Heijst et al., 1997) and recognized as a light domain ontology (*lightweight ontology*), with focus on specialization relationships. In this context, there are several standards for the exchange of information between customers and suppliers, and between different suppliers, in order to promote the non-ambiguous identification of goods and services for the global market.

The *United Nations Standard Products and Services Codes* (UNSPSC) is an example of ontology basically composes by taxonomic relationships. It is directed towards the general classification of goods and services from a global standard code which classifies. It coding system is organized in a taxonomy with five levels (Segment, Family, Class, Merchandise and Function), each of them with a two-character numerical value and a description text. In its sixth version, UNSPSC contains 20,000 products organized in 55 segments. Segment 43, for example, is related to computer equipment, peripherals and components, containing around 300 types of products. A fragment of this segment is shown in Figure 1.6.

Another example or electronic commerce ontology is eCl@ass, which contains 4 concept levels (segment, main group, groups and merchandise class), or material class, which contains a numbering system that is similar to that of UNSPSC. It contemplates around 12,000 products organized in 21 segments. Segment 27, for example, is related to electric engineering and contains around 2,000 products. The main group 27-23 contains 400 concepts and is related to process control systems and other computer devices. Figure 1.7 presents a fragment of eCl@ss, which provides a set of attributes for each product which is as a sheet in this classification system. The set of attributes along with the individual characteristics describe a merchandise relationship. For example, a computer has attributes such as *product-type* and *product-name*. The provider of this ontology makes an online software tool which allows the interaction with the system and the obtainment of results in different languages.

4. For example, the inverted property of isOwner and hasSon would, respectively, be isPropertyOf and hasFather.

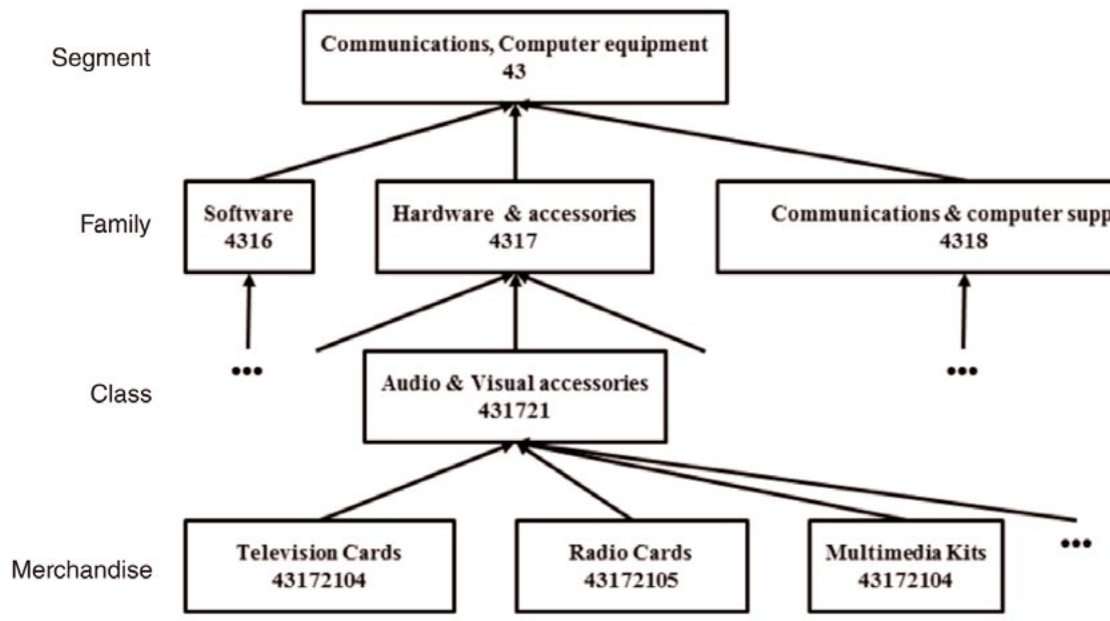


Figure 1.6: Part of UNSPSC classification for computer equipment
Source: Gómez-Pérez et al. (2004)

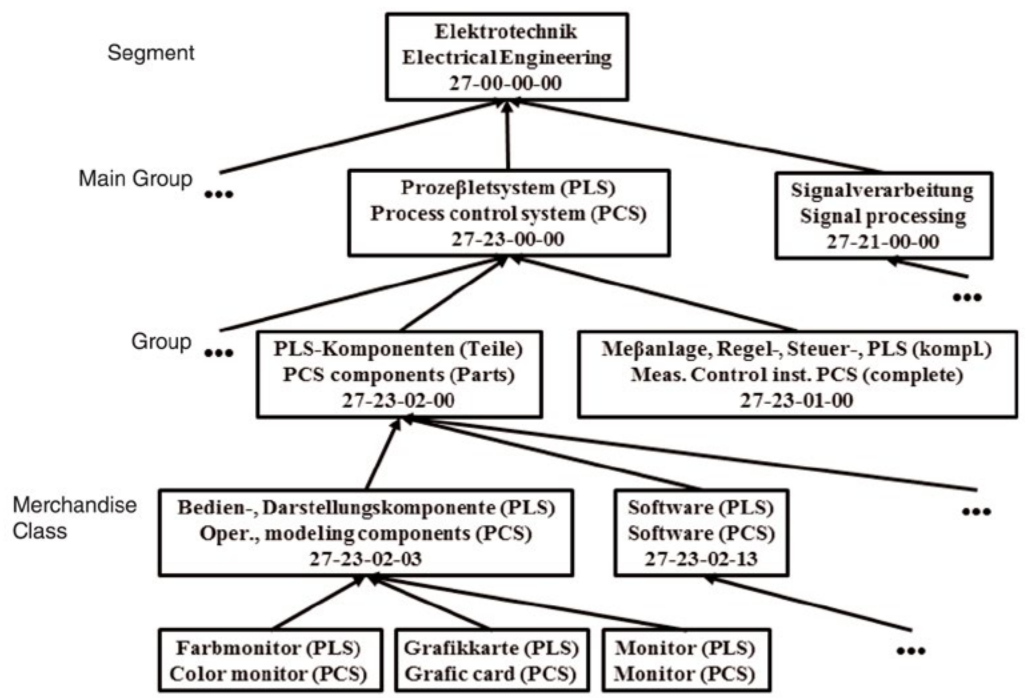


Figure 1.7: Part of eCl@ss classification for electric engineering products (German and English)
Source: Gómez-Pérez et al. (2004)

The *Enterprise Ontology* (Uschold et al., 1998), formed by 92 classes, 68 relationships, 7 functions and 10 individuals, of which one fragment is shown in Figure 1.8, gathers relevant terms and definitions related to companies.

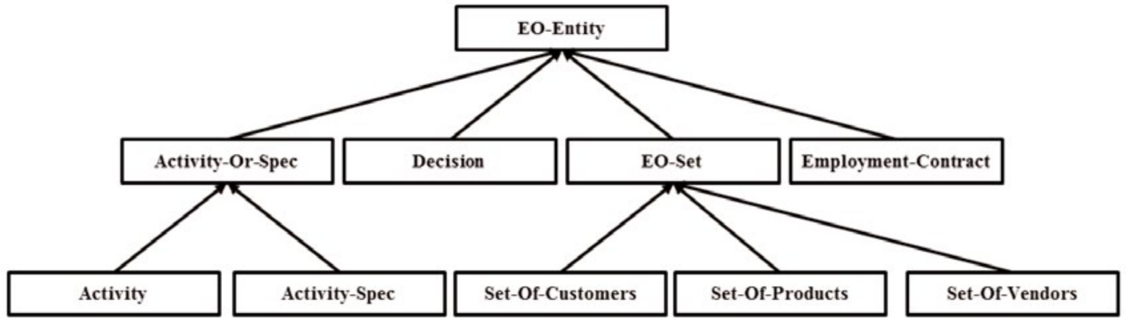


Figure 1.8: Taxonomy extract of Enterprise Ontology
Source: Gómez-Pérez et al. (2004)

Conceptually, *Enterprise Ontology* is divided into four sections:

- *Activities and processes.* The central term is the activity, which is intended to capture the notion of everything that involves actions. The concept of activity is closely connected to the idea of actor, which can be a person, an organizational unit or machine.
- *Organization,* the central concepts of which are: Legal Entity and Organizational Unit.
- *Strategy,* the central concept of which is Purpose, which captures the idea of something that a plan can help to reach, or the idea that an Organizational Unit can be responsible. In fact, this section includes any type of proposal, whether at an organization and time scale level (normally called strategic), or in a detailed and short mandate.
- *Marketing,* the central concept of which is Sale, which is an agreement between two Legal Entities for the exchange of a product by a sales price.

Normally, the products are the goods or services and the sale price is monetary, although there are other possibilities. The legal entities assume the supplier or customer roles. The sale may have been previously agreed, and a future potential sale can be foreseen, even if the real product is or not identified and if it exists or no

1.3.2 FOUNDATIONAL ONTOLOGIES

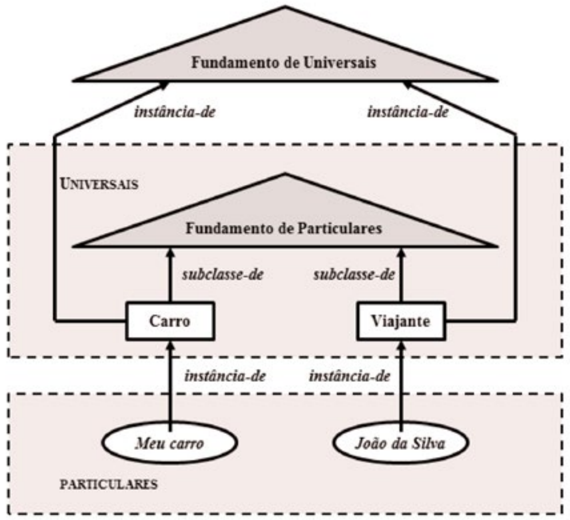
A century ago, philosopher Edmund Husserl defined formal ontology as a specific part of the philosophical Ontology subject (Zamborlini, 2011). Like the Formal Logics, which addresses formal logic structures (truth, validity, consistency, etc.), the formal ontology addresses formal ontological structures (theory of the part, theory of the whole, types and instantiation, identity, dependence, uniqueness, etc.), i.e., formal aspects of objects regardless of their particular nature (Guizzardi, 2005 apud Zamborlini, 2011). This concept has had a number of interpretations in the scope of Computing, being used in areas such as Information Systems, Software Engineering, Artificial Intelligence and Semantic Web.

Foundational ontologies incorporate the formal ontologies in a system of independent domain categories (concepts such as part, whole, role, event, etc.), which can be used for the construction of other types of ontologies, such as the domain (Gómez-Pérez et al., 2004) one. Among the characteristics of foundational ontologies, are:

(i) Universality. It must be “universal”, i.e., each concept imaginable in a specific ontology must be correctly connected to the foundational ontology category systems, regardless how much the domain concept is general or specific. For example, a foundational ontology, which only classifies entities in physical objects and mental objects is not universal, once it does not support concepts such as processes or situations.

(ii)Articulation. It must be “articulated”, i.e., there must be a justification for each foundational ontology concept. On the other hand, there must be sufficient concepts to allow the share of knowledge without ambiguities. For example, a foundational ontology which classifies entities into mortal and immortal could be, ideally, useful for Theology, but not for the other fields.

Guarino and Welty (2000a apud Gómez-Pérez et al., 2004) conceived two foundational ontologies called, namely, Universal Ontology and Particular Ontology, so as to allow the definition of concepts such as those presented in Figure 1.9. In this figure, the concepts and instances inherent to the domain on travels (Car, Traveler, ...) are derived from these two foundational ontologies.



UNIVERSALS FOUNDATION

instance-of *instance-of*

Universals

PARTICULARS FOUNDATION

subclass-of *subclass-of*

CAR TRAVELER

instance-of *instance-of*

MY CAR *JOÃO DA SILVA*

Particulars

Figure 1.9: Relationship between the private and universal ones. Source: Gómez-Pérez et al. (2004)

Each concept of foundational ontology considers the attributes mentioned in several value combinations for these attributes and each concept of a specific domain will be an instance of at least one of the foundational ontology sheets. Concerning the particulars (individuals), the foundational ontology maintains general concepts (for example, object) for which the domain concepts can be connected with the *subclass-of* relationship.

Examples of foundational ontologies are *Cyc's upper ontology*, developed by Lenat and Guha (1990 apud Gómez-Pérez et al., 2004, p. 76-77) in the scope of the Cyc project, the Sowa one (1999 apud Gómez-Pérez et al., 2004, p. 75-76), the *Standard Upper Ontology* – SUO (Pease and Niles, 2002 apud Gómez-Pérez et al., 2004, p. 77-78) and the *Unified Foundational Ontology* – UFO (Guizzardi et al., 2005, 2007, 2010).

The foundational ontology UFO is based on ontological metaproperties and defines, among other things, terms related to structural aspects such as general concepts of objects, its intrinsic and relational properties, the types they instantiate and the roles they play. UFO is divided into three fragments: *(i)* UFO-A (*Ontology of Endurants*), which defines the core of this ontology, systematizing concepts, for example, taxonomic types and structures, whole-part relationships, intrinsic properties and attribute value spaces, roles, relational properties; *(ii)* UFO-B *(Ontology of Perdurants)*, which aims at systematizing concepts such as states, processes, events, time relations, among others, and *(iii)* UFO-C (*Ontology of Social and Intentional Entities*), built on the previous fragments, it aims at systematizing concepts which include the agent, action, intentional states, delegation, commitments and social claims concepts, among others.

For example, from the foundational ontology UFO, Nardi et al. (2013) proposes the UFO-S, a reference ontology for services based on the commitment notion. The UFO-S refines the concepts of a foundational ontology, providing a conception for services that is independent from a particular application domain. Figure 1.10 represents a diagram with the main concepts and relationships involved in a product offer, according to UFO-S. A *service offer event* results in the establishment of a *service offer* between a *service provider* and a *target customers community*. A *service offer* is composed by a *service offer commitment* between the *service provider* with the target-customers community. The concepts used in this diagram (*category*, *rolemixin*, *event*, *collective*, *relator* and *mode*) are described later on, in section 2.2.1 (Chart 2.2).

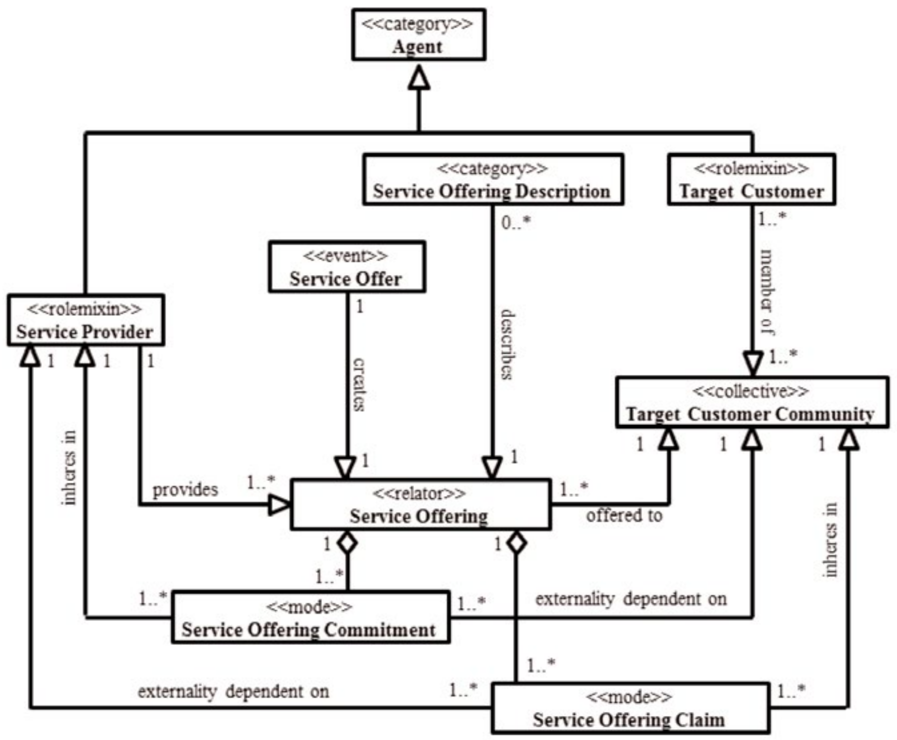


Figure 1.10: UFO-S: main concepts and relationships involved in a product offer
Source: Nardi et al. (2013)

1.4 Comparing knowledge organization systems

Among the presented knowledge organization systems, taxonomies and thesauri are more popular, although ontologies are gaining more and more space to conceptualize knowledge domains.

Concerning the taxonomies, the Porphyrian tree has not shown to be appropriate for some classification systems because the contents to be indexed are not normally part of a single knowledge domain (Campos and Gomes, 2003), on the contrary, they are complex and interrelate with each other.

Currently, the idea of the Banyan tree has been used for the construction of evolved taxonomy forms where each term is associated to as many classes and subclasses as required (polychotomous taxonomies[5]), considering the multidimensional mapping problem of any specialized area. In this case, the resulting knowledge organization system can be much more qualified as an ontology than a traditional taxonomy.

In the context of the product listing, taxonomies have been preferably adopted by the governments, although there is certain structural rigidity, making its application in different contexts than those they were conceived to (Borgo and Lesmo, 2008) more difficult. In addition, the taxonomic standards in use (CPV, eCl@ss, CNAE etc) are not always compatible between each other (Attström et al., 2012).

Concerning thesauri and ontologies, Ferneda (2013) mentions several works addressing the differences and similarities among them (Codina and Pedraza-Jiménez, 2011; Kless and Milton, 2010; Sales and Café, 2009; Jiménez, 2004;). Among the indicated similarities are: *(i)* both aim at representing and sharing the concepts or the vocabulary of a domain in order to allow an efficient communication; *(ii)* its basic structures are hierarchical, grouping terms or concepts into categories and subcategories (classes and subclasses); *(iii)* both can be used to list or organize information resources.

For Qin and Paling (2000), however, ontologies have a higher semantic level due to the class/subclass-type hierarchical relationships and to the "crossed" relationships. Ding and Foo (2002) highlight that ontologies allow communication between humans and computers, while controlled vocabularies, created in the context of Library Science, are tools used to facilitate communication between human beings. Although the relationship of these two instruments with the description or representation of something, its origins and proposals are different (Moreira, 2003 apud Ferneda, 2013). Thesaurus are oriented towards the needs of experts, being "practical instruments to help with the indexation and search for documents", while ontologies emerged from the "need to describe the digital objects and their relationships", an application that is more directed to the automated procedures, and to computer inferences by means of intelligent agents.

5. In the polychotomous taxonomies, an element is associated to as many classes and subclasses as required, within a specialized domain or task (Campos and Gomes, 2007).

1.5 Example of ontology

The domain of academic courses and professorships, used by Antoniou and van Harmelen (2004) in order to exemplify the description of a certain domain, addresses individual objects (professorships and academic courses) and object classes (personal, technical-administrative body, faculty, support personal, assistant professor, deputy professor, full professor). In this context, the hierarchical relationships can be represents as in Figure 1.11. In addition, some properties must be considered. For example, the courses can only be conducted by the faculty members.

Figure 1.11: Example of class hierarchy
Source: Antoniou and van Harmelen (2004)

Antoniou and van Harmelen (2004) state:

> Classes, heritage and properties are, of course, known in other computer fields, for example, in the object-oriented programming. But, while there are many similarities, there are differences too. In the object-oriented programming, an object class defines the properties applicable to it. In order to add new properties to a class it is necessary to change the class.
>
> However, in RDFS[6] properties are globally defined, i.e., they are not encapsulated as class definition attributes. It is perfectly possible to define new properties applicable to an existing class, without changing such class.
>
> On the on hand, this is a powerful mechanism with extended range consequences: we can use classes defined by others, and adapt them to our needs by means of new properties. But, on the other hand, this property engineering deviates from the standard approach which emerged in the object-oriented programming and modeling area.

Thus, while in object-directed scenarios the hierarchical relationships, only three classes are defined, with ontologies, these relationships are also possible between properties. For example, "itisTaughtBy" is a sub-property of "involves". If a course is conducted by a faculty member, then this course involves, among other things, a faculty member. The inverse is not necessarily true. For example, a professor can be the course organizer (full professor), and the other one can be a tutor who follow the teaching-learning process (assistant professor). This is represented in Figure 1.12.

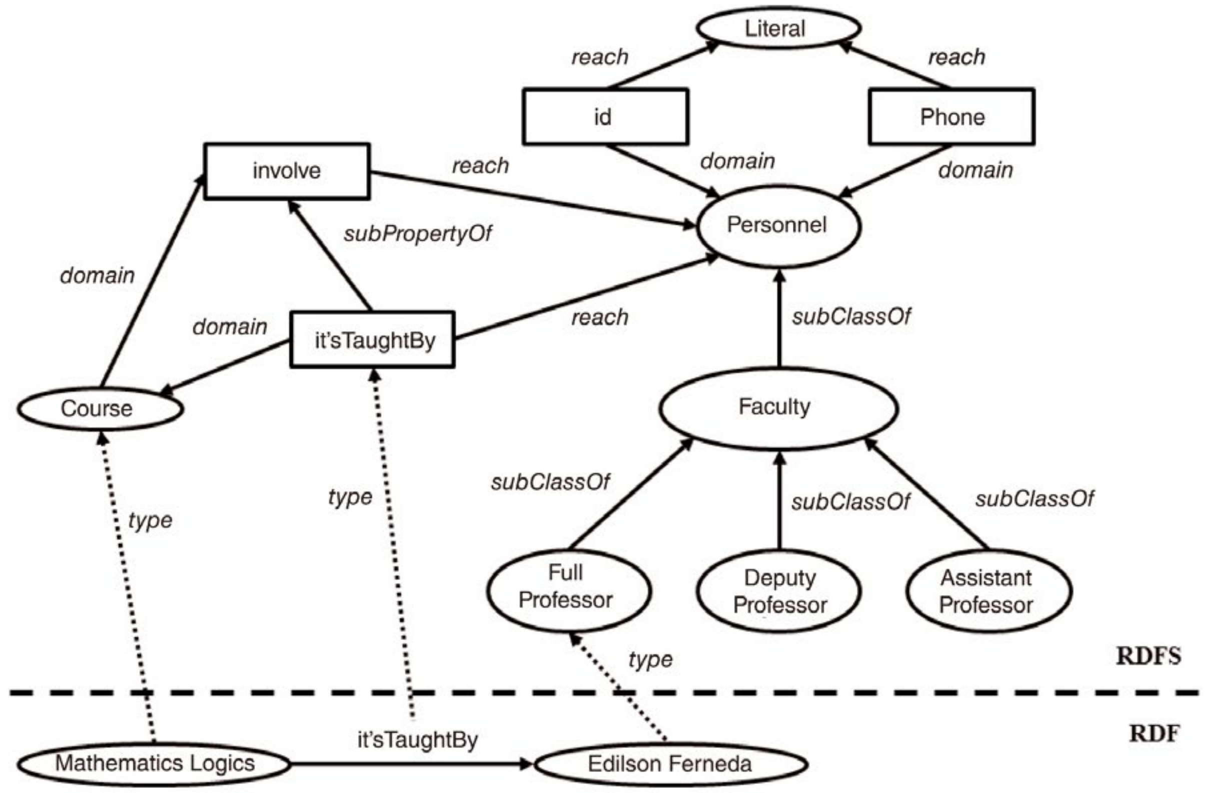


Figure 1.12: Example of RDF and RDFS levels representation. The blocks are properties, ellipses below the dashed line are instances. Source: Antoniou and van Harmelen (2004)

On a direct way, the "full professor" and "course" concepts are instantiated, as well as the property which relates them ("itisTaughtBy"). However, a question about the elements involved by this course can have an inferred answered from this property hierarchy.

6. "RDF (Resource Description Framework) is a standard model for data exchange in the Web. RDF has characteristics which facilitate the fusion of data, even if the underlying schemes are different, and specifically support the evolution over the time of schemes without needing that all the data consumers are changed" (http://www.w3.org/RDF). RDF properties can be seen as resource attributes and, with this respect, they correspond to traditional attribute-value pairs. RDF properties also represent relationships between resources. However, RDF does not foresee mechanisms to describe such properties, or provides any mechanism to describe the relationships between these properties and other resources. This is the role of RDF vocabulary description language, RDF Schema, or RDFS. RDFS defines classes and properties which can be used to describe the classes, properties and other resources. (https://www.w3.org/TR/2004/REC-rdf-schema-20040210/)

2. Conceptual modeling and ontologies

Conceptual modeling is “the activity of describing aspects of the physical and social world with the purpose of understanding and communication [...]. The adequacy of a conceptual modeling notation rests on its capacity to promote the understanding about this world between its human users (Mylopoulos, 1992).

This activity results in a model which represents a certain reality domain, according to some conceptualization. In its turn, a conceptualization can be defined as a set of concepts which articulate abstractions about referents or reality things. Such concepts are described in a model by means of a language which promotes the understanding of what is being described. This view follows the idea of the Ullmman’s triangle (1972) presented in Figure 2.1. In this figure, a concept is an abstraction of the referents in the physical and social world and the language is a way of representing this concept and refers to things noted in the interest domain (Albuquerque, 2013).

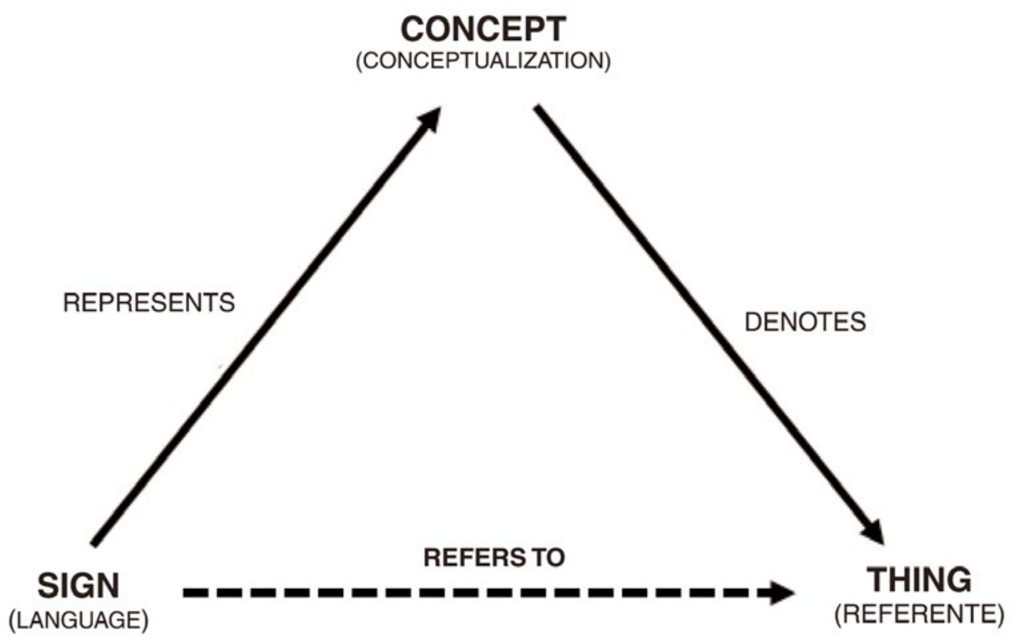


Figure 2.1: Ullmman’s Triangle
Source: adapted by Albuquerque (2013) from Ullmman (1972)

According to Guizzardi (2005), the models are concrete artifacts which, in order to be created, require modeling primitives from some modeling language. These, in their turn, normally have *(i)* an abstract syntax to represent the concepts identified in the domain of interest *(ii)* a concrete syntax to visually represent the concepts using lexical elements such as symbols, figures and connectors. For this purpose, there are several conceptual modeling languages available, especially in the Computer Science area, with the Entity-Relationship Model (MER) (Chen, 1976), a modeling language for the databanks, and the *Unified Modeling Language* (UML), oriented to system specification objects.

Almeida et al. (2009) presented the evolution of the conceptual modeling until the ontologies. According to them, the first data model specification initiatives date to the end of the 50s (Young and Kent, 1958; Bosak et al., 1962), while the creation of models was directed to computer data structures. In the following decade, surveys about data banks tool to the proposal of three main types of data models: hierarchical, in network and relational, known as logical models, once they do not refer to physical aspects (coding). However, these models are limited to the scope of the conceptual modeling. For example, modeling errors are common in the relational model (Codd, 1979), considering that the *relationship* model is used both to represent entities and relationships between entities (Peckhman and Maryanski, 1988).

The first semantic models for the conceptual modeling only appear in the 70s, in the context of the databank management systems.The so-called MER was the most important of them. The advantage of the semantic models compared to the logical models relates to the ease of understanding.

In the 90s, object-oriented modeling proposals emerge, presenting additional characteristics with respect to the data models, but, also, similarities. For example, the concepts of objects and entities; attributes and properties; relationships and associations (Milton, 2000) are equivalent. UML was an attempt to standardize the object-orientation notations (Booch, 1993; Rumbaugh et al., 1991; Jacobson et al., 1992).

However, such models count on a yet limited set of representation resources to create a conceptual model. MER, for example, assumes that the entity and relationship concepts are sufficient to represent the interest reality for an information system. However, MER lacks of several elements which would be useful for the representation of complex dependencies between the domain objects and values (Abrial, 1974;

Artale et al., 2007; Bernardi et al., 2005 apud Lembo and Poggi, 2013). Among these expressibility limits of MER, they are related, for example to the incapacity to (Lembo and Poggi, 2013):

(i) define sets of objects by means of complex properties. Assuming, for example, that the concept denotes courses on which at least one graduate student is enrolled. This could be expressed defining a restriction on the scheme presented in Figure 2.2. However, the only possibility to express such restriction would be by introducing a new relationship and a new cardinality restriction, as shown in Figure 2.3. This makes the number of relationships required to express the desired assertions to significantly increase and, consequently, unnecessarily increasing the system complexity.

*(ii)*Refine properties over an IS-A hierarchy. Assuming now that CursoAv denotes courses on which only graduate students can be enrolled. The model in Figure 2.3 does not cover this restriction, because there is now way to foresee that an undergraduate student is enrolled in an advanced course. For this purpose, resources are normally used such as the inclusion of a code in the system, not for the refinement in the conceptual model, what, once more, increases the system complexity and maintainability.

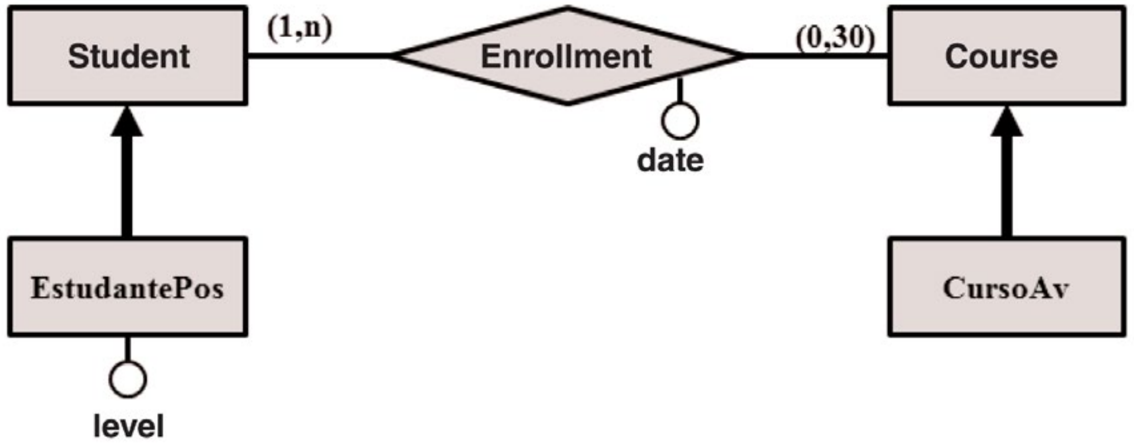


Figure 2.2: Example of Entity-Relationship model
Source: Lembo and Poggi (2013)

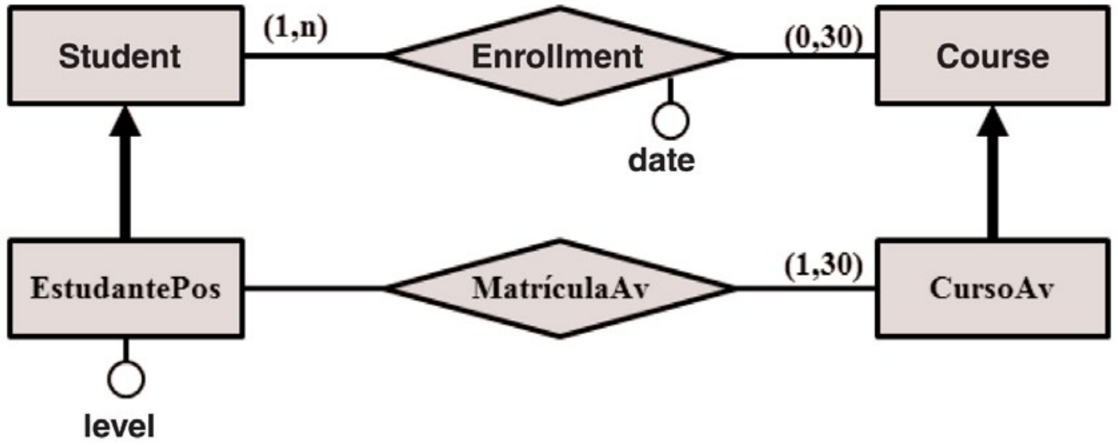


Figure 2.3: Example of Entity-Relationship model with the inclusion of a restriction
Source: Lembo and Poggi (2013)

(iii) Define polymorphic relationships. There are relationships, called polymorphic, which can be associated to distinct sets of objects. For example, both the class of people and the class of the animals can be associated to a place of birth. However, normally, while the place of birth of a person is associated to a city, the place of birth of an animal is related to some more specific thing, like a kennel. As in an ER scheme the relationships are defined between two or more entities and have a single name, it would be necessary to specify different relationships for these two cases and, possibly, the need for an addition relationships which generalizes them.

On the other hand, according to Smith and Welty (2001), the inconsistencies resulting from the limitations of these approaches in the conceptual modeling are also credited to the current interoperability difficulties in information systems.

The notion of interoperability is very wide (UKOLN, 2005 apud Farinelli et al., 2013). In the context of this work, the notions of *technical interoperability,* which covers communication, transportation, storage and information representation standards, and the *semantic interoperability,* related to the meaning of the information originated in different systems and involves the adoption of solutions that are able to guarantee uniform interpretations between the systems, like, for example, metadata, classification, thesauri and ontologies schemes.

For the *Institute of Electrical and Electronics Engineers* (IEEE)[7], interoperability is the capacity of a system or a product to work with other systems or products without special effort by the customer. It is possible thanks to the application of standards. Thus, the exchange of information between systems cannot occur without appropriate rules, policies and standards.

The models based on ontologies have been uses as an alternative to solve these types of problem, because they can serve as a robust and shared reference, with significant advantages over the *ad-hoc* methods and those based on the individual analysis used so far. A representation language based on ontological categories is a prerequisite for the modeling of a domain in the form of an ontology, in addition to foundational ontologies, which are categories independent on domains (Guizardi et al., 2004 apud Torres, 2012).

Guarino (1998) reinforces the need for a common conceptualization. According to the author (apud Almeida, 2009):

> The integration between systems is only possible if the languages underlying the models have conceptualizations[8] which overlap each other at some level. Many times, such overlapping does not exist between the conceptualizations, as it can seem by taking only the languages into account. In many cases, the languages overlap each other, but not the conceptualizations. Ultimately, only on share conceptualization provides conditions for the interoperability between information systems. Such conceptualization is the result of the human communication and consensus between groups of experts.

2.1 Ontological metaproperties

In order to support the construction of domain descriptions in an ontological language, Guarino (1992) proposes OntoClean, a methodology focused on the analysis of the meaning of a domain concepts and properties concerning the way they relate with each other in the reality. OntoClean serves as a guide for modeling by providing an ontological basis to represent the concepts in such a way to prevent misinterpretations, in addition to optimizing ontology construction process (Torres, 2012). Through this methodology, the modeling process starts by choosing the primitives which will represent the concepts form an analysis of a set of metaproperties which seeks for clarifying ontological restrictions at the knowledge level. The concept identification is facilitated by considering the rigidity, identity, uniqueness and dependence concepts.

The notion of *rigidity* is defined according to the idea of essence. A property is essential for an instance if, and only if, it is necessary for this instance. Thus, a property is rigid if, and only if, it is necessarily essential for all its instances. Thus, a property is anti-rigid if, and only if, it is not essential for all its instances. A property is non-rigid if, and only if, it is not essential for some of its instances. For example, the concept "person" is generally considered as rigid, once there is not an instance of people which is not really a person. The concept traveler is considered as being anti-rigid, once each traveler may, eventually, be a non-traveler after the end of the trip. Finally, the concept "seatable", in the sense of being possible to seat on, is non-rigid, once there are instances that are essentially seatable (for example, a chair) and instances which are not necessarily seatable, but which is possible to seat on (a box, a table).

Some concepts have a characteristic of providing an *identity* criterion or principle for its individuals, so that it is possible to distinguish and count them. For example, the concept "apple" provides an identity criterion to individuals classified by it. Therefore, it makes sense asking "how many apples are there in this box?" On the other hand, the concept red does not provide such criterion, once it does not make sense asking "how many reds are there in this box?"

The *uniqueness* seeks to identify if the object can be recognized from its parts and limits. This is conducted by analyzing the object composition identifying if it is single or the sum of single objects. The conditions required to determine if the properties have instances or single individuals is called uniqueness criterion. Thus, individuals are single when it is possible to identify their limits, such as for example, and individual of "being a sand strip". There are individuals which are single, but they have different uniqueness criteria, i.e., they do not carry uniqueness. For example, "being a quadruped" allows different uniqueness criteria among their individuals. There are properties for which the individuals are not single, i.e., the carry anti-uniqueness. For example, "being an amount of sand" has non-single individuals, once it is not possible to establish well define limits.

7. https://www.ieee.org/education_careers/education/standards/standards_glossary.html
8. A conceptualization is an abstract and simplified view of the world intended to be represented.

A concept c_1 relationally *dependent* on another concept c_2 if for instantiating c_1 the individuals need to participate in a certain relationship with the instances in c_2. Some types of dependences can be highlighted: *(i)* the dependence is *generic* if the relationship which characterizes the dependence can be changed. For example, the concept "heart" is generically dependent on the concept "person" once every person needs a heart to exist. However, the heart does not need to always be the same, considering the possibility of a transplantation. *(ii)* the dependence is *specific* if the relationship which characterizes the dependence cannot be changed. For example, a brain is specifically dependent on a person, instantiating a *part-of* relationship always with the same individuals of the person type which cannot be changes (the impossibility of brain transplant is considered, for example). *(iii)* the dependence is *existential* when the dependent individual exists only if the other individual it depends on also exists. For example, the concept "hole in the wall" is always dependent on the concept "wall".

These metaproperties are used in several ways by a number of foundational ontologies, as in case of BWW (Bunge-Wand-Weber) (Wand and Weber, 1995 apud Zamborlini, 2011) and UFO.

2.2 Ontological languages

The construction of consistent models assumes the use of representation languages which explicit formal, clear and correctly concepts and their relationships. Guarino (1995) establishes classification levels for formalisms according to the types of primitives they are based on. This classification is presented in Chart 2.1.

Levels	Primitive Constructs	Main Characteristic	Interpretation
Logical	Attributes and functions	Formalization	Arbitrary
Epistemological	Structural Relationships (Concepts and Roles)	Structure	Arbitrary
Ontological	Structural Relationships (they meet meaning postulates)	Meaning	Restrict
Conceptual	Conceptual relationships	Conceptualization	Subjective
Linguistic	Linguistic terms	Language dependence	Subjective

Chart 2.1. Main language level characteristics
Source: Guarino (1995)

In the *logical* level, the basic primitives are attributes and functions, which provide formal semantic in terms of the relationships between the domain objects. The relationships are generic and content-independent. This level formalizes the primitives, however, its interpretation is fully arbitrary.

In the *epistemological* level, it is sought to decrease the gap between the logical level, where the primitives are extremely generic, and the conceptual level, where meaning is given to the primitives. Therefore, while the logical level deals with abstract attributes and the conceptual level with specific concepts, the primitives declared at the epistemological level allow a concept to be seen as a knowledge structuring primitive. A concept in the epistemological level, which corresponds to a unary attribute in the logical level, ha an internal structure in terms of other concepts o binary relations (roles).

In the *ontological* level, the ontological commitments associated to the language primitives are explicitly specified by means of the semantic restriction of the primitives or meaning postulates directly introduced in the language. The objective is to restrict the possibility of diverging interpretations. Therefore, the ontological level is focuses on establishing the meaning of the domain in terms of primitives. It can be said that a language is ontologically appropriate if, in the syntactic level, it has sufficient granularity and reification capacity to express the meaning postulates of its own primitives, or if it is possible to provide formal ontological interpretation to its primitives.

In the *conceptual* level, the primitive have a defined cognitive interpretation, comprising language-independent concepts. The framework of the domain structure is defined and the knowledge is specified as a specialization of such framework. The definitions at this level can be accompanied by definitions in the ontological level.

In the *linguistic* level, the primitives are intended to directly refer to terms such as verbs and substantives. These definitions are important, once the knowledge abstraction level must be defined before the modeling step. A knowledge can be specified in the conceptual level and never be implemented in a formal representation language.

In this context, the foundational ontologies have been used in the evaluation and (re)engineering process of conceptual modeling languages once they describe categories used in the construction of conception which represents certain part of reality.

For the conceptual modeling of goods and services, below is a general view of the OntoUML languages – considered as being of ontological level – and OWL – seen as a more epistemological level and considered as standard for use in the Semantic Web.

2.2.1 ONTOUML

On a simplified way, it can be said that OntoUML, proposed by Guizzardi (2005), is a conceptual modeling language derived from UML added with ontological metaproperties. This strategy seeks to solve the lack of an exact definition of the formal semantic of UML language and extends the metamodel of this to be isoformic to UFO, making it ontologically consistent. UML has extension mechanisms allowing to change the language elements in such a way that a coherent set of such extensions constitutes a UML profile. The OntoUML is, therefore a UML profile composed by a set of stereotypes which represent the ontological categories of the universal types proposed in UFO, as well as by formal restrictions which reflect the UFO axiomatization in such a way that it restricts the set of grammatically valid models in OntoUML to those which represent admissible situations according to the UFO theory (Guizzardi, 2007). Thus, the explanation for applying the stereotypes is identical to the application of the corresponding universal types of UFO.

The quote of this language is due to the fact that the sheet-elements of the OntoUML language are the relevant stereotypes to meet the conceptual modeling of products in order to meet SMPE requirements which will be discussed later on in this document. As an example, Figure 2.4 present fragments of the OntoUML metamodel on which the representation element *Class* of UML is specialized according to the UFO Monadic Universal types. Also, UML representation element *Relationship* is specialized according to the UFO Relationship Universal types, and it is equally important for the products conceptual modeling (Figure 2.5). The basic form and meronimic types specialize a special type of *relationship,* namely, *Directed Binary Relationship*, while other types specialize *Association,* which refers to enary relationships. Particularly, the stereotype *Formal Association* refers to the UFO universal type *Domain Formal Relation.* Charts 2.2. and 2.3 present the descriptions and examples of OntoUML concepts and relationships, respectively.

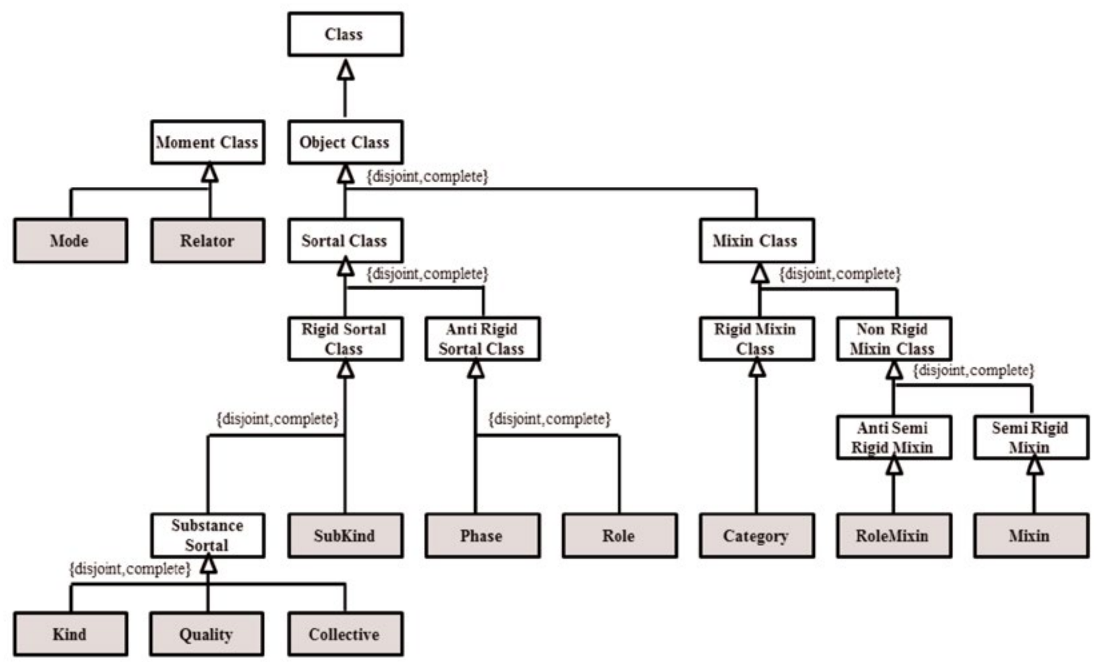


Figure 2.4: Fragment of metamodel OntoUML to describe classes
Source: Zamborlini (2011)

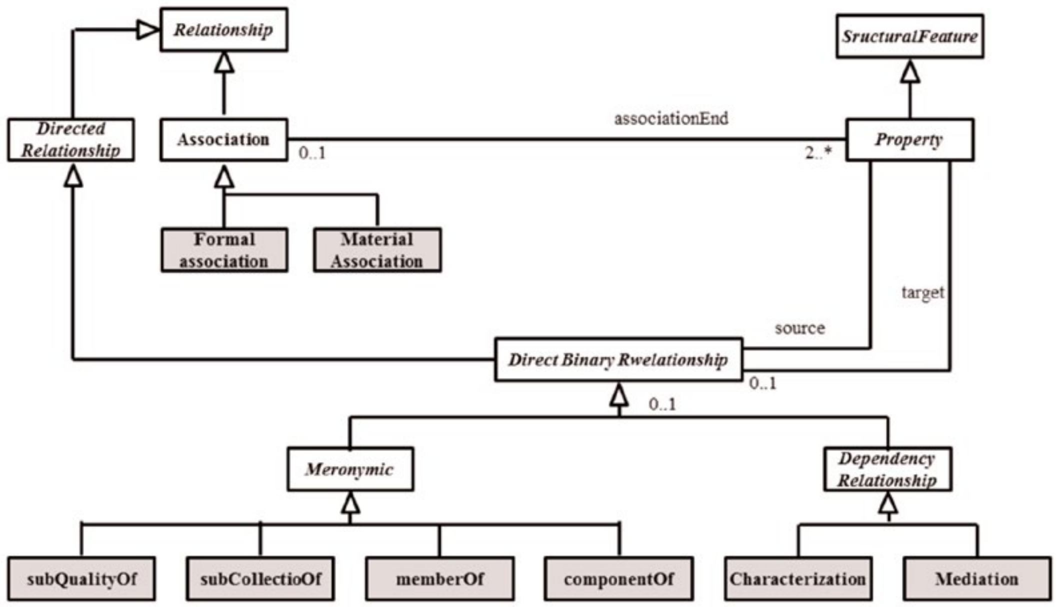


Figure 2.5: Fragment of metamodel OntoUML to describe relationships between concept instances
Source: Zamborlini (2011)

Constructs	Description
<<Kind>>	Represents a substance sortal* for which the instances are functional complexes. Examples: Person, Car, Chair, Television.
<<Quantity>>	Represents a substance sortal for which the instances are amounts. Examples: Clay, Wine, Gold.
<<Collective>>	Represents a substance sortal for which the instances are collectives or collection of complexes presenting a uniform structure. Examples: Forest, Cards, Football Team.
<<Subkind>>	Represents a rigid sortal which restricts a sortal substance which suppresses its identity. Examples: Carnivorous and Herbivorous (Animal subtypes).
<<Phase>>	Represents an anti-rigid sortal for which the instances are phases (of an individual). Phase are necessarily applicable in some possible world, but not in any possible world. Examples: Child, Adolescent, Adult, Elderly (person phases).
<<Role>>	Represents an anti-rigid sortal for which the instances are roles assumed by the individual who normally participates of some event or some relationship. Roles are necessarily applicable in some possible world, but not in each possible world. Examples: Father, Mother, Student, Teacher (Person roles).
<<Mixin>>	Represents a semi-rigid mixin which adds essential properties for some of its instances but is accidental for others. Examples: Colored, Seatable (essential for Chair and accidental for Table).
<<Category>>	Represents a rigid mixin which adds common essential properties for different substance sortals. Example: EntidadeRacional (as a generalization of the Person types and AgentInteligente).

Sortal is a concept associated to identity, persistence and change issues. The simplest property of a sortal is that it can be counted, i.e., it can have numerical values as modifiers. Thus, "sandwich" is a sortal in the sentence "I want a sandwich", but "water" is not a sortal in the sentence "I want some water". Sortal substances are sortal concepts applied to an object over its existence.

Chart 2.2: OntoUML concepts
Source: Adapted from Guizzardi (2005) by Albuquerque (2013)

Relationship	Description
<<characterization>>	Represents the formal relationship placed between intrinsic and universal moments characterized by them. *Example:* the mode Symptom characterizes Patient.
<<mediation>>	Represents the formal relationship placed between the reporters and the universal ones which are mediated by them. *Example:* the reporter Wedding mediates the Husband and Wife mediation relationships.
<<derivation>>	Represents the formal relationship placed between the reporters and the material relationships which are derived from the reporters. *Example:* the material relationship married-with derived from reporter Wedding.
<<material>>	Represents the material relationship placed between the universal mediated maintainers and are inserted by reporters. *Example:* married-with (between husband and Wife)
<<formal>>	Represents the formal relationship which can be both a comparative relationship (derived from the concerned entities' qualities) and an internal relationship. *Example:* older-than (between People, based on property age)

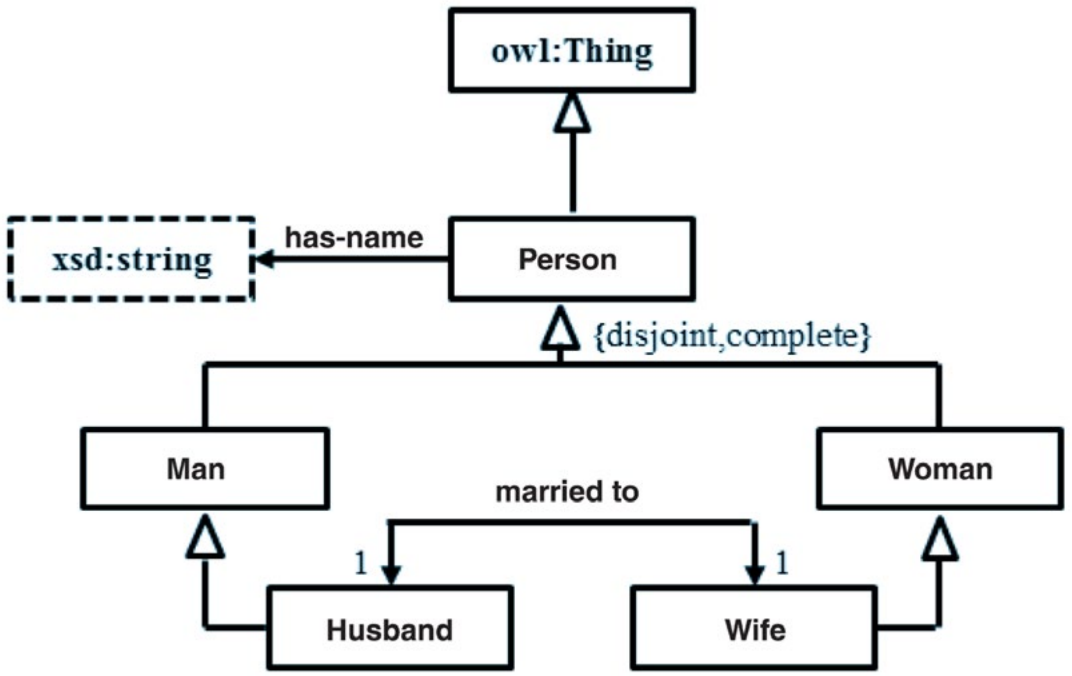
Chart 2.3: OntoUML Relationships
Source: Adapted from Guizzardi (2005) by Albuquerque (2013)

2.2.2 OWL

The OWL (*Web Ontology Language*) language is a more epistemological conceptual modeling language recommended by the *World Wide Web Consortium* (W3C) for the representation of ontologies in the Semantic Web. For this purpose, it was designed to represent object categories, their relationships and the description of the objects themselves.

The OWL objective is to meet the following restrictions: *(i)* maintaining the compatibility with RDF standard of Web information representation, extending, however, the capacity to express the said "ontological" knowledge; *(ii)* having well-defined syntax and semantic, as well as the expressiveness power which maintains desirable computational properties. (Antoniou and Harmelen, 2008).

The OWL serves to represent complex objects, making use of RDF to create ontologies as a form of knowledge representation (Staab and Studer, 2009). It is currently in version 2 (OWL 2.0), which solves some limitations noted in previous versions. Figure 2.6 shows the same example represented by an UML diagram and in OWL, constructed by Zamborlini (2011).



Classes: Person, Man, Woman, Husband, Wife
Object Properties: married-to
Data Properties: has-name

(1) **SubClassOf** (Man Person)
(2) **SubClassOf** (Woman Person)
(3) **DisjointClasses** (Woman Man)
(4) **EquivalentClasses** (Person **ObjectUnionOf** (Woman Man))
(5) **SubClassOf**(Person **DataExactCardinality** (1 has-name xsd:string))
(6) **SubClassOf** (Husband Man)
(7) **EquivalentClasses** (Husband **ObjectSomeValuesFrom** (married-to Wife))
(8) **SubClassOf** (Wife Woman)
(9) **EquivalentClasses** (Wife **ObjectSomeValuesFrom** (married-to Husband))
(10) **FunctionalObjectProperty** (married-to)
(11) **SymmetricObjectProperty** (married-to)
(12) **ObjectPropertyDomain** (married-to **ObjectUnionOf** (Husband Wife))
(13) **ObjectPropertyRange** (married-to **ObjectUnionOf** (Husband Wife))
(14) **DataPropertyDomain** (has-name Person)
(15) **DataPropertyRange** (has-name xsd:string)

Figure 2.6: Example of context described in UML for representation in OWL
Source: Zamborlini (2011)

The main distinction in this example is that the individuals classified as Person are men or women. This distinction is represented in the diagram by the generalization set which contains the classes Man and Woman, which specialize the class Person, labeled with terms *disjoint* (there is no intersection between individuals of these two classes) and *complete* (there is no other classification possibility in addition to the specified ones).

In Zamborlini's (2011) words:

> The specializations, represented in lines 1 and 2 of the OWL model, mean that the interpretation of the class Man, as well as Woman, is contained in the interpretation of class Person, i.e., every individual who is a man or woman is also a Person.
>
> The term *disjoint,* represented in line 3, states that classes Man and Woman are disjoined, and means that the intersection of their interpretation is empty, i.e., a man cannot be a woman and vice-versa. The term *complete,* represented in line 4, states that class Person is partitioned on a complete way by classes Man

and Woman, and means that the interpretation of class Person is equivalent to the interpretation of the joint of classes Man and Woman, i.e., every person is a man or a woman. In addition, a Man can be a Husband, and a Woman can be a Wife. The explanation of these specializations is similar to the other one already performed in this paragraph.

It is also true that every person has exactly a name, which is of *string* type. This is represented in the diagram by the relationship named *has-name* between class Person and the data type *xsd:String* with a cardinality restriction equal to 1. In the OWL model, this refer to line 5, and it means that the interpretation of class Person is equivalent to the interpretation of the class expression which corresponds to all the individuals which instantiate the data property *has-name* with exactly one data domain element of *String* type.

Lines 14 and 15 further restrict, respectively, the domain and the image of this property. The first one means that, for each instance of property *has-name*, the property's domain element belongs to the interpretation of class Person, for example, if *has-name*(João, "João da Silva"), then, the interpretation of João belongs to the joint of class Person interpretation. Analogously, the second one means that, for each property tem-name's instance, the image element belongs to the interpretation of the data type String, for example, if *has-name*(João, "João da Silva"), then, "João da Silva" has to be a string.

An individual of type Husband is necessarily married with exactly one of Wife type, and vice-versa. This situation is represented in the diagram by the bidirectional relationship *married-to* between classes Husband and Wife. In the OWL mode, this refers to lines 7, 9 and 10 to 13. The two first ones mean that the interpretation of class Husband (Wife) is equivalent to the interpretation of the class expression corresponding to all the individuals which instantiated the property of objects *married-to* with some element in the domain which belongs to the interpretation of class Wife (Husband). Also, line 10 means that if the property *married-to* is instantiated twice for the same individual, for example, *married-to*(João, Maria) and *married-to*(João, Ana), then, Maria and Ana are the same person, i.e., they have the same interpretation.

Line 11 further means that whenever the property *married-to* is instantiated for an ordered pair, it is also true for its symmetric pair, for example, if *married-to*(João, Maria), then, *married-to*(Maria, João). Finally, lines 12 and 13 respectively restrict domain and image of property *married-to*, and, being similar, they mean that, for each instance of property *married-to*, the domain (or image) element of the property belongs to the interpretation of the class expression that all the individuals who belong to the joint of the interpretation of classes Husband and Wife, for example, if *married-to*(João, Maria), then, the interpretation of João belongs to the joint of the interpretation of classes Husband and Wife, as well as Maria.

An example of syntax and formal semantic of OWL constructs are presented by Zamborlini (2011) in Chart 2.4, on which some of the elements used to describe the context presented in Figure 2.6 are.

Syntax	Semantics
DECLARATIONS	
Declaration(Class(C **))**	$C^I \subseteq \Delta^I$
Declaration(ObjectProperty(OP **))**	$OP^I \subseteq \Delta^I \times \Delta^I$
Declaration(Datatype(DT **))**	$DT^D \subseteq \Delta^D$
Declaration(DataProperty(DP **))**	$DP^I \subseteq \Delta^I \times \Delta^D$
CLASS EXPRESSIONS	
ObjectComplementOf (C) ,	$\{x \mid x \in \Delta^I \setminus C^I\}$
ObjectIntersectionOf (C_1 .. C_n) ,	$\{x \mid x \in C_1^I \cap .. \cap C_n^I\}$
ObjectUnionOf (C_1 .. C_n) ,	$\{x \mid x \in C_1^I \cup .. \cup C_n^I\}$
ObjectSomeValuesFrom (OP C)	$\{x \mid \exists y : (x,y) \in OP^I \wedge y \in C^I\}$
ObjectAllValuesFrom (OP C)	$\{x \mid \forall y : (x,y) \in OP^I \rightarrow y \in C^I\}$
ObjectMinCardinality (n OP C)	$\{x \mid \#\{y \mid (x,y) \in OP^I \wedge y \in C^I\} \geq n\}$
ObjectMaxCardinality (n OP C)	$\{x \mid \#\{y \mid (x,y) \in OP^I \wedge y \in C^I\} \leq n\}$
ObjectExactCardinality (n OP C)	$\{x \mid \#\{y \mid (x,y) \in OP^I \wedge y \in C^I\} = n\}$
DataSomeValuesFrom (DP DT)	$\{x \mid \exists y : (x,y) \in DP^I \wedge y \in DT^D\}$
DataMinCardinality (n DP DT)	$\{x \mid \#\{y \mid (x,y) \in DP^I \wedge y \in DT^D\} \geq n\}$
DataMaxCardinality (n DP DT)	$\{x \mid \#\{y \mid (x,y) \in DP^I \wedge y \in DT^D\} \leq n\}$
OBJECT PROPERTIES EXPRESSIONS	
InverseObjectProperty (OP)	$\{(x,y) \mid (y,x) \in OP^I\}$
CLASS EXPRESSION AXIOMS	
SubClassOf (SC C)	$SC^I \subseteq C^I$
EquivalentClasses (C_1 C_2)	$C_1^I = C_2^I$
DisjointClasses (C_1 .. C_n)	$C_j^I \cap C_k^I = \varnothing, 1 \leq j < k \leq n$
ObjectPropertyDomain (OP C)	$\forall x,y\ (x,y) \in OP^I \rightarrow x \in C^I$
ObjectPropertyRange (OP C)	$\forall x,y\ (x,y) \in OP^I \rightarrow y \in C^I$
OBJECT PROPERTIES EXPRESSION AXIOMS	
SubObjectPropertyOf (SOP OP)	$SOP^I \subseteq OP^I$
SubObjectPropertyOf(**ObjectPropertyChain** (OP_1 .. OP_n) OP)	$\forall y_0 .. y_n\ (y_0, y_1) \in OP_1^I \wedge .. \wedge (y_{n-1}, y_n) \in OP_n^I \rightarrow (y_0, y_n) \in OP^I$
FunctionalObjectProperty (OP)	$\forall x,y_1,y_2\ (x,y_1) \in OP^I \wedge (x,y_2) \in OP^I \rightarrow y_1 = y_2$
IrreflexiveObjectProperty (OP)	$\forall x,y\ (x,y) \in OP^I \rightarrow x \neq y$
SymmetricObjectProperty (OP)	$\forall x,y\ (x,y) \in OP^I \rightarrow (y,x) \in OP^I$
DATA PROPERTIES EXPRESSION AXIOMS	
SubDataPropertyOf (SDP DP)	$SDP^I \subseteq DP^I$
DataPropertyDomain (DP C)	$\forall x,y\ (x,y) \in DP^I \rightarrow x \in C^I$
DataPropertyRange (DP DT)	$\forall x,y\ (x,y) \in DP^I \rightarrow y \in DT^D$
FunctionalDataProperty (DP)	$\forall x,y_1,y_2\ (x,y_1) \in DP^I \text{^} (x,y_2) \in DP^I \rightarrow y_1 = y_2$

Chart 2.4: Functional and semantic syntax of some OWL constructs
Source: Zamborlini (2011)

In this chart, the functional syntax is adopted as defined by Motik et al. (2012b). The Semantic Theory, corresponding to the description logic SROIQ(D) (Horrocks et al., 2006), is presented as defined by Motik et al. (2012a), however, on a simplified way.

> The OWL semantics is defined by using a model semantics (*Model-Theoretic Semantics*) or the so-called *Tarskian Semantics*, in honor of Polish logician Alfred Tarski. In this theory, a called *model* includes a set (called domain or universe) D^I and an interpretation function$.^I$. Domain D^I is the set of elements for a certain interpretation, and the interpretation function $.^I$ is a mapping with the names of individuals in the domain elements, names of classes in the domain subsets, and names or properties individuals in sets of ordered pairs in the domain elements. The interpretation role can be extended to expressions of classes or properties. A separate domain D^D is used for the data type values, as well as an interpretation role $.^D$ which maps a data type name in this domain's elements. Finally, an interpretation meets an ontology if, and only if, it meets each axiom and ontology fact; and an ontology is consistent if, and only if, it is met by at least one interpretation. (Zamborlini, 2011)

3. Electronic commerce: from the traditional web to the semantic web

The last century saw the number of types and varieties of products traded at global level vertiginously grows[9]. As a consequence, the complexity and the costs associated to seeking for such products increased at the same ratio. According to Wallis and North (1986), more than 50% of the Gross Domestic Product of the United States in 1970 was related to the "transaction services" provision activity, i.e., the support for seeking specific products (for example, real estate broker). I.e., specificity is the key to size the search courses (Hepp, 2010a), once it increases the search space and the complexity of the demands.

The advent of Web has drastically changed this scenario, once the search costs were considerably reduced. Despite of this, there is still a great space for improvements, particularly under the perspective of the electronic commerce, once some of the characteristics of the traditional Web create some obstacles to the electronic commerce. The fact that the searches in the traditional Web are conducted by means of syntactic

9. In 1920, the number of item types marketed in the world was 5,168, according to Beythien and Dreßle (2000 apud Hepp, 2010b).

comparisons make the linguistic aspects, such as homonymy[10] and polysemy[11], to significantly increase the search efforts. In addition, although the connection of Web pages with databank systems is increasingly frequent, due to the lack of connection between this data, the integrated use of these bases is not trivial. This makes the integration/synthesis of this data to be under the user's responsibility. In the context of the electronic commerce, this means that a user who, for example, seeks for best offer conditions for a certain product, has to check a number of sites to find what he/she is looking for, without, however, a guarantee that the best option is found.

This has lead to alternatives being proposed in order to make the search for products feasible with higher efficiency and efficacy. Among these alternatives is the use of what has been called Semantic Web[12].

3.1 Semantic Web

While the current Web can be classified as syntactic and the process for interpreting the contents made available is generally under the user's responsibility, the Semantic Web[13], extension of the current Web, is able to associate explicit meaning to the contents of the documents available at the Internet, being that its main goal is to allow computer systems to automatically process and interpret such documents.

The Semantic Web aims at providing the computers with the capacity to access structured data and to define inference rules, generating useful information from large volumes of data (Berners-Lee et al., 2001). This adds semantics to the Web pages by means of some technologies: *(i) Extensible Markup Language* (XML)[14] (Harold and Means, 2004) and *Uniform Resource Identifier* (URI)[15], for a simplified representation of the Web information, *(ii) Resource Description Framework* (RDF) (Powers, 2003), for the structured representation of data, *(iii) Web Ontology Language* (OWL)[16] (Motik et al., 2012a, 2012b), in order to associate meaning to data, and *(iv) SPARQL Protocol and RDF Query Language* (SPARQL), semantic search language able to recover and manipulate data stored in RDF.

An ontology provides a semantic structure for the annotation of Web pages. It is expected that, with the structure provided by the Semantic Web, it is possible to obtain more precise searches (once semantics will be placed in formal language) and to provide increased capacity to the software agents who use/manipulate the Web content. Figure 3.1 represents the framework proposed by Berners-Lee for the Semantic Web. In the Semantic Web, the data is represented with increases richness, once they are associated to vocabularies and have a meaning. Thus, the data does not need to be isolated, being shared by several applications. In the context of the Semantic Web, the data is called resources, they are unique and have their own identification in the Web.

In their turn, relational databank systems represent data sets in the form of tables, accessible only for those who connect with the pages making reference to them. For the compatibility with the Semantic Web objectives, the silos of data residing in these tables have been converted and made available in RDF open format, in a process known as triplification. For example, each element of the table shown in Figure 3.2 and their respective row and column is converted into a RDF sentence which, in its turn, results in a RDF triples' table published in the open form in the Web, as illustrated in Figure 3.3. In this case, the data representation in the Semantic Web occurs on a global manner in order to reference the rows (subject), the columns (object) and the values (attributes) of a table, i.e., in the form of a triple. As it can be seen in this figure, a number of triples can refer to the same entity and can be represented by means of graphs (Figure 3.4). Triples of different sources can be combined in the same graph (Figure 3.5). In the Semantic Web, these triples are represented in RDF.

10. Two or more words are homonymous when they have distinct origin and meanings, but with the same spelling or the pronunciations.

11. Polysemy occurs when a certain word or expression acquires a new sense in addition to its original sense, maintaining a sense relationship that is different between them.

12. Semantics is the study of the meaning. It concerns about the relationship between signifiers, such as words, sentences, signs and symbols, and what they represent, their denotation. (Wikipedia)

13. "The Semantic Web is the extension of the Web which allows the people to share content in addition to the application limits and websites. It is described in very different ways: as a utopic view, like a data network, or only like a natural paradigm change in our daily use of the Web." (semanticweb.org)

14. "Extensible Markup Language (XML) is a markup language which defines a set of rules for the coding of documents in a format that is both readable by humans and machines. It is defined by W3C XML 1.0 Specification and by various other related specifications, all of them being free open standards." (Wikipedia)

15. Uniform Resource Identifier (URI) is a compact characters' chain used to identify or denominate an Internet resource. The main purpose of this identification is to allow the interaction with the resource representations through a network, typically the Worldwide Network, using specific protocols. URIs are identified in groups defining a specific syntax and associated protocols. (Wikipedia)

16. "Web Ontology Language (OWL) is a Semantic Web language designed by W3C in order to represent the rich and complex knowledge about the things, groups of things and the relationships between the things. OWL is a computer logics-based language so that the knowledge expressed in OWL can be explored by computer programs, for example, in order to verify the consistency of this knowledge or to make the implicit knowledge explicit. Ontologies can be published in the World Wide Web and may refer or be referred by other OWL ontologies." (http://www.w3.org/2001/sw/wiki/OWL)

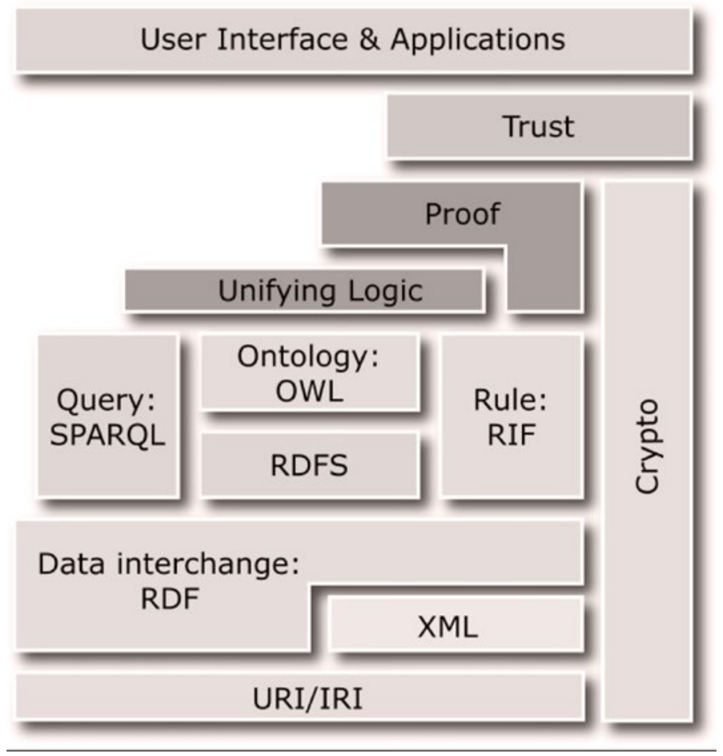


Figure 3.1: Berners-Lee's Framework for the Semantic Web
Source: http://www.w3c.tut.fi/talks/2007/1031-soa-ws-web20-on/index-rdf.html

Nº	Title	Author	Year	Genus
01	Mar morto	Jorge Amado	1936	Romance
02	A Estrela do Mar	Jorge Amado	1938	Poetry
03	O mundo da paz	Jorge Amado	1951	Travel Report
04	O quinze	Raquel de Queiroz	1930	Romance
05	Tieta do Agreste	Jorge Amado	1977	Romance

Figure 3.2: Example of a database system table
Source: Lóscio (2012)

Subject	Predicate	Object
Jorge Amado	escreveu	Mar morto
Zélia Gatai	casadaCom	Jorge Amado
Jorge Amado	viveuEm	Salvador
Salvador	éEm	Bahia
Bahia	parteDe	Brazil
Ceará	parteDe	Brazil
Jorge Amado	escreveu	Tieta do Agreste

Figure 3.3: Representation of a dataset in the form of triples
Source: Lóscio (2012)

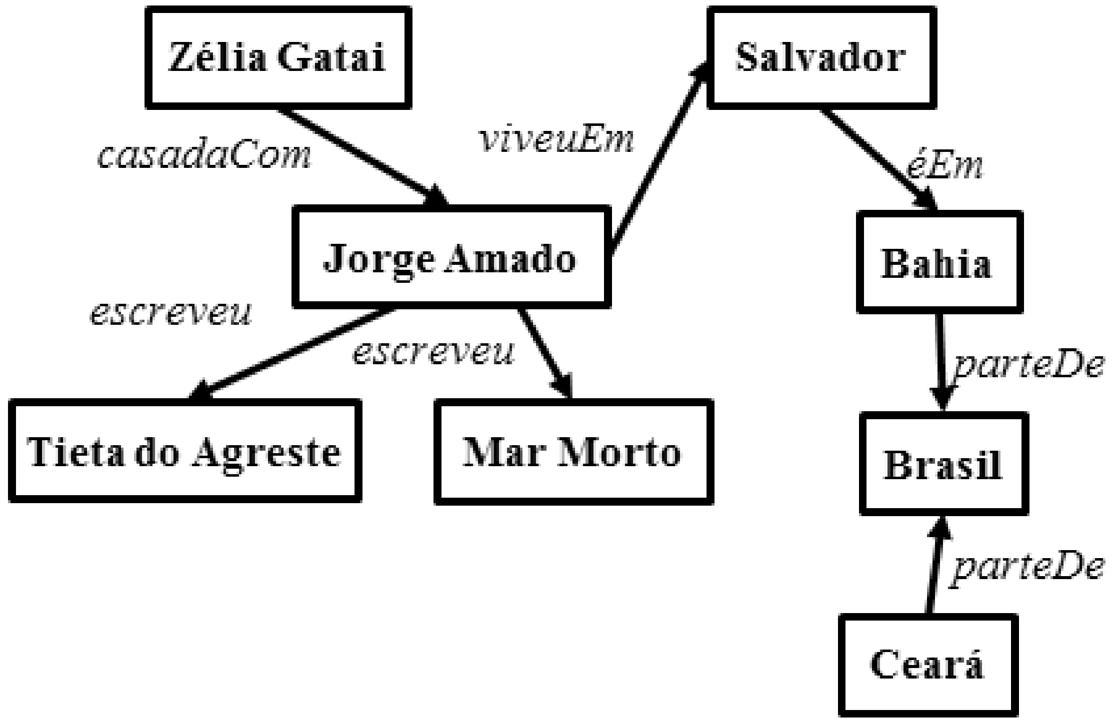


Figure 3.4: Representation of a set of triples as a graph
Source: Lóscio (2012)

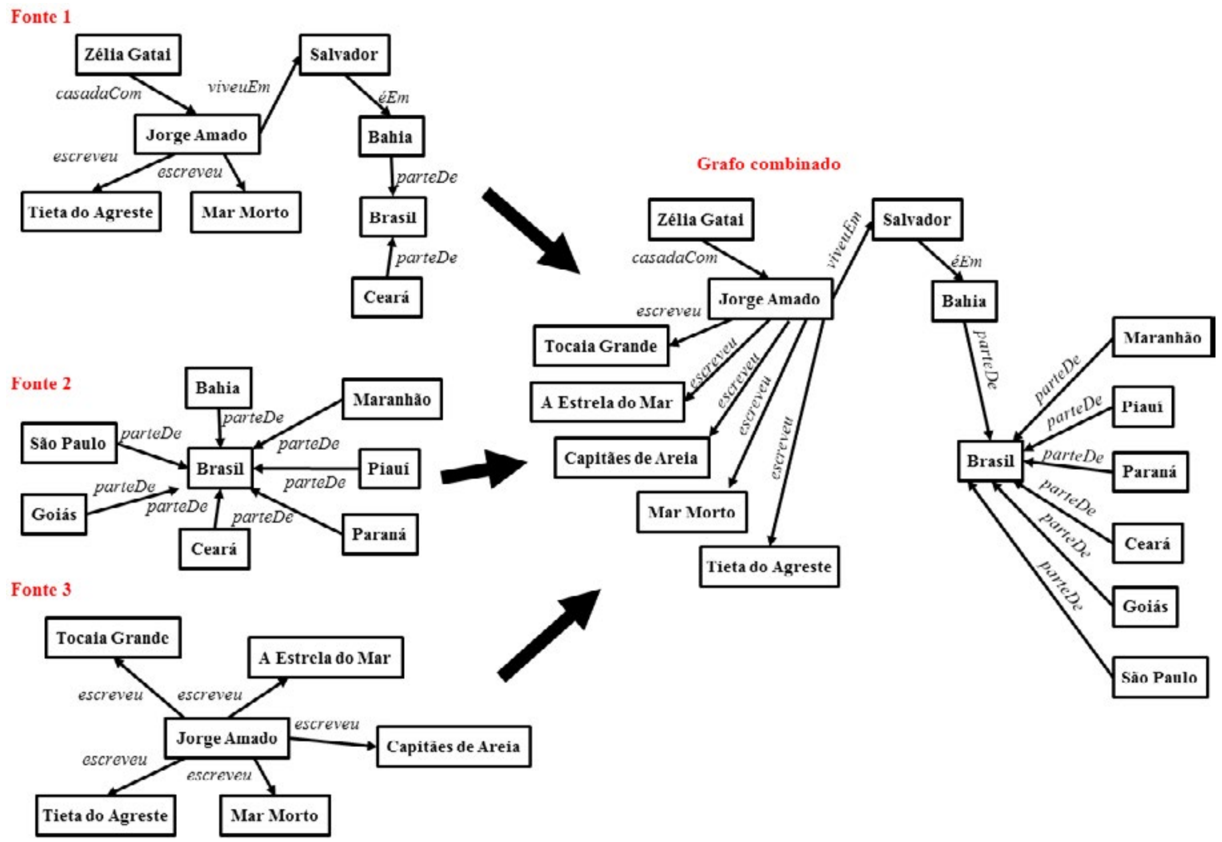


Figure 3.5: Graph generated by integrating data from several sources
Source: Lóscio (2012)

RDF format constitutes a homogeneous form to publish data in the Web. The publication of data in this format is being proposed by initiatives such as *Linked Open Data.* As informed by LOD Brasil[17]:

> The publication of data in the Web is not a novelty, but the traditional approaches did not usually pass from the simple availability of heterogeneous data in formats such as CSV, XML or HTML pages, with little concern about its integration, of semantic structure. As well as in the traditional web, formed by hypertext links, the data network in *Linked Data* results from the connection between the documents present in the web. However, the connections are made using hyperdata links with information expressed in RDF, on which not only the document can be linked, but also the information presented in them.

17. http://lodbrasil.com.br/

3.2 Linked Data

Linked Data[18] is a proposal to reach the global data space, the Data Web, where: *(i)* any person can publish his/her data, *(ii)* data items are connected by links, *(iii)* applications may discover new sources browsing the links and*(iv)* the data is self-descriptive. With this respect the initiative *Linked Open Data* (LOD)[19] emerged. This term refers to a set of good practices (or principles) intended to publish and establish links between sets of open data intended to add in the Web space a kind of global level data bank (or *Linked Open Data Cloud*). Thus, the *Linked Data* is not more than the creation of links between data from different sources connected to the following principles: *(i)* using URIs as names for resources; *(ii)* using URIs HTTP for the people to find these names; *(iii)* when somebody searches for a URI, through this, providing useful information, by means of RDFs; *(iv)* including RDF sentences connecting other URIs so that they can discover more resources.

Currently, there are several *datasets* made available this way. Figure 3.6 represents the state of the LOD in August 2014. In this figure: *(i)* the nodes are datasets published in the *Linked Data* format; *(ii)* the size of the circles (nodes) corresponds to the number or triples of each dataset; *(iii)* the arrows indicate the existence of at least 50 links between two datasets; *(iv)* a link is a triple RDF, where subject and object are in different datasets.

Figure 3.6: Linked Open Data's datasets Cloud connected in august 2014
Source: http://data.dws.informatik.uni-mannheim.de/lodcloud/2014/ISWC-RDB/LODCloudDiagram.png

3.3 Electronic commerce in Semantic Web

By understanding the Web as a large bank of linked data, in RDF format and with semantic outlined by ontologies, in addition to the possibility of having data management software which accept search languages like SPARQL, sophisticated applications for the electronic commerce become possible. In fact, the electronic commerce is increasingly incorporating the Semantic Web tools for better developing its activities. Among these advantages is the possibility to compare prices and

18. "*Linked Data* relates to the use of the Web to connect correlate data which were not previously linked, or to the use of the Web to overcome barriers for the link of data normally connected by means of other methods." (linkeddata.org) / "This is a term used to describe a set of best practices recommended to publish, share and connect data parts, information and knowledge in the Semantic Web using URIs and RDF." (Wikipedia)

19. "The objective of the community project W3C SWEO Linking Open Data is to extend the Web so that the data is common good by means of publication in the Web of a number of data sets open in RDF form and defining RDF links between data items from different data sources" (http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData)

conditions of different suppliers in the same search with an increasingly better level of search details and answer and consequent user satisfaction. This satisfaction occurs through the higher inference capacity made available, increasing the capacity of the computer systems to provide assistance in the user decisions. I.e., the searches are made available with increased richness of information both in the demands and in the answers to it. It becomes possible, for example, the incorporation of functionalities such as the verification of the product compatibility, manufacturers, among other details.

Below is an example of the use of the Semantic Web for the knowledge organization to meet the electronic commerce needs of a scientific magazines publisher (Antoniou and Van Harmelen, 2004):

Elsevier horizontal information products

Scenario: Elsevier is one of the main science publishers worldwide. Elsevier products, and of many of its competitors, were organized, particularly over the traditional lines: magazine subscriptions. The availability of these magazines online had not changed the organization of the products line. Although the articles were available online, these were in the same form on which they appeared in the magazine, and the article collections were organized according to the magazine where they appeared. Elsevier customers could do online content subscriptions, but, again, these subscriptions were organized according to the traditional products lines: magazines or groups of magazines.

The problem: These traditional magazines could be described as "vertical products": the products would divide all the sciences in a number of separate "columns" (for example, biology, chemistry, medicine), and each product would cover a column (or, more likely, part of such column). However, with the fast evolution of science in a number of areas (information sciences, life sciences, physical sciences), the traditional division of the science into fields covered by distinct magazines is no longer satisfactory. In contrary, Elsevier customers are interested in covering certain thematic areas spread through traditional science subjects. The pharmaceutical company wants to purchase from Elsevier all the information it has, for example, about Alzheimer's disease, regardless this comes from a biology magazine, a medicine journal or a chemistry magazine. Thus, there was a great demand for "horizontal" products: all the information Elsevier has about a certain subject, crossing all the traditional disciplines. It was difficult, if not impossible, for great publishers like Elsevier, to provide such "horizontal" products. The information published by Elsevier was blocked inside each magazine, each one with its own indexation system, organized according to different physical, syntactic and semantic standards. Concerning the physical and syntactic heterogeneity, Elsevier was translating much of its content to a XML format, allowing crossed searches to a number of journals. However, the semantic problem was not solved. Of course, it is possible to search several magazines seeking for articles containing the same key words, but give the extensive homonym and synonym problems in and (even more) between several disciplines, it is unlikely to have satisfactory results. It was required a manner to search the various magazines about a coherent set of concepts on which all these magazines are indexed.

The technological contribution of Semantic Web: ontologies and thesauri (which can be seen as very light ontologies) have proven to be key technologies for the access to efficient information, helping to overcome some of the free text search problems, relating and grouping relevant terms in specific domain, as well as providing a controlled vocabulary for indexation information. Dictionaries were developed in different specialization domains. These dictionaries were already used to access information sources such as MBASE or Science Direct, however, there were no links between the different information sources and the specific dictionaries used to index and consult these sources. At first, Elsevier implanted a system of access to multiple information sources in the life sciences area through a single interface, using a single underlying ontology through which all the "vertical" information sources were indexed. The Semantic Web Technology performed a number of roles in this architecture. First, RDF is used as an interoperability format between the heterogeneous data sources. Second, an ontology presented in RDF. Each of the separated data sources is mapped for this unifying ontology, which is, then, used as a single input point for all these data sources.

Part II – A product classification and offer proposal

In order to provide subsidies for the Electronic Business Market to become a "meeting point between the Brazilian company and its business partners", it is necessary to find a product knowledge organization system both for product classification and listing. With this respect, this part of the document conducts a critical analysis of the currently used systems and their adherence degree concerning the interoperability requirements desired for the commercial relationships between micro and small entrepreneurs.

First, an analysis is performed about the product classification systems concerning the requirements presented by SMPE. Subsequently, in face of the limitations found in these traditional systems, a tool proposal is presented for modeling the production chains, supply chains or value chains considering the use of knowledge organization systems supported by ontologies. In addition, a modeling methodology is presented for the construction of the desired classification and listing scheme. Finally, a connection of this type of approach with another Electronic Business Market related to the product offer is presented.

4. Conceptual modeling of products with foundational ontology support

This chapter presents an alternative for the classification and listing of products in the scope of the Electronic Business Market. In this case, these functionalities will be proposed from a conceptual modeling of the production chains supported by ontologies. In addition, a contextualization is conducted in order to present the limitations of the taxonomic systems commonly used for this type of demand.

4.1 Contextualization

As previously mentioned, the taxonomies have been used for the knowledge organization about production activities, covering all the goods and services areas in order to generate information which help to indicate the government decisions about such activities. Such mechanisms are defined and coordinated by specific sector and generated useful taxonomies for the systematization of information about the production chains, as in Brazil, of CNAE and the Brazilian Nomenclature of Services (NBS), used and connected with some of the Federal Public Administration structuring systems.

CNAE aims at "enlarging the comparability between the economic statistics from distinct national sources, and the statistics of the Country in the international plan", providing, thus, Brazil "with an updated classification of economic activities with the changes to the structure and composition of the Brazilian economy and synchronized with the changes introduced in version 4 of the *Clasificación Industrial Internacional Uniforme de todas las Actividades Económicas* – CIIU[20]" (IBGE, 2007). Version 2.0 of CNAE is structured on a graded way (5 levels, with 21 sections, 87 divisions, 285 groups, 673 classes and 1301 subclasses, being that the level of the subclasses was defined for use by the Public Administration). However, CNAE "is a classification per type of economic activity and not a product classification – goods and services" (IBGE, 2007). For the production of product statistics, "IBGE developed and maintains a detailed nomenclature of products for application in researches, called Products List - PRODLIST". There are several PRODLISTs (Agriculture/Fishing, Industry, …) each of them with their particularities. For example, the PRODLIST-Agriculture/Fishing is articulated with NCM (Mercosur Common Nomenclature) and with other international classification standards (CPC – *Central Products Classification*, ICC / FAO 2006, FAO commodities 1990).

However, the differences of criteria to justify the formation of its five levels (Section, Division, Group, Class and Subclass) and the fact that there is no biunivocal correspondence with the CIIU/ISIC rev. 4 codes, make these classification identification rules appropriate for a certain economic activity unity (company) very complicated. Aiming at "enabling the elaboration, inspection and evaluation of public policies on an integrated way", the Brazilian government still uses the NBS (Brazilian Nomenclature for Services, Intangibles and other Operations which Produce Variations in Equity) along with NEBS (Explanatory Notes), a national classifier for the identification of the services and intangibles as Products.

20. Or *International Standard Industrial Classification of All Economic Activities* – ISIC (http://unstats.un.org/unsd/cr/registry/isic-4.asp) is a systematic classification of all the economic activities which seeks for the harmonization between the industrial products classification system of each country so as to allow the countries to produce data according to the internationally comparable categories.

In the search for alternatives for the description of products in the Electronic Business Market, the Brazilian government has considered alternatives of knowledge organization systems in the form or taxonomies, such as GPC, UNSPSC, eCl@ss and CPV. The latest, in particular, has deserved more attention, considering its structural proximity concerning the CNAE. The CPV is a European classification system directed to public contracts with view to the standardization of the references used by the contracting entities in the description of the objects of the contracts. Attström et al. (2012) identified a number of problems in the CPV: *(i)* the coverage is, sometimes too specific; *(ii)* sometimes the coverage is too general; *(iii)* some codes do not correspond to the current market demands; *(iv)* a significant percentage of categories does not have a correspondence in other classification systems (eCl@ss, GPC, UNSPSC), as shown in Table 4.1. For example, in 23% of the cases, there is at least one category in eCl@ss without an available class in CPV. In contrary, in 34% of the cases, there's at least one category in CPV without class available in eCl@ss.

	CPV	eCl@ss	GPC	UNSPSC
CPV	-	23%	13%	33%
eCl@ss	34%	-	14%	35%
GPC	35%	33%	-	37%
UNSPSC	28%	30%	9%	-

Table 4.1: Quantitative comparison of the CPV coverage extension
Source: Attström et al. (2012)

The recommendation for improving the CPF given by Attström et al. (2012) was the following:

- To the roles already offered by systems like TED/SIMAP[21] interactive instruments, such as Web seminars, *wiki* and feedback functionalities must be added. The objective is to provide the exchange of experienced of the portion of society related to the subject and a debate about the ways for the evolution of this classification system.

21. The Système d'Information pour les Marchés Publics (SIMAP) is the main information system directed to public bids in Europe.

- There are several possibilities to integrate the CPV in *e-procurement* environments. The current CPV could be reinforced, and could collaborate with other classification systems. Each possible classification has advantages and disadvantages and a decision between them could be taken.

- An update policy for CPV must be defined and there must be a distinction between higher or lower importance updates. The users must be involved in the maintenance of CPV on a structured way. In addition, a reflection about the normative nature of CPV must be started in order to allow increased flexibility.

These recommendations can also be applied to Brazilian classification systems, CNAE, PRODLIST and NBS, once there is great proximity between these taxonomies and CPV. One of the identified problems is that, given the great coverage of goods and services, there is no consensus about the interpretation of many of the concepts employed in all the government spheres. This misalignment, the smallest as it may be, negatively affects the discussions about the appropriate form to define a product taxonomy, being an obstacle for the interoperability aspirations between product listing systems. In addition, as it will be seen further on, a taxonomic approach leads to other limitations for the expressiveness of production chains.

4.2 An essential concepts base

It is a fact that the aspects involved in a proposal such as the listing of products for the Electronic Business Market are poorly addressed in the technical or scientific literature. Making the producers and consumers in different positions in a production chain to be found more easily is submitted to a critical review of the currently existing proposals for the product classification, predominantly based on a taxonomic structure.

A way to meet the presented requirements is to adopt the conceptual modeling supported by foundational ontologies for the description and classification of the Brazilian production chains products. Although it is yet unknown for many people, this approach has been adopted in several fields of the chemical, manufacturing and natural resources production industry due to its benefits, especially to solve interoperability problems, in addition to represent composition relationships (*part-of*) and views of a product allowing to insert it into different positions of the production chain.

Similar initiatives have been presented in the *Formal Ontologies Meet Industry* (FOMI)[22], annual event which gathers the main government bodies, companies and members of the academic community interested in classification systems for goods and services in the industrial context. In fact, it is noted a consensus that this approach provides a more favorable approach than the taxonomies for sharing knowledge, exchange of reliable information, in addition to making the elements required for the interaction between information systems of distinct companies feasible, in *business-to-business* (B2B)[23] or *business-to-govern* (B2G)[24] contexts (Borgo and Lesmo, 2008).

The foundational ontologies metaproperties discussed before substantially contemplate the presented requirements and guarantee the flexibility of relationships required to meet the trade demands of the micro and small entrepreneurs. Under this perspective, instead of adopting a strict taxonomic structure, the proposal herein is the construction of an ontological conceptual model of the production chain products, called Essential Concepts Base (BCE), for the classification and descriptions of goods and services in the Electronic Business Market context. Figure 4.1 shows BCE with the respective characteristics which must be met.

Thus, BCE must contemplate the configuration or composition of products and, at the same time, be able to absorb other relevant definitions for the electronic commerce, such as the offer of products and other types of functionalities which increase the interoperability between commercial partners. In addition, BCE is being proposed as an instrument which facilitates the dialogue between the involved parties over the products domain, once the models constructed based on a foundational ontology tend to have clearer and more outlined descriptions and classifications.

22. The last edition of this event can be found at http://www.csw.inf.fu-berlin.de/fomi2015/

23. The term B2B (business-to-business) is associated to transactions of products or services between companies. B2B is related to the transactions between companies for the purchase of services, technologies, equipment, components and financial transactions (Paoliello and Furtado, 2004). In the B2B perspective, the electronic commerce facilitates the business applications, benefiting the management of suppliers, stock, distribution, channel and payment.

24. The term B2G (business-to-govern) is associated to the use of the electronic commerce technologies, particularly, to the relationship of the public power with the other economic agents for the government purchases, for example, by means of bidding processes, or the search for efficiency of the public collection systems.

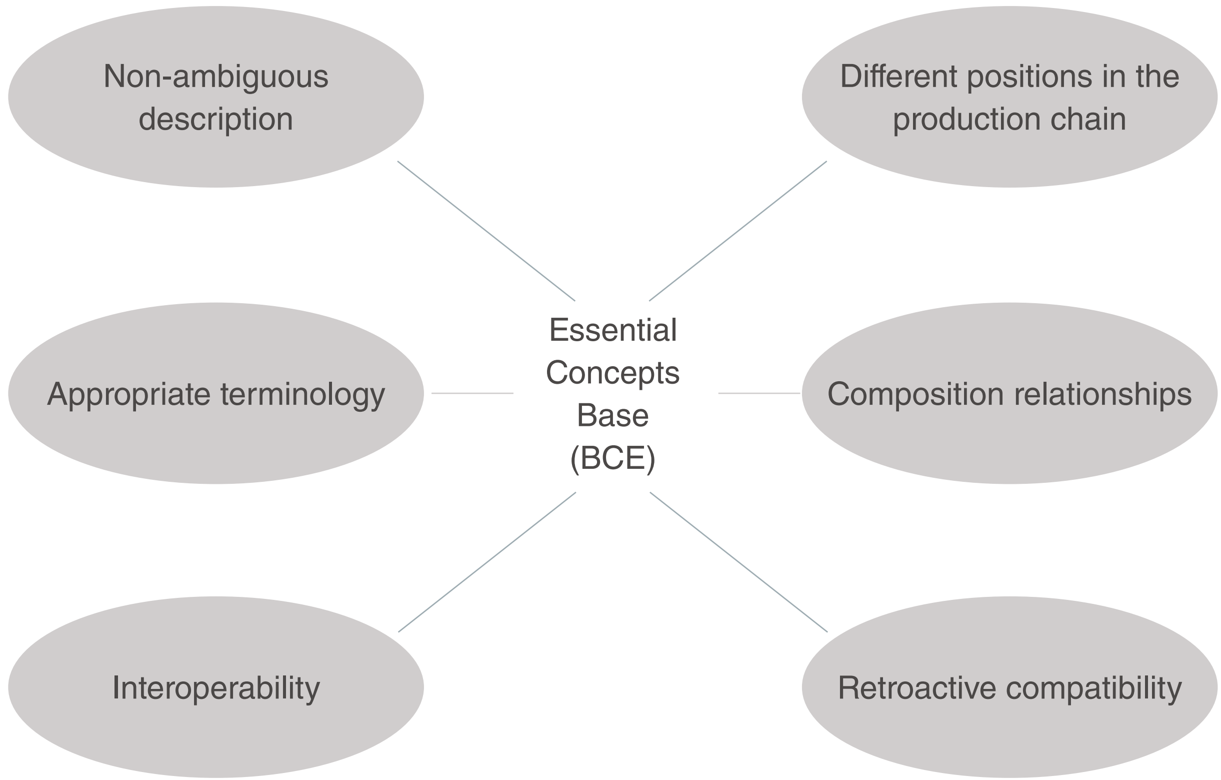


Figure 4.1. Essential Concepts Base: involved requirements

BCE is a description of the production chain in the form of an ontology, which can be seen as a graph. In this graph, each node is a product concept which, in its turn, can have several forms of relationship with other production chain concepts, characterizing a higher flexibility of relationships. With this respect, the composition relationships refer to a product which can be described as a combination of a series of other products from the chain. Due to the implicit flexibility of the relationships of the proposed model, the practical effect is that a product can be positioned in more than one position of the production chain.

BCE solves the requirements problems mentioned because it is a model which supports types of concept relationships which go beyond the specialization relationships, typical of the taxonomies. In this case, products must be described on a non-ambiguous way, making use of the ontological suppositions and the metaproperties inherent to the foundational ontologies. Therefore, BCE designer must conceive descriptions which characterize a product on a univocal way, attributing a sole identifier for each item. This way, products can be described based on its parts and essential characteristics and, at the same time, having varied classification codes for the product, preserving, therefore, the retroactive compatibility requirements or adherence to the taxonomic definitions of interest by the Government (CNAE, PRODLIST, etc.).

Concerning the linguistic differences due, for example, to regionalisms, it is foreseen that BCE concepts are connected to a Thesaurus of Product Terms (TTP), in such a way that a certain product can have varied and compatible denominations with the usual language of the entrepreneur, regardless of the region where the product is being offered. This guarantees that, in a negotiation process, the parties have the clarity about the trade object at the same time on which the desired interoperability is met. I.e., associated to the sole identifier of the product concept, terms from a thesaurus which must be foreseen in the project in order to five different denominations to the same product in order to guarantee the equalization of the regional terms about the same product must be linked.

One of the necessary requirements in this project is that BCE product description takes into consideration the traditional product codings. In order to meet this demand, a retroactive compatibility is foreseen, where each describe product contains, among its attributes, the alternative codes of the desired taxonomies, such as CNAE, PRODLIST and others.

The interoperability is one of the essential requirements to make the product negotiation feasible, particularly in distributed environments where the entrepreneurs can be using different computer platforms, too. BCE, due to its inherent characteristics, and for being a complete ontology of products, (and in continuous maintenance to guarantee such completeness), tends to be the aggregating element which enables the dialogue between distinct parties. As this specification tends to be used for the description of the entire Brazilian products chain, it is reasonable to think that there will be compatibility of definitions and, therefore, of interoperability between commercial partners in Brazil. In its turn, the interoperability with suppliers and external parties can be followed by aggregating other codes, such as CPV, UNSPSC or eCl@ass, to BCE concepts.

However, for this requirement to be reached, it is important to have a political and administrative positioning by SMPE in order to guarantee a standardization of description about products and adoption of BCE (or a variation of it) as a reference. In other words, in addition to the effort of designing, constructing and maintaining BCE, it is necessary an articulation in order to sensitize the concerned community for the adoption of a single structure to product description.

In addition to the benefits mentioned for BCE, the strategy of adopting the conceptual modeling supported by foundational ontologies produces additional relevant results: *(i)* for the generation of relational databases, *(ii)* for the code of applications, and *(iii)* for the specification of domain ontologies described in OWL, a Semantic Web standard which admits inferences and the acquisition of new knowledge, as represented in Figure 4.2.

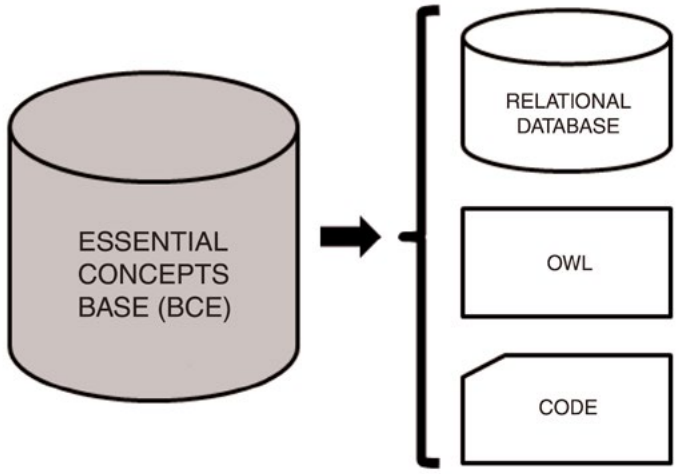


Figure 4.2. Essential Concepts Base: implementation benefits

When the conceptual modeling supported by foundational ontologies is compared with respect to the specifications made with the Entity-Relationship model (Chen, 1976), the literature presents evidences of gains compared to the computer structures involved[25]. MER, for example, assumes that the entity and relationship concepts are sufficient to represent the interest reality for an information system, but several elements which would be useful for the representation of complex dependencies between the domain objects and values.

The same way, applications which manipulate these data tend to be leaner and more modular and, for this reason, the conceptual modeling based on ontologies can be seen as a good Software Engineering tools (Calero et al., 2006).

25. In relational models (ER), the entities (classes) and relationships give origin to the physical model tables. This approach generates a considerably higher number of tables when compared to the conceptual modeling supported by ontologies.

5. Methodology for the development of the essential concepts base

In order to serve the electronic commerce in the Web context, the conceptual modeling supported by ontologies (BCE) needs to be done considering, among other things, the product creation dynamism, the variability present in the production chain and the endless range of details about each of them. For this reason, the following needs are identified:

- A representation language appropriate for the development of this conceptual model;
- An definition of a team of *experts* in the subject in order to guarantee the growth and adaptation of such specification in order to make it the most complete one in the shortest possible time;
- A work methodology in order to reach a good conceptual model.

Concerning the language to be used in BCE description, it is recommended the use of OntoUML (Guizzardi, 2005) to describe the conceptual model, once this is an essentially ontological language which, although there are restrictions for the effect of computing and inference, it is appropriate for the description of the products with the due detailing degree. In addition, the OntoUML contains a notation which favors the dialogue between the pairs and, most important, it has been adopted in other similar initiatives to this project to the Federal Government[26].

The conceptual modeling *expert* must produce a BCE so that it can provide meaning to the definitions about configuration and product offer. He/she is a technician or researcher with domain on philosophical concepts and on the category systems used in foundational ontologies, such as the all-part relationships, events, formal and material relationships, depending entities and value spaces, required to define BCE and its connection with the term thesauri and with taxonomies.

With this respect, some more relevant work groups were identified, which can be contacted in order to perform this activity. They are: *(i)* the Conceptual Modeling and Ontology Studies Center[27], of the Federal University of Espírito Santo, coordinated by Prof. Giancarlo Guizzardi, *(ii)* the Laboratory for Applied Ontology (LOA)[28], of the Italian National Research Council (CNR), in Trento, Italy, coordinated by Nicola Guarino, and *(iii)* Ontology Engineering Group (OEG)[29], of the Universidad Politécnica de Madrid, directed by Prof. Asunción Gómez-Pérez.

Concerning the methodology to be adopted to obtain BCE, the suggested approach is not start from zero to define the model concepts, but from existing taxonomies in order to facilitate the connection of the relevant definitions. I.e., the proposal is to generate BCE by means of a process of conversion of one or more existing taxonomies for a foundational ontology model. Figure 5.1 indicates a possible way for obtaining the ontological model of products from a view of using foundational ontology.

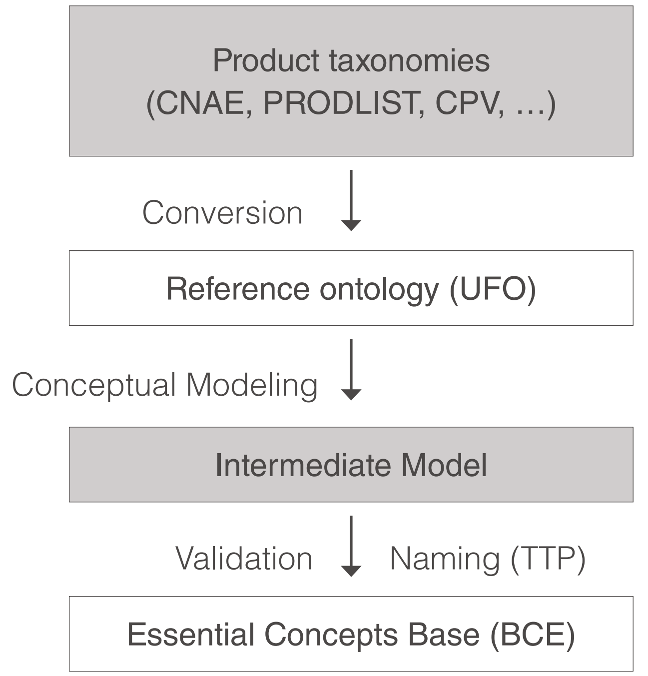


Figure 5.1. Methodology proposal for the construction of the Essential Concepts Base

26. See ontologies available in the Federal Government vocabulary repository: http://vocab.e.gov.br.

27. http://nemo.inf.ufes.br
28. http://www.loa.istc.cnr.it/
29. http://dia.fi.upm.es/

In this figure, the conversion phase refers to the process of choosing in-use taxonomies such as CPV, CNAE and PRODLIST and its consequent cleaning (analysis under the view of essential metaproperties of foundational ontologies) in order to convert them for a representation which uses a category system associated to the definitions from the original taxonomies, adding them with new representation forms. For this case, there are already appropriate tools, such as OntoClean (Guarino and Welty, 2000a, 2000b, 2000c; Masolo et al., 2003; Guizzardi, 2005), which is a methodology already consolidated in the literature and which allows the generation of conceptual models from taxonomic structures. Therefore, BCE adaptability to a taxonomy must be submitted to a methodological evaluation and the result of this approach is a re-design of the original taxonomy added with the philosophical suppositions of a foundational ontology (refers to the UFO - *Unified Foundational Ontology* box, of Figure 5.1).

The following step is the conceptual modeling of the products from the UFO in an ontological language, such as the case of OntoUML, in order to produce a foundational ontology about products, called here as intermediate model once it does not have validations of links with product thesauri. However, although still in intermediate stage, this result can be used by humans in tasks such as communication, domain analysis, negotiation of meaning, establishment of consensuses and problem solutions.

The reliability of the intermediate model of Figure 5.1 can be measured with the use of testing tools such as the free software Isabelle (Nipkow et al., 2002), an interactive tester associated to a set of efficient functionalities of automatic logical reasoning. A number of important Mathematics and Computer Science theorems were formalized with the help of this tool.

Another validation mechanisms subject to use in this context is Alloy (Jackson, 2006), which contemplates a formal structure description language and a tool to explore them. Alloy has been used in a wide range of applications seeking for conceptual inconsistencies in software and hardware projects. Both Isabelle and Alloy have been successfully used in the formal validation of ontologies by research centers, among them the Conceptual Modeling and Ontology Studies Center (NEMO), of the Federal University of Espírito Santo, one of the groups suggested in this document in order to conduct the conceptual modeling work of the Brazilian production chains.

After the validation process, the intermediate model also passes through a process for naming the concepts of each product, which involves the attribution of a single identification to each product, considering the connections of the concepts with the listing system codes of products of interest such as CNAI, PRODLIST, etc. Another measure in the methodology is to link the concept of the intermediate model to the terms of a Thesaurus of Product Terms (TTP), specifically conceived to serve the marketable products chain.

Therefore, BCE is the intermediate model in OntoUML duly validated and connected to TTP. It is further important to remember that the TTP deserves special attention, once it is an expensive activity and which demands addition development efforts and adaptation to BCE. In fact, thesauri in general are structures designed by specialists in the Information Science area and this concept also requires a considerable work time until a compatible version with BCE concepts are contemplated. As BCE and TTP are dynamic structures, it is foreseen the need for a maintenance team for these structures until a desirable version is reached for use in the Electronic Business Market..

6. Characterization of the essential concepts base

As previously discussed, the product description must consider an ontological perspective with relationships which are more complex than those provided by the taxonomies. For example, Figure 6.1 shows a product, in this case a printer, the classification of which was performed considering its role, i.e., an equipment that is essentially used in offices. However, a printer has an implicit concept and this does not have a necessary relationship with a work office. Thus, the same printer could be classified as a component from a larger equipment, such as lathe which had a printing module supplied by the same printer. Therefore, the separation of the essential concept of a product with respect to its functionality is important to allow the product to be declared just once (without redundancies) and, at the same time, it can be linked to different portions of the production chain, guaranteeing increased product positioning flexibility.

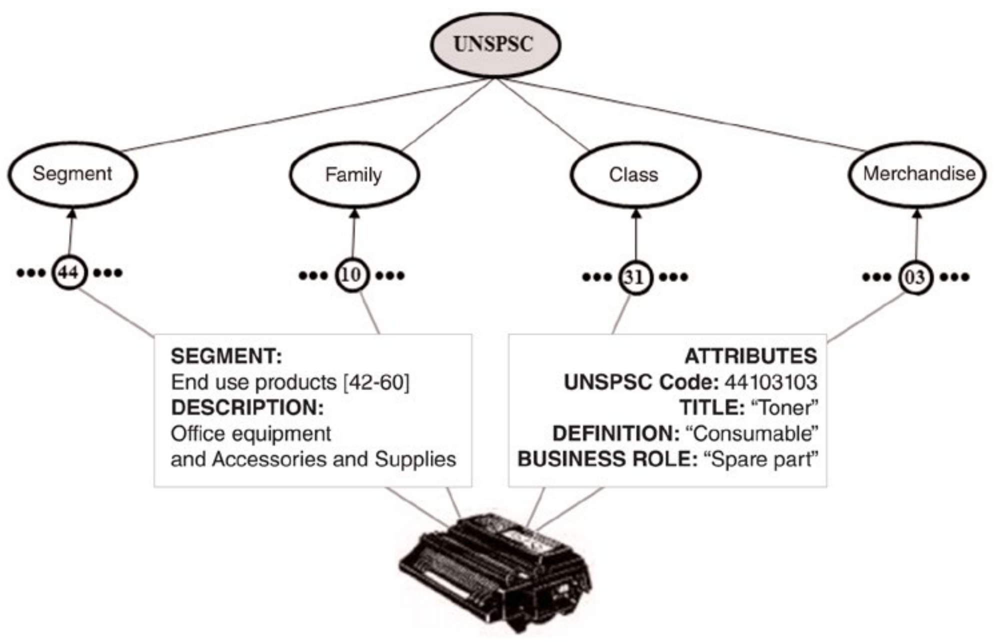


Figure 6.1: Example of product classified with a taxonomic bias
Source: Adapted from the European Committee for Standardization (2010)

BCE seeks to solve this problem and meet the previously listed requirements through the proposal of a product description supported by foundational ontologies, with varied concepts and relationships which provide the flexibility and adaptability required for the products to be composed on a varied way. In a more specific character, BCE is formed by a series of definitions about the concepts of the origin production chain (CNAE, PRODLIST, etc.), where each product concept is given based on its raw materials and components. The raw materials or components, in their turn, are also described on a recursive way by their respective raw materials and parts so as to meet all the constituent elements of the production chain which can be marketed in some measure.

An example of supply for the manufacturing of a product can be seen in Figure 6.2, where four products are presented (cotton, line, cloth and shirt) which are in different positions in the production chain and which can be marketed by different MPEs. While the connection between who makes the offer and who needs the product is not possible in the traditional taxonomic classification (in this case, UNSPSC), BCE prioritizes this type of relationship exactly because it allows the meeting between pairs in order to enable the trade of these raw materials. Due to its importance, the *input-of* relationship of the figure must be present in the product's conceptual model (BCE).

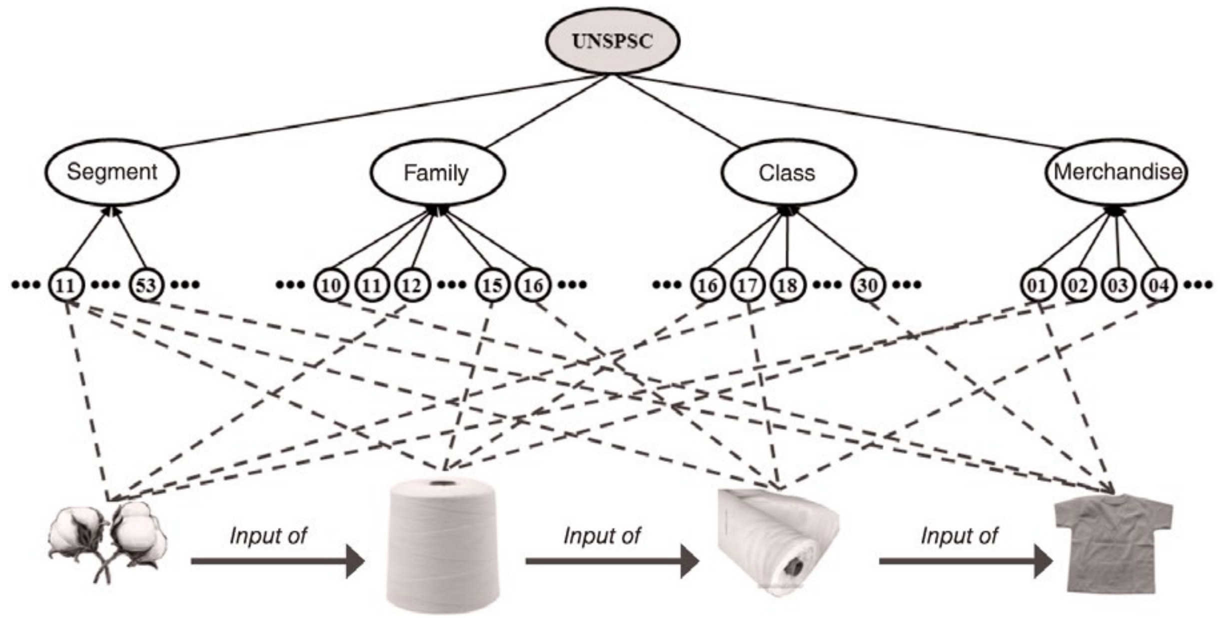


Figure 6.2 Configuration of products from the input-of relationship

Another type of relationship absent in the traditional taxonomies and which potentializes the commercial relationships is the *part-of* relationship on which a product is described as the sum of juxtaposed parts. Thus, the entrepreneur who trades a certain product can automatically identify potential commercial partners who have registered the offer of products which are part of its composition. A clear example of this approach is represented in Figure 6.3, where a bicycle is defined as a product composed by a juxtaposition of other products. It is noted, with this approach, that the *part-of* relationship is the greatest responsible for the possibility of meeting between the commercial partners who sell both the largest product (bicycle), and the products which are part of the configuration (ratchet, tire, frame, etc.).

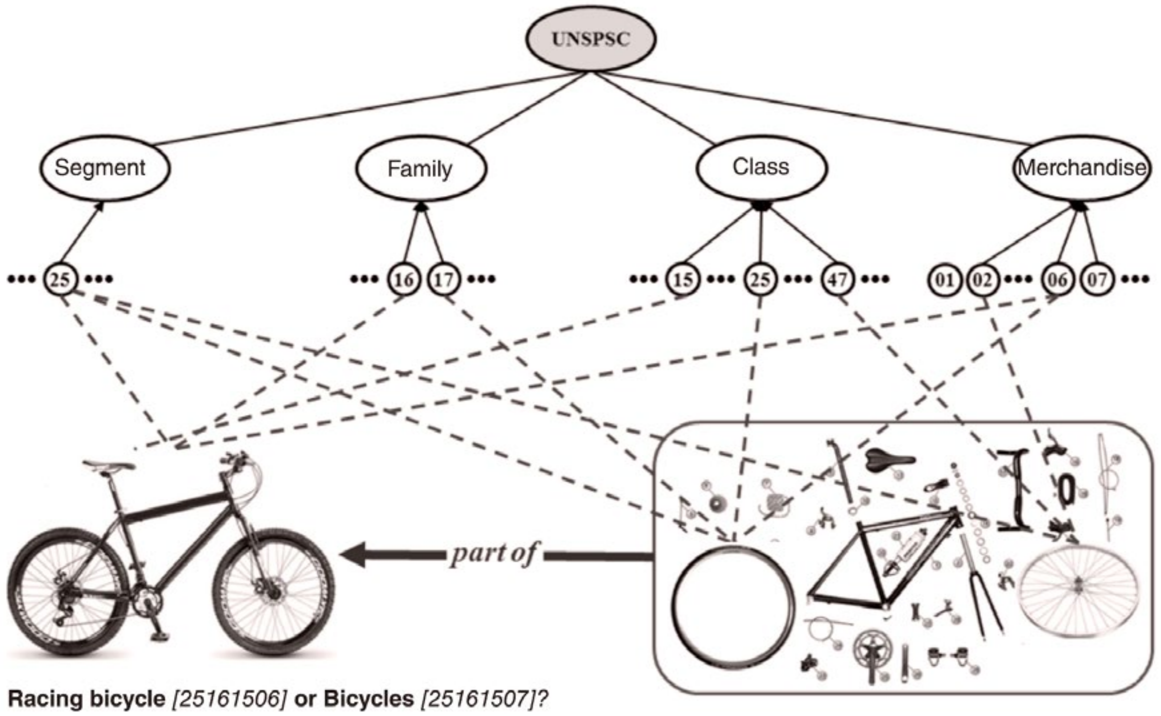


Figure 6.3 Configuration of products by using the *part-of* relationship

Due to the importance of the presented composition relationships, the description of products and their raw materials or parts is one of the greatest characteristics of BCE. After all, any product must be described based on its parts and this work involves a considerable effort by assuming this approach for all the elements of the Brazilian chains of production, supply chains or value chains.

A procedure suggestion to describe these products in BCE is shown in Figure 6.4 and refers to the process of including new product specifications in BCE. In this case, this new product can be derived from a previous definition which is present in the base, or is conceived from zero, with no previous relationship with other products from the concept base. At first, the derivation relationships can be of *(i)* specialization, when the new product is being conceived from an existing product, or *(ii)* of composition, when the new product is conceive by the sum of the parts of products already existing in the base.

In the figure, it is assumed that an *expert* interacts with the system searching for some definition of concept which is already present in the products base. If the product is not found, it must be included by a recursive action of inclusion and description of its parts. If the product was found by the search process, and already contains an established description, the next step is to verify if there is any change to be performed in the referred product.

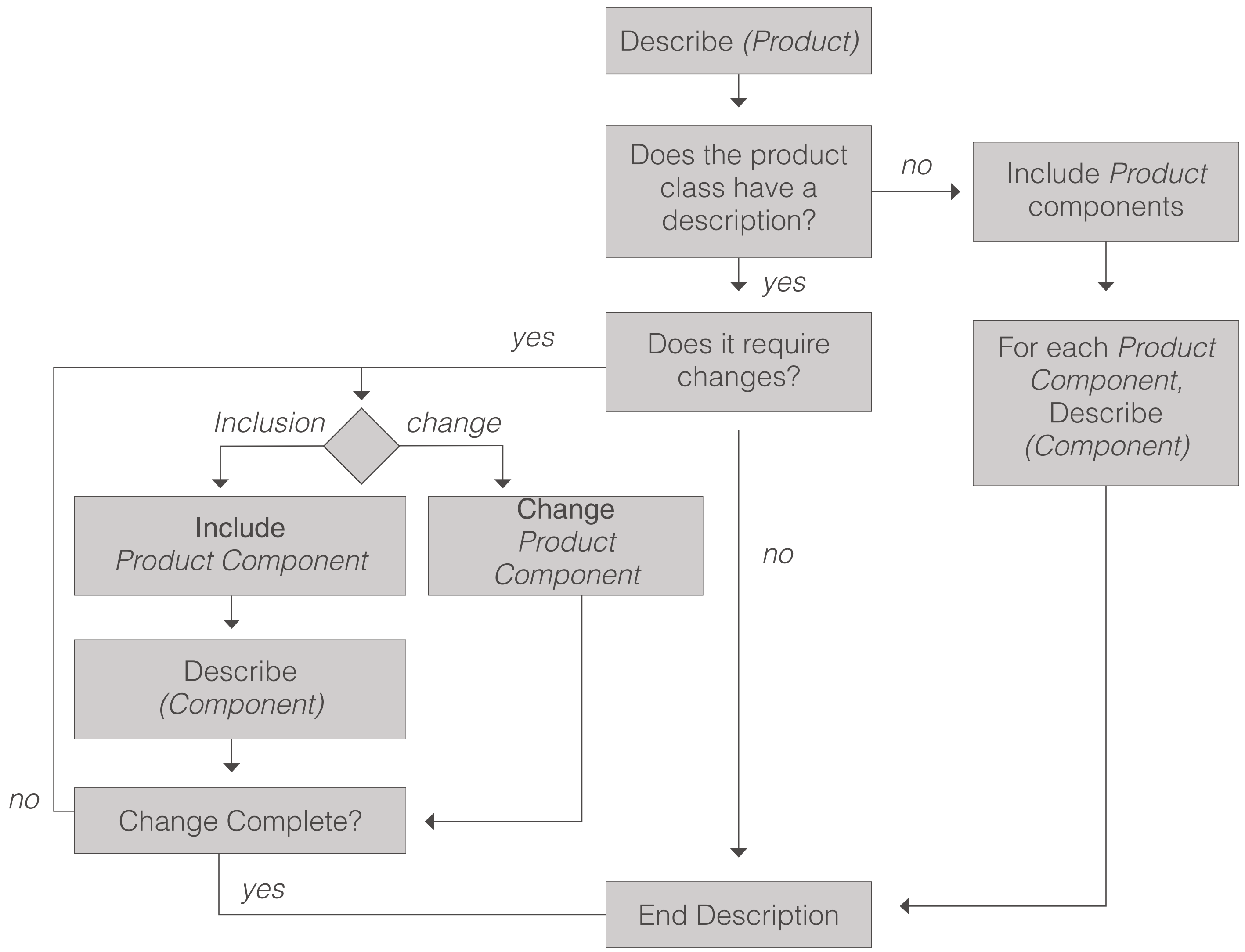


Figure 6.4 Suggestion of product description logics

In case there is no new description to be performed, it is considered the product description found to compose a product registration and the description process is ended. The eventual changes to a product description involve inclusion and change. In the first case, the inclusion of a product component and its consequent description (attribution of the classifier code and description text, for example) occur. Supposing that a product can be formed by at least one component, the process for the inclusion of new components is repeated until the product is completely configured by the sum of its parts, according to the recursion shown in the figure. If the change of the main product which was being initially conceived involves only the change to one of its components, such change is performed and the new derived product is recorded in the base, remembering that the part change process must admit only those which do not strip the conceptual definition of this product of its original characteristics.

7. The essential concepts base and the semantic web

One of the essential requirements to be met is that producers and consumers in different chain of production, supply chain or value chain can be found more easily in order to enable the objectives of the Electronic Business Market. On the other hand, it is know that the traditional Web does not favor this type of approach, because the existing electronic commerce sites form data silos with relative isolation, making goods and services difficult to be found by the users with the desired facility. In addition, the searches based on syntactic comparison promoted by the search drivers hamper a unified view on prices and conditions about goods and services.

Therefore, although it is possible to use BCEs in conventional applications, it is suggested that the adoption of the Semantic Web principles, more compatible with the interoperability and homogeneity requirements desired in the context of the Electronic Business Market. In the semantic Web, it is assumed that the ontologies and the data, described, as recommended by the World Wide Web Consortium (W3C)[30], in OWL[31] and RDF[32] format, respectively, are considered as essential pillars to allow the desired interoperability. In other words, it is defended here the idea that BCE's definition in ontological language created by the expert are converted into the OWL format in order to guarantee the benefits of the Semantic Web[33], as shown in Figure 7.1.

30. https://www.w3.org

31. https://www.w3.org/2001/sw/wiki/OWL

32. https://www.w3.org/RDF/

33. It is interesting to emphasize that the editor OntoUML, already mentioned, has a direct conversion module of conceptual models for the OWL format, in version 2.0, and it also allows the conduction of proving tests.

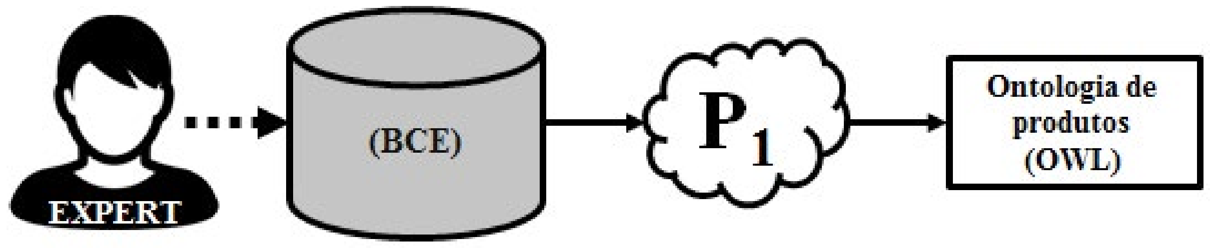


Figure 7.1. Conversion of BCE in OWL format

In this figure, process P_1 refers to an engineering suggestion for the construction of an ontology by using the previously mentioned tool OntoUML which contains functionalities which allow the conversion of BCE to the OWL format. Concerning the data or instance of products described in RDF format, there are some possible alternatives. One of them is the use of the heuristics implicit in tool PCS2OWL[34] (Stolz et al., 2014) in order to obtain RDF triples directly from the conceptual model, as shown in Figure 7.2. In this case, the description of the products in RDF format is obtained by the processes of: *(i)* validation – conversion of the traditional classification systems (eCl@ss, CPV etc.) in OWL format, *(ii)* transformation – obtainment of the individuals from the OWL format, and *(iii)* serialization – description of the individuals in one of the several RDF representations available (in this case, RDF/XML). In the case of this proposal, this tool can be adapted to use only the described transformation and serialization processes, once the OWL modeling has already been obtained by already previously described processes.

34. PCS2OWL is a generic method and a set of tools to derive valid OWL ontologies from product classification standards and catalog group standards. The resulting ontologies are compatible with the GoodRelations ontology, presented further on in this document, and can be used to enrich the product model data and provide information in Semantic Web systems. Currently there is PCS2OWL support for the main product classification systems (http://www.ebusiness-unibw.org/ontologies/pcs2owl/).

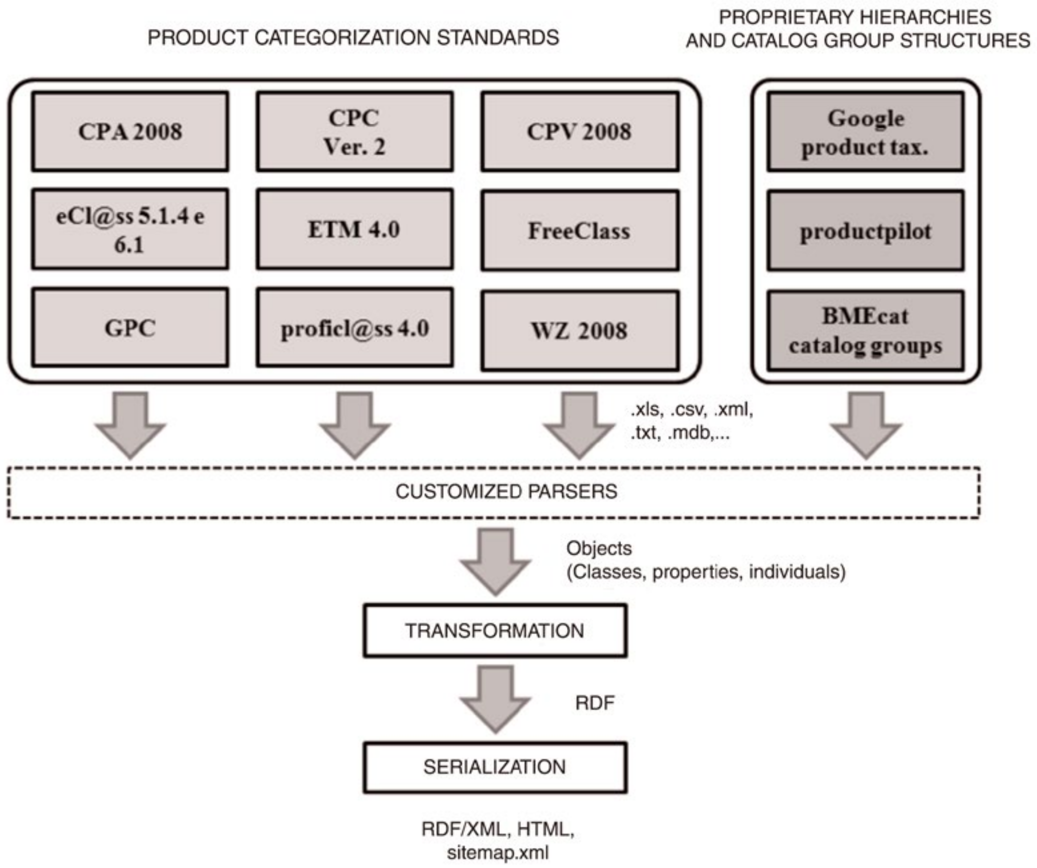


Figure 7.2 Scheme for obtaining RDF model by using the PCS2OWL tool
Source: Stolz et al. (2014)

Another possibility is to obtain the products configuration from the products description made by the final user himself/herself (the entrepreneur), as presented in Figure 7.3. In this figure, while process P_1 refers to the conversion of BCE into OWL format, process P_2 is an automated (or semi-automated) process suggestion intended to obtain information about the products (goods and/or services) configuration/composition from those descriptions made by the entrepreneurs. In this type of application, in general, the entrepreneurs must have conditions to describe his/her products, considering its natural limitations which must be resolved with technological support (friendly interfaces and software resources allowing an interaction compatible with the type of user).

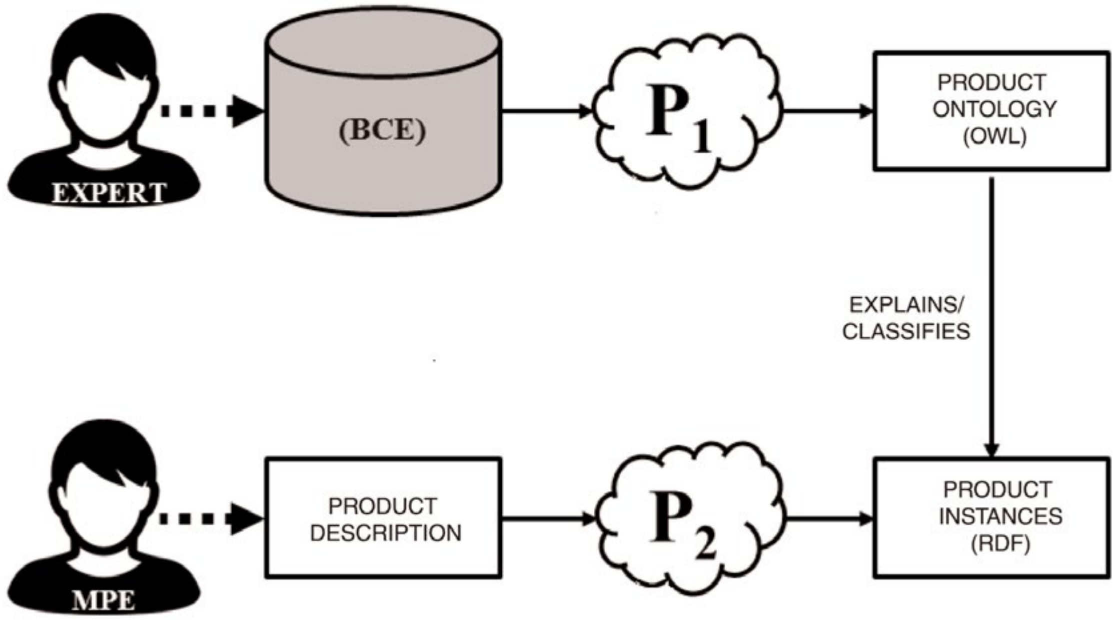


Figure 7.3. Essential concepts base and instances of OWL/RDF products

One of the alternatives to receive the specifications from the entrepreneur is to allow him/her to construct a descriptive text and the consequent use of Natural Language Processing (PLN) techniques, implicit in process P_2 of Figure 7.3, in order to generate the data structure in RDF format. Another possibility to receive the products specification from the entrepreneur is to allow him/her to access BCE defining elements (browsing by the products ontology) in order to produce RDF instances from this browsing process[35].

It is important to emphasize that the description logics of a product must be followed by the expert in BCE definition, but it must also be considered in the processing of the descriptions made by the entrepreneurs (process P_2 of Figure 7.3). In the latest case, due to the PLN strategy used, it can be required that the results obtained are evaluated by a curatorship process, what would require the organization of an infrastructure to be organized by SMPE to solve this type of demand. In one way or another, the ontological categories system built-in in the conceptual modeling helps the user to identify if what is being conceived is, in fact, a new product or if it is a feasible change

35. It is assumed that the product ontology browsing (from BCE) must be conducted with the help of some appropriate graphic interface in order to facilitate the interaction with the entrepreneur.

to an existing product. For example, assuming that in the base there is previously configured product (peanut, for example), and the user wants to define a new product (in this case, roasted peanuts). If a new product is a derivation of an existing product (as in the example), the underlying data model must reflect this reality, preventing record redundancies in the databank. However, these product variations can continue to receive taxonomic codes with the advantage that this codification process can be supported by ontological inferences, facilitating the identification where a certain product must be positioned in the taxonomic chain in use.

The suggestion made for process P_2 supposes certain processing complexity and largely requires that BCE is as complete as possible so that the inferences and suggestions can be indicated by the involved reasoning drivers. As the construction of BCE involves a significant time, it is important to consider alternative development plans in order to enable solutions that are useable by the entrepreneurs in phases before the completion of the full BCE design. Anyway, the final design of the description information about products would be reflected in a base composed by an OWL file of concepts about product configuration/composition and a set of data about products from the Brazilian production chains in RDF format.

The OWL format suggested for BCE: *(i)* is compatible with the Semantic Web principles, *(ii)* is more appropriate to describe the meaning of the products base relationships in RDF format, and *(iii)* facilitates the process for reusing public domain ontologies, such as the case of GoodRelations (Hepp, 2008), mentioned further on in this document.

7.1 Product offer in electronic commerce

Electronic commerce applications in the traditional Web are normally directed to the products offer, involving, here, an additional discussion to the products configuration/ composition theme implicit in BCE. In fact, the offer can be understood as a description about a public commitment of the offerer (i.e. the prospective supplier) with respect to the consumers. It represents, therefore, a plan or intention of producers or sellers, and not the sale itself. A possibility of representation of the offer is shown in Figure 7.4, which considers the following basic offer structure: for a certain value, a company promises (commits itself to deliver) a certain product (good or service) to a customer, directly of by means of an intermediate (commercial representative).

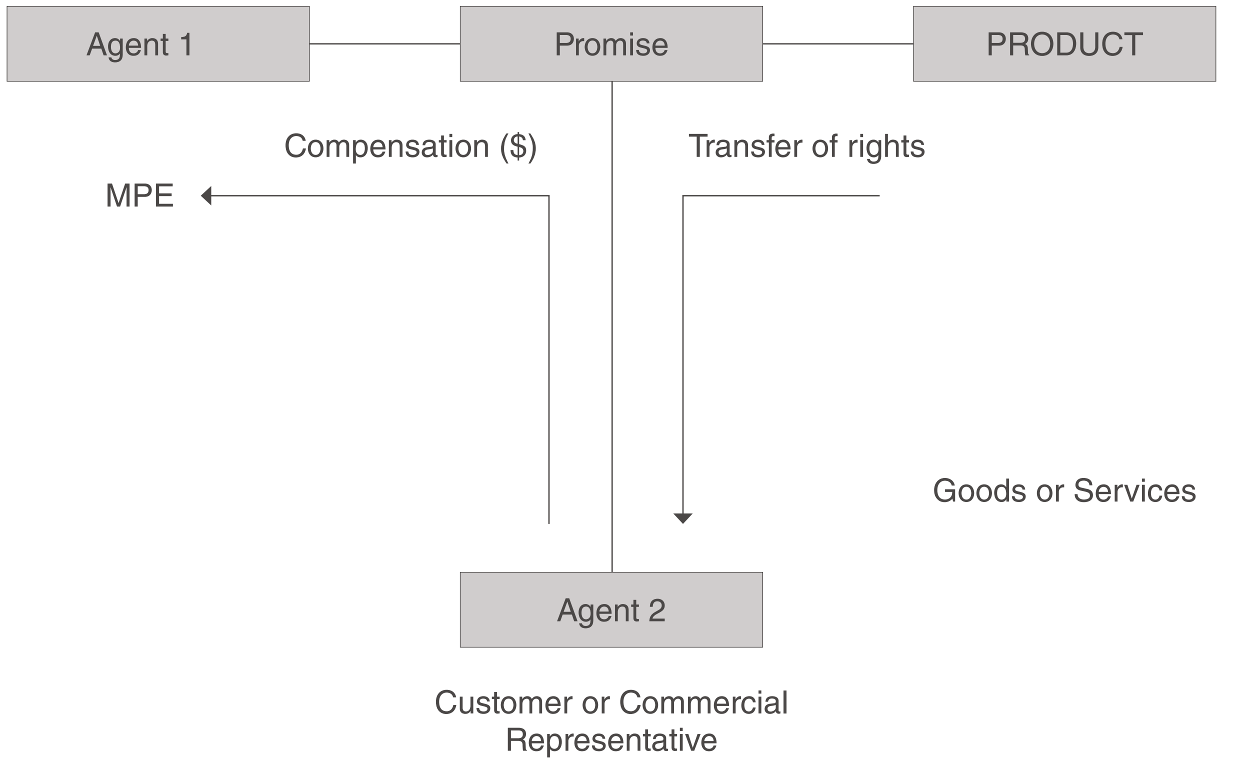


Figure 7.4. Basic offer structure: the Agent-Promise-Object principle
Source: Adapted from Hepp (2010a)

In this context, for any commerce (traditional or electronic), a minimum scenario for the offer of products involves an offering company, points of sale, operating time and payment options (Hepp, 2010a). For specific products, this minimum scenario would further involve the variety of offered products, the business functions (process or operation conducted on a routine basis to perform a part of an organization mission), regions or types of eligible consumers and delivery options. A more comprehensive scenario could further involve the offer of individualized products, with specific characteristics of prices and discounts of availability.

In general lines, the electronic commerce systems contemplate elements such as: *(i)* different types of businesses (purchase, sale, rental,…), *(ii)* specification of product packages in combination with any types of measurement units, *(iii)* price specifications, both for individual values and for price intervals, due to the type of customer, his/her geographic location or the desired amount, and *(iv)* flexibilization for the delivery or payment form. In the Semantic Web perspective, these elements are represented in the form of an ontology. An example of ontology in this context is GoodRelations (Hepp, 2008), an ontology which contemplates the mentioned functionalities. GoodRelations is based on Semantic Web standards and is compatible with standardized ontologies and market taxonomies such as eCl@ss.

The product offer specification with GoodRelations includes the following information (e-Business + Web Science Research Group, 2008): *(i)* the offering company, being requested specific information (URI of its main page, name, address, telephone and electronic mailing), of the company and its point(s) of sale, *(ii)* the product being offered, when a textual product description is made, is included in a specific URI for the product and is associated to the types of transactions (sale, leasing, repair, ...) possible for the range of concerned products, *(iii)* the offer validity, where the period, the region, types of partnership (final user, companies, resellers, public service) are defined, and, if necessary, *(iv)* payment options and *(v)* forms of delivery. Finally, the automatic generation of the corresponding RDF code is possible.

For the purposes of illustration, Figure 7.5 shows (in the graphic form), an offer of a product described with GoodRelations. In this case, the product is a 30 cm screen TV set which costs 200 Euros, offered by company Electronics.com. In the figure, the elements in blue represent concepts from the database related to the products offered by the offerer, in green are the element from a product ontology (in the example, called toy) and the elements in yellow are related to the GoodRelations structure.

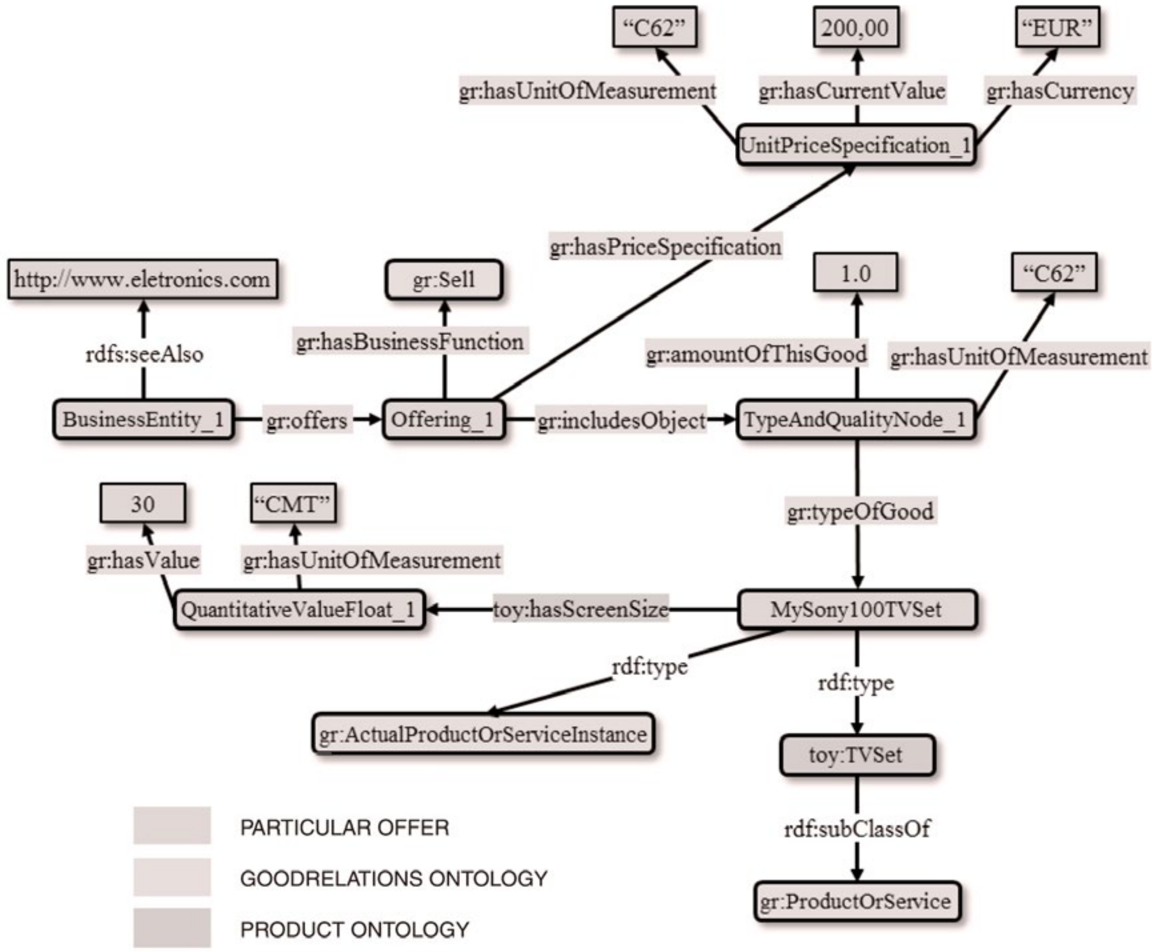


Figure 7.5. An OWL scheme for electronic commerce
Source: e-Business + Web Science Research Group (2008)

7.2 Suggestion for product offer by using BCE

The offer of products in electronic commerce requires that the products are presented to the consumer with an appropriate language. This type of presentation normally involves a simplified description of the products in the text form, including the code, the name and other product details and illustrative images, which can be informed by the entrepreneur himself/herself. This is the most popular case and has been adopted by a number of electronic commerce platforms and cloud computing services available in the Web. In this scenario, some electronic commerce platforms and the search drivers themselves such as Google and Yahoo have invested in the use of ontologies in order to facilitate the location of the offered products for these conventional systems.

Although the mentioned approach is interesting and has improved the searches for offered products in the Web, for the purposes of SMPE it is important that such offer is linked with BCE products configuration/compositions in order to enable the Electronic business Market objectives. On the other hand, if this offer is conducted in the Semantic Web context, the offer itself is also described by an ontology in OWL format and the information about trade can also be seen as RDF triples. Considering this perspective, the proposal of a trade model is presented in Figure 7.6.

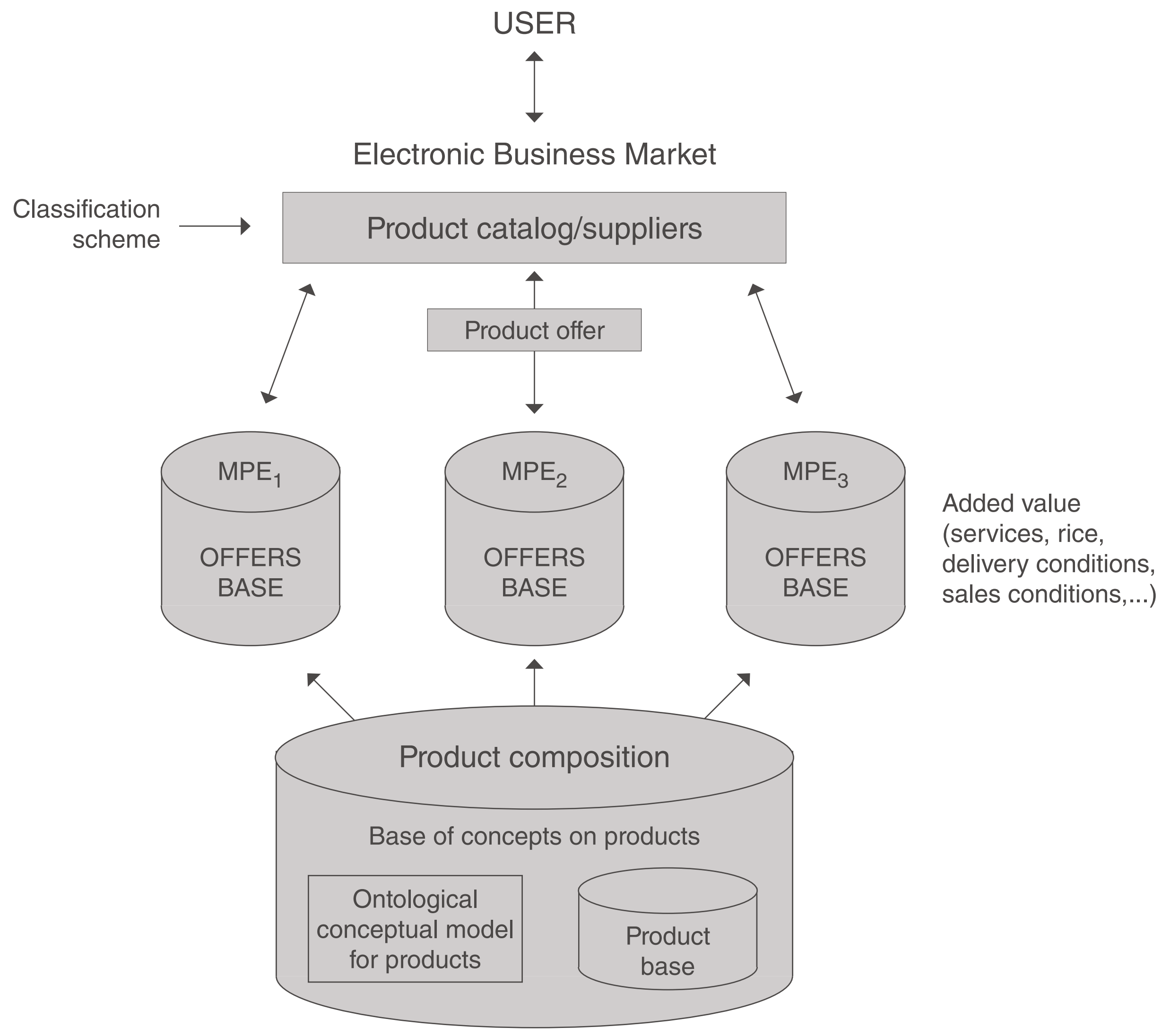


Figure 7.6. Offer suggestion of products from the Essential Concepts Base

In this model, it is considered that the information about configuration and product offer are provided in pairs, in the OWL/RDF format. In this case, these elements can be manipulates for the construction of a typical application of Semantic Web with the use of inference drives for decision making based on OWL specifications, using a first order logics or any alternative which facilitates the decision making about the product offer and configuration. Therefore, instead of having the offer of products described in text format, the proposal here is that the description is from BCE in order to allow the interested parties to meet more easily and promote and interoperability environment favorable for the project objectives.

It is suggested that the user can interact with the Electronic business Market by means of a portal. This, in its turn, contains a product catalogue which, in practice, is an interface for the OWL ontology where the offered products are made available and easily located by using a graphic interface allowing the browsing guided by the terms of these or the thesaurus of terms linked to it. In the same way, this application may contain a special databank, called triplestore, intended to store RDF triples, and an interface for the SPARQL language, which allows the conduction of consultations which may involve other databanks in addition to the products triplestore. In this practice, such technologies can be used in the Electronic Business Market portal and allow a typical dialogue from a virtual store where the user (final consumer or entrepreneur) directly interacts with an environment of products and services to perform the purchases.

For the effect of constructing the portal application in the context of the Semantic Web, it is suggested the use of specific libraries, as in the case of Jena[36], a Java interface with a series of functionalities (shown in Figure 7.7) allowing to process RDF triples and generate processings which include gains related to: *(i)* product location, *(ii)* inference of product codes from the OWL specification, *(iii)* suggestion of potential markets for products and other functionalities directed to the virtual purchase and sales market.

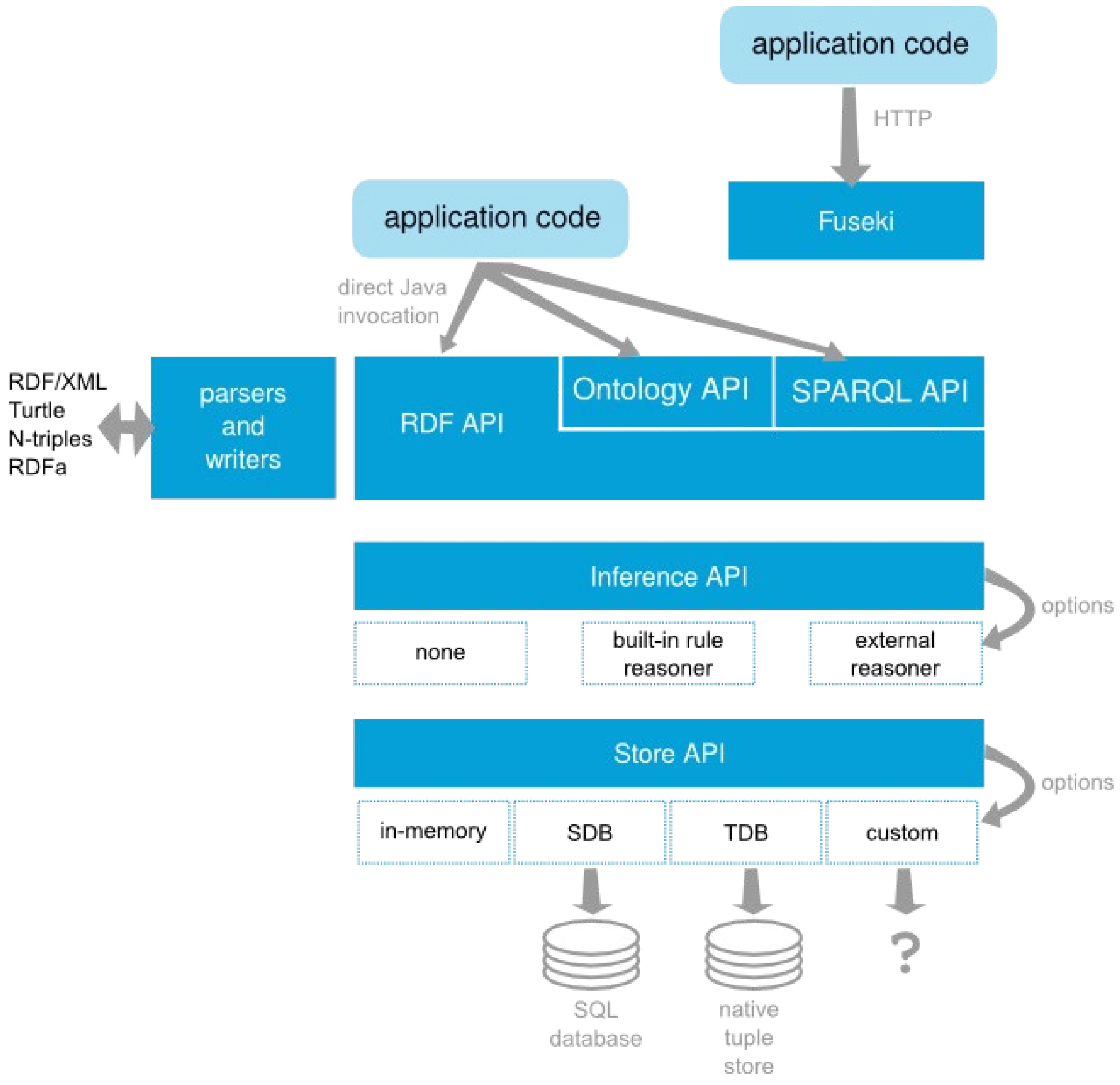


Figure 7.7. API Jena for the development of Semantic Web applications
Source: https://jena.apache.org/getting_started/

36. https://jena.apache.org/

It is further suggested that the portal contains a catalogue of product offers from the entrepreneurs and that these products are described by using vocabularies (ontologies) which guarantee their positioning in appropriate markets, with easy location, also by common Web browsers (Google, Yahoo, ...). In this case, it is recommended that each product has its description performed according to *schema.org*[37], which gathers a number of ontologies (including GoodRelations, suggested in this document) and which is used by these browsers as a search strategy and content recovery.

It is important to note that, in a first moment, it should be foreseen that the negotiations are directly conducted in the Electronic Business Market, where all the entrepreneurs should record their offers of products. However, the conception by BCE and the adopted model admits a migration to context B2B on which MPEs can have their own sites directly interacting with SMPE by means of communication protocols directed for this purpose or access interfaces to the services, as in case of REST[38].

As previously mentioned, the product specifications by the entrepreneur are not only related to a public offer commitment, but also over the configuration elements which define the product based on its parts and the essential characteristics implicit in the essential concepts base and in the corresponding OWL. In this case, the product description process must be preceded by a profile record of the entrepreneur himself/herself in the Electronic Business Market so that the products have a property link with who specified them.

Therefore, it can be said that the offer base is composed by several specifications made by each entrepreneur, which, in their turn, are derived from RDF/OWL product base, presented in the lower part of Figure 7.6. For this reason, it is important that BCE and the equivalent OWL is always updated by the experts team (Figure 7.3) with sufficient definitions to allow meeting the entrepreneurs' demands.

It is interesting to note that, in the product description process, the system can not only infer its parts (previously recorded in BCE), but it can also indicate potential suppliers of these parts. I.e., if an entrepreneur registered himself/herself as a supplier of a certain product, another entrepreneur who registers another product using that product as a raw material automatically receives an indication that the first entrepreneur is a potential supplier for him/her. This way, each entrepreneur composes his/her potential suppliers bank as he/she registers his/her products and this result occurs without the need for an explicit search. This is just one of the additional forms the system admits, in addition to the traditional product and supplier localization forms practiced in the Web.

Therefore, the portal associated to BCE suggested in this proposal may include total gains, such as: *(i)* products are presented in contexts where there is higher negotiation probability, *(ii)* the products can be described in different ways (meets the last mentioned requirement), *(iii)* the offered products can be located by the Web browsers (*search engines*), *(iv)* the solution admits that one or more taxonomic classifiers (traditional products classification systems, such as CNAE, PRODLIST or CPV) can be used to identify the products, and *(v)* the listing process allows to reach the same object through different knowledge organization systems.

7.3. A practical example

Although the conceptual modeling with the use of an ontological level language has been proposed as an essential requirement for product classification, the Semantic Web scenario is considered as an interesting environment for the electronic commerce supported by BCE. In this case, the domain descriptions are performed using standardized epistemological languages, as in case of OWL, which enable interoperability and, at the same time, allow the computer to make inferences about the described model.

With this respect, below a simple (and picturesque) example intended to show how the ontologies can be used in environments such as the Electronic Business Market portal, considering a system directed to the Semantic Web. In order to understand the example, it is assumed here that BCE and the Products Bases must have a refinement degree above the elementary.

Let us consider a pizza restaurant in São Paulo which provides a special *paçoca*[39] pizza. As well as it had done for its other products, by configuring this product in particular in a Web environment for the purposes of offer, the owner described it as follows:

1. **Company**
   Website: www.pizzariadaesquina.com.br

37. http://schema.org/

38. *Representational State Transfer* (REST) is an abstraction of the Web architecture, an architectural style which consists of a coordinated set of architectural restrictions applied to data components, connectors and elements within a distributed hypermedia system. The REST ignores the protocol component and syntax implementation details in order to focus in the components roles, in the restriction about their interaction with other components and in their interpretation of significant data elements. (Wikipedia)

39. Depending on the region in Brazil, the term *paçoca* may refer to a candy made with peanuts or a meat dish.

Name: Pizzas Ltda
Country: Brazil
City: São Paulo

2. **Point of sale**

Name of the store: www.pizzariadaesquina.com.br
Country: Brazil
City: São Paulo
CPF: 01.000-000
Address: Av. Berrini, 999
Logo URL: www.pizzariadaesquina.com.br/logo.jpg

3. **Product**

Name: *Paçoca* pizza
Category: Pizza
Description: Pizza with banana and *paçoca*
UFL of the details: www.pizzariadaesquina.com.br/pizza/pizzadepacoca.htm
URL of the image: www.pizzariadaesquina.com.br/pizza/pizzadepacoca.jpg
Standard price: 20.00
Payment methods: Cash / credit card (Visa, MasterCard,…)

A description like this in similar environments to GoodRelations would already make the company, the store or the product was found from the search engines of the main Web browsers or applications such as Kekanto[40], which connects data about Google's restaurant offers with georeferenced information.

In case the system already counts with BCE, on which there were already ontologies of goods and services for the repair area and, more specifically, and ontology about pizzas, the user could deepen the product description, including structured details of its composition. A possible interaction would be, in case of pizzas, the definition of the components of its coverage which, for the concerned example, would be banana and *paçoca.* As the term "*paçoca*" generates a semantic ambiguity due to linguistic regionalisms[41], the system must make a clarifying request to the user. As the company is located in São Paulo, the initial suggestion is to use the interpretation given in the Southeast region. This description could be refined according to the entrepreneur's interest. For example, specifying the types of dough (normal/whole, thick/thin), size (small/medium/large), details about the coverages (types of banana, types of *paçoca*, …) This detailed description will make it possible for the users, for example, producers/suppliers of banana or peanut *paçoca,* to find or be found by this pizza place, assuming that these components will also already be previously indexed

40. Available for download at https://play.google.com/store/apps/details?id=com.kekanto.android&hl=pt_BR.

41. In some regions of Brazil, like the Northeast, *paçoca* is as dish made of flour and *carne de sol* [sun-cured beef], while in other regions it is a candy made with peanuts.

Conclusions and work agenda

Although the taxonomies, as in case of CNAE, are efficient to produce government statistics about the industrial product, they do not show to be appropriate to describe concepts and relationships which meet the new trade demand for the Electronic Business Market. In fact, a taxonomy considerably decreases the elaboration effort, but, in contrast, it preserves a certain structural rigidity, making it difficult to be used in a context that is different from the one initially foreseen. Although the taxonomies are more intuitive and allow the use of classifiers which position the product on an unequivocal way in a hierarchical tree, it is noted that this method is not able to represent all the necessary relationships to describe products in order to meet market needs and requirements. In addition, bibliographic review show that the adoption of taxonomies as knowledge organization system aggravates interoperability problems.

In order to meet the requirements set forth by SMPE for the Electronic Business Market, this report proposed the adoption of a conceptual modeling based on foundational ontologies as a strategy for the description, classification and listing of goods and services. This solution, called Essential Concepts Database (BCE, acronym in Portuguese), seems to meet the demand made and includes instruments for the homogenization of the terminology which covers the linguistic differences in and out of the Country. In addition, BCE guarantees the mapping with the traditional knowledge representation structures such as CNAE, PRODLIST and other taxonomies in use by the Government.

The proposal herein presented involves a semantic approach for the construction of a conceptual model by using ontological languages and cope with the dynamism and varieties inherent to any chain of production. This approach seems to be compatible with all the presented requirements and, in addition, it provides the flexibility required to conduct interoperability projects between more advanced commercial partners than those presented here.

Moreover, it is proposed a methodology for the construction of BCE, taking as reference the experiences collected from other research groups for the construction of ontologies from traditional product classification systems. Finally, it is presented the possibility of integration of the presented solution with an electronic commerce scenario in the context of the Semantic Web, considering product offer aspects.

Concerning the development process of the proposed solution, in face of its complexity, it is recommended a strategy based on successive refinements separated into phases. An initial version of the Electronic Business Market must provide functionalities that are similar to that of the traditional electronic commerce environments with respect to the offer of goods and services, but which already incorporates some of the Semantic Web advantages by incorporating the proposed structure for BCE.

In this first version, concerning the *product classification,* the following is suggested: *(i)* by means of tool OntoClean, making a "cleaning" in the CPV and eCl@ss, by deepening in some areas with the application of multilevels and composition, as well as PRODLIST, constructing a more flexible taxonomy and considering other types of relationships, in addition to the taxonomic ones, and *(ii)* electing some sectors/areas for the application of multilevel structures (increasing the detailing level) and composition (i.e., approaching the classification issue with more sophistication). In this case, it is suggested that, for the *product base,* a restrict set of product is elected for the purposes of concept testing forming a pilot project. Concerning the *product offer,* the following is recommended: *(i)* the adoption/adaptation of a product description and offer specification environment, such as GoodRelations, compatible with the Brazilian reality; *(ii)* the integration of this environment with the adopted product classification and *(iii)* its connection with a system for product description by the entrepreneurs.

Considering that BCE must be materialized in an electronic commerce application, it is further suggested that the ontologists team work in parallel and in line with the application developers (software factory) particularly in the initial phase of the project related to the product description and offer. Although the first steps of this project require considerable effort to be accomplished, it will be essential to form a work and improvement detection culture so that the project expansion can follow without major problems during the subsequent phases of the project.

From this first version, it is possible, in subsequent phases, to enrich the elements allowing the detailed configuration of products and services, and the consequent improvement of the product description and offer support system. With this respect, an intensive work is foreseen for the conceptual modeling group to conduct activities such as:

- Extension of the basic ontology of classifiers, configuration and offer of goods and services in order to contemplate the other typical concepts of the production chain which remained out of this pilot project.
- Construction of new domain ontologies based on conceptual modeling (or exploitation of any existing ontology) in order to contemplate functionalities which were not conducted in the pilot project. A typical example is the linguistic ontologies issue for use in the goods and services annotation schemes (tag attribution).
- Maintenance of the existing functionalities and development of new functionalities in order to contemplate new items included in the conceptual modeling of the project.

It is still worthy emphasizing the need of a wide disclosure of the project, once reaching the Electronic Business Market depends on the engagement of a number of related sectors of the society. SMPE is also responsible for paying attention to the need to provide a structure in the form or curatorship for the continuous evaluation and adjustments of the system to be offered. With this respect, it is interesting, like the Workshops promoted by the *European Committee for Standardization,* the *CEN Workshops Agreements* (CWA), meetings with the concerned community (developers, managers, etc.) in order to make them aware of the problem and discuss the progress of the standards, systems and ontologies related to the product description and offer.

REFERENCES

Abrial, J. R. (1974). Data semantics. In: Klimbie, J. W.; Koffeman, K. L. (Eds.). Proceedings of the IFIP Working Conference Data Base Management, p. 1-60. Amsterdam: North-Holland.

Albertazzi, L. (1996). Formal and Material Ontology. In: Poli, R.; Simons, P. (Ed.). Formal Ontology. Dordrecht: Kluwer. p.199-232.

Albuquerque, A. F. (2013). Ontological Foundations for Conceptual Modeling Datatypes. Dissertação de mestrado em Informática, Universidade Federal do Espírito Santo.

Almeida, M. B.; Souza, R. R.; Coelho, K. C. (2009). Estudo exploratório sobre ontologias aplicadas a modelos de sistemas de informação: perspectivas de pesquisa em Ciência da Informação. IV Encontro Ibérico EDIBCIC 2009, Coimbra, Portugal.

Alonso, L. B. N.; Ferneda, Edilson; Braga, L. V. (2011). Governo eletrônico e políticas públicas: análise sobre o uso da certificação digital no Brasil. Informação & Sociedade, v. 21, p. 13-24.

ANSI/NISO (2010). Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies – ANSI/NISO Z39.19-2005 (R2010). American National Standards Institute National Information Standards Organization.

Antoniou, G.; van Harmelen, F. (2004). A Semantic Web Primer. Cambridge, Massachusetts: The MIT Press.

Artale, A.; Calvanese, D.; Kontchakov, R.; Ryzhikov, V.; Zakharyaschev. M. (2007). Reasoning over extended ER models. In: Proc. of the 26th Int. Conf. on Conceptual Modeling (ER 2007). Lecture Notes in Computer Science, v. 4801, p. 277-292.

Attström, K.; Kröber, R.; Jähne, J. C.; Bergschneider, J.; Thomä, J.; Swysen, T.; Scholzen, F.-S.; Trautenmüller, T. (2012). Review of the functioning of the CPV codes/ system. European Commission DG Internal Market and Services.

Bernardi, D.; Calvanese, D.; De Giacomo, G. (2005). Reasoning on UML class diagrams. Artificial Intelligence, v. 168, n. 1, p. 70-118.

Berners-Lee, T.; Hendler, J.; Lissila, O. (2001). The Semantic Web. Scientific American, May 2001, p. 29-37.

Best, J. (1995). Cognitive Psychology, Ed. West Publishing Company, 4. Ed.

Beythien, A.; Dreßle, E. (Eds.) (2000). Mercks Warenlexikon für Handel, Industrie und Gewerbe. Alemanha: Manuscriptum Waltrop.

Booch, G. (1993). Object-oriented analysis and design with applications, 2nd ed. Redwood: Benjamin Cummings.

Borgo, S.; Lesmo, L. (2008). The Attractiveness of Foundational Ontologies in Industry. In: Borgo, S.; Lesmo, L. (Eds.). Formal Ontologies Meet Industry, p. 1-9. Frontiers in Artificial Intelligence and Applications, v. 174. IOS Press.

Bosak, R.; Richard, F.; Clippinger, R. F.; Dobbs, C.; Goldfinger, R.; Jasper, R. B.; Keating, W.; Kendrick, G.; Sammet, J. E. (1962). An information algebra: Phase 1 report – language structure group of the CODASYL development committee. Communications of the ACM, v. 5, n. 4, p. 190-204.

Brousseau, E. (2003). Les trois défis de l`administration électronique. In: Chatillon, G.; du Marais, B. L'administration électronique au service des citoyens: actes du colloque organisé à Paris les 21 et 22 janvier 2002 par le Conseil d'Etat et l'Université Paris I Panthéon Sorbonne. Bruxelles: Bruylant.

Calero, C.; Ruiz, F.; Piattini, M. (2006). Ontologies for software engineering and software technology. Berlin: Springer.

Campos, M. L. de A.; Gomes, H. E. (2003). Organização de domínios de conhecimento e os princípios ranganathianos. Perspectiva da Ciência da Informação, v.8, n.2, p.150-163.

Campos, M. L. de A.; Gomes, H. E. (2007). Taxonomia e classificação: a categorização como princípio. Anais do VIII Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB).

Carlan, E.; Medeiros, M. B. B. (2011). Sistemas de Organização do Conhecimento na visão da Ciência da Informação. RICI: Revista Ibero-americana de Ciência da Informação, v. 4, n. 2, p. 53-73, ago./dez.2011.

Chandrasekaran, B.; Josephson, J.; Benjamins, V. (1999) What Are Ontologies, and Why Do We Need Them?. IEEE Intelligent Systems, v. 14, n. 1, p. 20-26.

Chen, P. (1976). The entity-relationship model: towards a unified view of data. ACM Transactions on Database Systems, v. 1, n. 1, p. 9-36.

Codd, E. F. (1979). Extending the database relational model to capture more meaning. ACM Transactions on Database Systems, v. 4, 4, p. 397-434.

Codina, L.; Pedraza-Jiménez, R. (2011). Tesauros y Ontologías em Sistemas de Información Documental. El profesional de la Información, v. 20, n. 5.

Ding, Y.; Foo, S. (2002). Ontology research and development.Part 1- a review of ontology generation. Journal of Information Science, v. 28, n. 2.

e-business + Web Science Research Group (2008). GoodRelations: An Ontology for Describing Web Offers. Primer and User's Guide, Draft 2008-11-27. http://www.heppnetz.de/projects/goodrelations/primer/

Embley, D.; Campbell, D.; Jiang, Y.; Ng, Y.; Smith, R. (1998). A Conceptual-Modeling Approach to Extracting Data from the Web. In Li, T. W.; Ram, S.; Lee, M. (Eds.). Proceedings of the 17th International Conference on Conceptual Modeling – ER'98, Berlin: Spring Verlag, p. 78-91.

European Committee for Standardization (2010). Classification and Catalogue systems used in electronic public and private procurement. CEN Workshop Agreement. CWA 16138.

Farinelli, F.; Melo, S.; Almeida, M. B. O papel das ontologias na interoperabilidade de sistemas de informação: reflexões na esfera governamental. XIV Encontro Nacional de Pesquisa em Ciência da Informação – XIV ENANCIB 2013. http://mba.eci.ufmg.br/downloads/Interoperab_Enancib_2013_camera-ready.pdf

Ferneda, Edilson; Alonso, L. B. N.; Braga, L. V. (2011). Digital certification in the brazilian e-government. Journal of Information Systems and Technology Management, v. 8, n. 2, p. 331-346.

Ferneda, Edberto (2013). Um método de expansão automática de consulta baseada em ontologia. Anais do XIV Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB) – GT 8: Informação e Tecnologia.

Gómez-Pérez, A.; Fernández-López, M.; Corcho, O. (2004). Ontological Engineering with examples from the areas of knowledge management, e-commerce and the semantic web. London: Springer.

Guarino, N. (1992). Concepts, attributes and arbitrary relations: some linguistic and ontological criteria for structuring knowledge bases. Data & Knowledge Engineering, v. 8, p. 249-261.

Guarino, N. (1997). Understanding, Building and Using Ontologies: A Commentary to Using Explicit Ontologies in KBS Development, by van Heijst, Schreiber, and Wielinga. International Journal of Human and Computer Studies, v. 46, n. 2/3, p. 293-310.

Guarino, N. (1998). Formal Ontology and Information Systems. In: Guarino, N. (Ed.). Formal Ontology in Information Systems. pp. 3-15, Amsterdam, Netherlands: IOS Press.

Guarino, N.; Giaretta, P. (1995). Ontologies and KBs, Towards a Terminological Clarification. In: Mars, N. (Ed.). Towards a Very Large Knowledge Bases; Knowledge Building and Knowledge Sharing. [S.l.]: IOS Press, p. 25-32.

Guarino, N.; Welty, C. (2000a). A Formal Ontology of Properties. In: Proceedings of EKAW-2000: The 12th International Conference on Knowledge Engineering and Knowledge Management. Berlin: Springer, LNCS, Vol. 1937, pp. 97-112.

Guarino, N.; Welty, C. (2000b). Ontological Analysis of Taxonomic Relationships. In: Proceedings of ER-2000: The 19th International Conference on Conceptual Modeling, Springer-Verlag.

Guarino, N.; Welty, C. (2000c). Identity, Unity, and Individuation: Towards a Formal Toolkit for Ontological Analysis. In: Proceedings of ECAI-2000: The European Conference on Artificial Intelligence. Amsterdam: IOS Press. pp. 219-223.

Guizzardi, G. (2005). Ontological Foundations for Structural Conceptual Models. PhD Thesis, University of Twente, The Netherlands.

Guizzardi, G. (2007). On Ontology, Ontologies, Conceptualizations, Modeling Languages, and (Meta)Models. In: Vasilecas, O.; Edler, J.; Caplinskas, A. (Eds.). Frontiers in Artificial Intelligence and Applications, Databases and Information Systems IV, Amsterdam: IOS Press.

Guizzardi, G.; Wagner, G. (2005). Some Applications of a Unified Foundational Ontology in Business Modeling. In: Rosemann, M.; Green P. (Eds.). Ontologies and Business Systems Analysis, IDEA Publisher.

Guizzardi, G.; Wagner, G. (2010). Using the Unified Foundational Ontology (UFO) as a Foundation for General Conceptual Modeling Languages. In: Poli, R.; Healy, M.; Kameas A. (Orgs.); Theory and Applications of Ontology: Computer Applications. p.175-196. Germany: Springer Netherlands.

Harold, E. R.; Means, W. S. (2004). XML in a Nutshell: a desktop quick reference. 3rd ed., O'Reilly.

Heijst, G. van; Schreiber, A. T.; Wielinga, B. J. (1997). Using explicit ontologies in KBS development. International Journal of Human-Computer Studies, v. 46, n. 2-3.

Hepp, M. (2008). GoodRelations: An Ontology for Describing Products and Services Offers on the Web. In Gangemi, A.; Euzenat, J. (Eds.): Proceedings of the EKAW – Knowledge Engineering and Knowledge Management, Berlin Heidelberg: Springer-Verlag. Lecture Notes on Computing Sciences, Vol. 5268, p. 329-346.

Hepp, M. (2010a). Linked Data in E-Commerce – The GoodRelations Ontology. ISKO UK Conference on Linked Data, 2010. http://www.iskouk.org/sites/default/files/Hepp_isko2010clean.pdf

Hepp, M. (2010b). The GoodRelations Ontology for E-Commerce. 3rd KRDB School on Trends in the Web of Data (KRDB-2010). Brixen-Bressanone, Italy, 17-18 September 2010. http://www.slideshare.net/mhepp/krdb2010goodrelations.

Horrocks, I.; Kutz, O.; Sattler, U. (2006). The Even More Irresistible SROIQ. In: Proc. of the 10th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR2006). p. 57-67. AAAI Press.

IBGE (2007). Classificação Nacional de Atividades Econômicas, Versão 2.0. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, 2007.

IBICT (1993). Diretrizes para o estabelecimento e desenvolvimento de Tesauros Monolíngues. Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 1993.

Jackson, D. (2006). Software Abstractions: Logic, Language, and Analysis. Cambridge, Mass.: The MIT Press.

Jacobson, I.; Christerson, M.; Jonsson, P.; Overgaard, G. (1992). Object-oriented software engineering: A use case driven approach. Boston: Addison-Wesley.

Jiménez, A. G. (2004). Instrumentos de Representación del Conocimiento: tesauros versus ontologias. Anales de Documentación, n.7. Universidad de Murcia.

Kless, D.; Milton, S. (2010). Comparison of thesauri and ontologies from a semiotic perspective. In: Proceedings of the Sixth Australasian Ontology Workshop. Conferences in Research and Practice in Information Technology.Advances in Ontologes. Adelaide, Australia: Australian Computer Society.

Lembo, D.; Poggi, A. (2013). Conceptual Modeling: ER and beyond. Material de Curso. Great Ideas in Computer Science & Engineering, Roma, Sapienza Universitá di Roma. Available at: http://www.dis.uniroma1.it/~querzoni/corsi_assets/1213/GreatIdeasInComputerScienceAndEngineering/GIiCSaE-Lembo_Poggi.pdf.

Lenat, D. B.; Guha, R. V. (1990). Building Large Knowledge-based Systems: Representation and Inference in the Cyc Project. Boston, Massachusetts: Addison-Wesley.

Lima, V. M. A.; dos Santos, C. A. C. M.; de Lara, M. L. G.; Orth, G. P. (2015). Tesauro: um tipo de ontologia. Seminário de Pesquisa em Ontologias do Brasil – Ontobras 2015, São Paulo: Universidade de São Paulo. http://www.ime.usp.br/~ontobras/wp-content/uploads/2015/09/Ontobrasworkshop_final.pdf

Lóscio, B. F. (2012). Linked Data: da Web de documentos para a Web de dados. Cin-UFPE, Material de Curso.

Masolo, C.; Borgo, S.; Gangemi, A.; Guarino, N.; Oltramari, A. (2003). WonderWeb Deliverable D18, Ontology Library (final). Technical report, ISTC-CNR.

Milton, S. (2000). An ontological comparison and evaluation of data modelling frameworks. PhD Thesis, University of Tasmania, Hobart, AU, School of Information Systems.

Moreira, A. (2003). Tesauros e ontologias: estudo de definições presentes na literatura das áreas das ciências da computação e da informação, utilizando-se o método analíticosintético. Dissertação de Mestrado em Ciência da Informação. Belo Horizonte: Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

Morville, P.; Rosenfeld, P. (2007). Information architecture for the World Wide Web. O'Reilly, 3. Ed.

Motik, B.; Patel-Schneider, P. F.; Parsia, B. (Eds.) (2012a). OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax. 2nd Edition, W3C. Available at: http://www.w3.org/2012/pdf/REC-owl2-syntax-20121211.pdf.

Motik, B.; Patel-Schneider, P. F.; Grau, B. C. (Eds.) (2012b). OWL 2 Web Ontology Language: Direct Semantics. 2nd Edition, W3C. Available at: http://www.w3.org/TR/owl2-direct-semantics.

Mylopoulos, J. (1992). Conceptual modelling and telos. In: Loucopoulos, P.; Zicari, R. (Eds.). Conceptual modelling, databases and case: An integrated view of information systems development. New York: John Wiley & Sons.

Nardi, J. C.; Falbo, R. A.; Almeida, J. P. A.; Guizzardi, G.; Pires, L. F.; van Sinderen, M. J.; Guarino, N. (2013). Towards a Commitment-based Reference Ontology for Services. In: 17th IEEE International on Enterprise Distributed Object Computing Conference (EDOC), p. 175-184.

Neches, R.; Fikes, R.; Finin, T.; Gruber, T.; Patil, R.; Senator, T.; Swartout, W. R. (1991). Enabling Technology for Knowledge Sharing. Artificial Intelligence Magazine. v. 12, n. 3, p. 36-56.

Nipkow, T.; Paulson, L. C.; Wenzel, M. (2002). Isabelle/HOL – A Proof Assistant for Higher-Order Logic, Springer, Lecture Notes in Computer Science, Vol. 2283.

Noy, N. F.; McGuinness, D. L. (2001). Ontology Develepment 101: A Guide To Creating Your Furst Ontology. Knowledge Systems Laboratory Technical Report KSL-01-05, Stanford University. http://www-ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf.

Paoliello, C. M.; Furtado, A. L. (2004). Sistemas de Informação para Comércio Eletrônico. PUC-Rio Inf. MCC 27/04. Rio de Janeiro, July 2004. Technical Report.

Pease, R.A.; Niles, I. (2002). IEEE Standard Upper Ontology: A Progress Report. The Knowledge Engineering Review, v. 17, n. 1, p. 65-70.

Pinto, H. S.; Gómez-Perez, A.; Martins, J. P. (1999). Some Issues on Ontology Integration. In Proceedings of IJCAI99's Workshop on Ontologies and Problem Solving Methods: Lessons Learned and Future Trends.

Powers, S. (2003). Practical RDF. Sebastopol, California: O'reilly Media.

Qin, J.; Paling, S. (2000). Converting a controlled vocabulary into an ontology: the case of GEM. Information Research, v. 6, n. 2.

Ranganathan, S. R. (1967). Prolegomena to library classification. Bombay: Asia Publishing House.

Robredo, J. (2005). Documentação de hoje e de amanhã. 4ª ed. Brasília: Edição de autor.

Rumbaugh, J.; Blaha, M.; Premerlani, W.; Eddy, F.; Lorensen, W. (1991). Object-oriented modeling and design. NewYork: Prentice Hall.

Sales, R.; Café, L. (2009). Diferenças entre tesauros e ontologias. Perspectivas em Ciência da Informação. v. 14, n. 1.

Smith, B.; Welty, C. (2001). Ontology: Towards a new synthesis. In Smith, B.; Welty, C. (Eds.). Proceedings of the International Conference on Formal Ontology in Information Systems, p. 3-9. New York: ACM Press.

Smith, P. J.; Smythe, E. (2004). Globalization, citizenship and new information technologies: from the MAI to Seattle. In: Mälkiä, M.; Anttiroiko, A.; Savolainen, R. (Orgs.). Etransformation in governance: new directions in government and politics. Hershey: Idea Group Publishing.

Sowa, J. F. (1999). Knowledge Representation: Logical, Philosophical, and Computational Foundations. Pacific Grove, California: Brooks Cole Publishing Co.

Staab, S.; Studer, R. (Eds.) (2009). Handbook on Ontologies. 2nd Ed, International Handbooks on Information Systems, Springer Verlag.

Sternberg, R. (2000). Psicologia Cognitiva, Ed. Artes Médicas Sul, p. 305-338.

Stolz, A.; Rodriguez-Castro, B.; Radinger, A.; Hepp, M. (2014). PCS2OWL: A Generic Approach for Deriving Web Ontologies from Product Classication Systems. In: Proceedings of the 11th Extended Semantic Web Conference (ESWC 2014), May 25-29, 2014, Anissaras/Hersonissou, Crete, Greece, pp. 644-658.

Torres, G. M. (2012). Construção colaborativa de ontologias para domínios visuais utilizando fundamentação ontológica. Dissertação de Mestrado em Ciência da Computação, Universidade Federal do Rio Grande do Sul.

UKOLN (2005). Interoperability focus: looking at interoperability. University of Bath, UK. <http://www.ukoln.ac.uk/interop-focus/about/flyer-interoperability.pdf >.

Ullmann, S. (1972). Semantics: an Introduction to the Science of Meaning. Basil Blackwell, Oxford.

United Nations (2005). Unlocking the human potential for public sector performance. World public sector report 2005. New York: Department of Economic and Social Affairs.

Uschold, M.; Gruninger, M. (1996). Ontologies: Principles, Methods and Applications. Knowledge Engineering Review, v. 11, n. 2, p. 93-155.

Wache, H.; Vogele, T.; Visser, U.; Stuckenschmidt, H.; Schuster, G.; Neumann, H.; Hubner, S. (2001). Ontology-Based Integration of Information; a Survey of Existing Approaches. In: Goméz-Pèrez, A.; Grüninger, M.; Stuckenschmidt, H.; Uschold, M. (Eds.). Proceedings of IJCAI-01 Workshop: Ontologies and Information Sharing, p. 108–117. Seattle, WA.

Wallis, J. J.; North, D. C. (1986). Measuring the Transaction Sector in the American Economy, 1870–1970. In: Engerman, S. L.; Gallman, R. E. (Eds.). Long-Term Factors in American Economic Growth. University of Chicago Press, 1986. p. 95-162.

Young, J. W.; Kent, H. K. (1958). Abstract formulation of data processing problems. Journal of Industrial Engineering, v. 9, n. 6, p. 471-479.

Zamborlini, E. C. (2011). Estudo de alternativas de mapeamento de ontologias da linguagem OntoUML para OWL: abordagens representação de informação temporal. Dissertação de Mestrado em Informática, Vitória: Universidade Federal do Espírito Santo.

Zeng, M. L.; Salaba, A. (2005). Toward an International Sharing and Use of Subject Authority Data. In: FRBR Workshop, OCLC. Kent State University.

European Union

SECTOR DIALOGUES EUROPEAN UNION BRAZIL

SECRETARIAT OF MICRO AND SMALL ENTERPRISES

MINISTRY OF GOVERNMENT SECRETARIAT

MINISTRY OF PLANNING

BRASIL BRAZILIAN GOVERNMENT