UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO

FERNANDA FARINELLI

REALISMO ONTOLÓGICO APLICADO A INTEROPERABILIDADE SEMÂNTICA ENTRE SISTEMAS DE INFORMAÇÃO: UM ESTUDO DE CASO DO DOMÍNIO OBSTÉTRICO E NEONATAL

Belo Horizonte
2017

FERNANDA FARINELLI

REALISMO ONTOLÓGICO APLICADO A INTEROPERABILIDADE SEMÂNTICA ENTRE SISTEMAS DE INFORMAÇÃO: UM ESTUDO DE CASO DO DOMÍNIO OBSTÉTRICO E NEONATAL

Tese apresentada ao Programa de Pós-Graduação em Gestão & Organização do Conhecimento da Escola de Ciência da Informação, Universidade Federal de Minas Gerais, como requisito para obtenção do grau de Doutor em Ciência da Informação.

Linha de Pesquisa: Gestão & Tecnologia

Área de concentração: Representação do Conhecimento

Orientador: Maurício Barcellos de Almeida

BELO HORIZONTE

2017

F225r

Farinelli, Fernanda.

Realismo ontológico aplicado a interoperabilidade semântica entre sistemas de informação [manuscrito] : um estudo de caso do domínio obstétrico e neonatal = Improving semantic interoperability in the obstetric and neonatal : domain through an approach based on ontological realism / Fernanda Farinelli. – 2017.

256 f., enc. : il.

Orientador: Maurício Barcellos de Almeida.

Tese (Doutorado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

Referências: f. 220-225.

Apêndices: f. 225-256.

Em inglês; folha de rosto e resumo em português.

1. Ciência da informação – Teses. 2. Ontologias (Recuperação da informação) – Teses. 3. Sistemas de recuperação da informação – Teses. 4. Prontuário eletrônico do paciente – Teses. I. Título. II. Almeida, Maurício Barcellos de. III. Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CDU: 025.4:61

Ficha catalográfica: Biblioteca Profª Etelvina Lima, Escola de Ciência da Informação da UFMG.

FERNANDA FARINELLI

IMPROVING SEMANTIC INTEROPERABILITY IN THE OBSTETRIC AND NEONATAL DOMAIN THROUGH AN APPROACH BASED ON ONTOLOGICAL REALISM

Thesis submitted to the Graduate Program in Knowledge Organization and Management of the School of Information Science at the Federal University of Minas Gerais, in partial fulfilment of the requirements for the degree of Doctor in Information Science.

Line of Research: Management & Technology

Area of concentration: Representation of Knowledge

Supervisor: Maurício Barcellos de Almeida

BELO HORIZONTE

2017

UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO

FOLHA DE APROVAÇÃO

REALISMO ONTOLÓGICO APLICADO A INTEROPERABILIDADE SEMÂNTICA ENTRE SISTEMAS DE INFORMAÇÃO: UM ESTUDO DE CASO DO DOMÍNIO OBSTÉTRICO E NEONATAL

FERNANDA FARINELLI

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, como requisito para obtenção do grau de Doutor em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, área de concentração REPRESENTAÇÃO DO CONHECIMENTO, linha de pesquisa Gestão e Tecnologia.

Aprovada em 30 de outubro de 2017, pela banca constituída pelos membros:

Prof(a). Mauricio Barcellos Almeida - Orientador
ECI/UFMG

Prof(a). Renato Rocha Souza
FGV/RJ (por videoconferência)

Prof(a). Fabrício Martins Mendonça
UFJF

Prof(a). Zilma Silveira Nogueira Reis
Medicina/UFMG

Prof(a). Renata Maria Abrantes Baracho Porto
ECI/UFMG

Belo Horizonte, 30 de outubro de 2017.

ATA DA DEFESA DE TESE DA ALUNA
FERNANDA FARINELLI

Realizou-se, no dia 30 de outubro de 2017, às 09:00 horas, Sala 1000 - ECI/UFMG, da Universidade Federal de Minas Gerais, a defesa de tese, intitulada *REALISMO ONTOLÓGICO APLICADO A INTEROPERABILIDADE SEMÂNTICA ENTRE SISTEMAS DE INFORMAÇÃO: UM ESTUDO DE CASO DO DOMÍNIO OBSTÉTRICO E NEONATAL*, apresentada por FERNANDA FARINELLI, número de registro 2013708429, graduada no curso de CIÊNCIA DA COMPUTAÇÃO, como requisito parcial para a obtenção do grau de Doutor em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, à seguinte Comissão Examinadora: Prof(a). Mauricio Barcellos Almeida - ECI/UFMG (Orientador), Prof(a). Renato Rocha Souza - FGV/RJ (por videoconferência), Prof(a). Fabrício Martins Mendonça - UFJF, Prof(a). Zilma Silveira Nogueira Reis - Medicina/UFMG, Prof(a). Renata Maria Abrantes Baracho Porto - ECI/UFMG.

A Comissão considerou a tese:

(X) Aprovada

( ) Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.

Belo Horizonte, 30 de outubro de 2017.

Prof(a). Mauricio Barcellos Almeida

Prof(a). Renato Rocha Souza

Fabrício M Mendonça

Prof(a). Fabrício Martins Mendonça

Prof(a). Zilma Silveira Nogueira Reis

Prof(a). Renata Maria Abrantes Baracho Porto

DEDICATÓRIA

Dedico esta tese …

Aos meus pais Antonio e Nilza, e que dignamente me apresentaram à importância da família e o caminho da honestidade e persistência.

Às mães e gestantes que são o alvo das melhorias que esta tese pode trazer no que tange à continuidade dos cuidados de saúde prestados durante o pré-natal, intraparto e pós-parto.

AGRADECIMENTOS

Esta tese não é resultado apenas de um esforço individual. Ela é resultado de significativas contribuições, diretas ou indiretas, recolhidas durante minha trajetória profissional, acadêmica e pessoal, ao lidar com pessoas, instituições e situações fundamentais a essa construção. Neste momento percebo que uma etapa importante da minha vida se completa, trazendo o desafio de expressar minha gratidão a todos aqueles que contribuíram e me apoiaram nesta longa caminhada.

Primeiro, eu gostaria de expressar minha sincera gratidão ao Professor Maurício Barcellos de Almeida, por supervisionar meu trabalho, pelas suas sugestões, questionamentos, requisições, e revisões ao longo destes anos que trabalhamos juntos. Tive a sorte e o prazer de receber sua orientação, encorajamento, tolerância, liberdade, confiança, hospitalidade e sua amizade. Nossa parceria apenas começou...

Agradeço à equipe do PPGCI e PPGGOC da Escola de Ciência da Informação da Universidade Federal de Minas Gerais pelo apoio em todas as dificuldades de âmbito administrativo.

Agradeço à Universidade Estadual de Nova York em Buffalo (UB), em particular aos Professores Barry Smith e Peter L. Elkin, meus co-orientadores durante o intercâmbio de pesquisa. Ambos contribuíram pacientemente com minha pesquisa com valiosas ideias, conselhos, críticas, sugestões e revisões. Preciso mencionar que desfrutar da companhia e ensinamentos dos Professores Smith e Elkin foi um imenso prazer e uma grande honra. Sua gentileza, direção e apoio foram inestimáveis, tanto no âmbito da ciência como para minha vida pessoal. Eu me tornei uma pessoa melhor depois desta convivência. Estendo ainda este agradecimento a todos os colegas dos Departamentos de Informática Biomédica e Filosofia da UB, pelo acolhimento, gentileza e trocas enriquecedoras.

Agradeço também a minha família, Yóris, meu marido, por todo seu apoio, resistiu as minhas ausências e apoiou meus planos. Aos meus pais e irmãos, pela motivação, apoio, e por entenderem meus momentos de afastamento e reclusão. Meus sobrinhos, em especial à Luíza, Rafaela e Hugo, sua existência e sorrisos me deram força para acreditar e avançar em meu sonho.

Finalmente, agradeço o apoio financeiro das instituições brasileiras de fomento à pesquisa, CNPq e Fapemig pela bolsa de estudos no Brasil, e a CAPES pela bolsa de estudos durante meu intercâmbio na SUNY. Sem apoio financeiro, não há ciência e inovação no Brasil. Por favor, não parem de investir na pesquisa Brasileira!

Todos os homens, por natureza, desejam conhecer. (Aristóteles)

All men by nature desire to know. (Aristotle)

RESUMO

As instituições de saúde operam em ambientes de *Big Data*, onde são produzidos grandes volumes de informações em velocidade cada vez mais altas além de serem criadas em uma variedade de formatos. São gerados milhares de registros de saúde por segundo desde uma simples consulta médica, à exames complexos e sofisticados. Os profissionais de saúde usam o registro do paciente para registrar dados do estado físico e mental dos pacientes e apoiar a comunicação entre eles. O desenvolvimento da tecnologia da informação torna esse registro eletrônico e conhecido como prontuário eletrônico do paciente (PEP). A continuidade dos cuidados aos indivíduos depende de tais registros. Assim, a continuidade do atendimento tanto para mulheres quanto para recém-nascidos deve agrupar registros de cuidados de saúde prestados durante o pré-natal, intraparto e pós-parto. Em relação ao Sistema Único de Saúde, um cuidado pode ser prestado dentro de uma pluralidade de instituições de saúde, localizadas em diferentes locais geográficos ou jurisdições políticas. Observa-se uma fragmentação na informação dos PEPs em vários sistemas de informação, com padronização semântica mínima entre eles. Tal fragmentação com lacunas de padronização semântica dificulta a troca de dados entre os sistemas, exigindo maior esforço para unificar os dados de PEP e assim promover a continuidade da saúde para a mãe e a criança.

A promoção da troca de dados entre PEPs depende de uma solução semântica capaz de representar este domínio do conhecimento sem ambiguidade. Para superar a falta de interoperabilidade semântica, observamos várias iniciativas em torno de ontologias, devido à sua capacidade de representar o conhecimento. As investigações sobre ontologias como ferramenta de representação e organização do conhecimento ganham espaço no campo da Ciência da Informação.

Mas ainda não está claro qual a contribuição das ontologias na interoperabilidade, principalmente considerando as diferentes abordagens existentes. Desta forma, buscando preencher lacunas sobre o papel e as possibilidades oferecidas pelas ontologias na representação semântica de um domínio do conhecimento na promoção da interoperabilidade, o objetivo geral deste estudo é demonstrar uma alternativa para a interoperabilidade semântica entre os sistemas de informação usando uma representação do domínio obstétrico e neonatal do conhecimento por meio da abordagem do realismo ontológico. Esta tese apresenta uma pesquisa qualitativa aplicada e exploratória, envolvendo um estudo de caso que investiga a lacuna semântica de troca de dados no campo médico e as possibilidades de ontologias formais que promovem a interoperabilidade em tais trocas de dados. Durante a fase exploratória, estudamos o domínio obstétrico e neonatal e determinamos a metodologia para construir a ontologia alvo. Na fase empírica, construímos a ontologia do domínio obstétrico e neonatal e a validamos identificando a possibilidade de interoperabilidade. Busca-se oferecer um cenário simples para o uso da ontologia na resolução de problemas de interoperabilidade semântica entre sistemas de informação por meio da apresentação de uma arquitetura de integração de dados orientada à ontologia.

Palavras chave: Ontologia biomédica; ontologia formal; realismo ontológico; ontologia do domínio obstétrico e neonatal; interoperabilidade; prontuário eletrônico do paciente, arquitetura de dados orientada à ontologia.

ABSTRACT

Healthcare institutions operate in environments with large volumes of information, with a simple healthcare encounter sometimes generating thousands of records per second. Healthcare professionals use a patient's record – now, with the growth of Information Technology, becoming an *electronic* health record (EHR) – to register information concerning the physical and mental state of patients and to support the communication of such information amongst themselves. Continuity of care to individuals depends on such records. Thus, to foster continuity of care for both women and newborn one would need to group together records relating to the healthcare provided during prenatal, intrapartum and postpartum care. Regarding the Brazilian Unified Health System, such care can be provided in a plurality of institutions located in different geographic locations or political jurisdictions. This scenario brings about a fragmentation of information in EHRs using different information systems with minimal semantic standardization among them. Such fragmentation makes it difficult to exchange data between EHR systems, requiring significant effort to bring together EHR data in the sort of unified form that is needed to ensure the continuity of quality healthcare for both mother and child.

Promotion of data exchange among EHRs depends in part on a semantic solution – in order words a solution involving specification of the meanings and logical interrelations of the terms used – that is capable of representing this domain of knowledge without ambiguity within a single system. Such a system would enable semantic interoperability across the separate EHRs, and we observed several initiatives along these lines employing ontologies and relying on the capacity of the latter to represent knowledge in a uniform way. Investigations on ontologies as a tool for representing and organizing knowledge have been gaining more and more ground in the field of Information Science.

However, it is still not clear how ontologies could be used to promote interoperability, given that there are multiple existing ontologies of different types. The objective of this study is to fill this gap in our understanding of the possibilities offered by ontologies in promoting interoperability through the semantic representation of a knowledge domain. We introduce a new paradigm for achieving semantic interoperability among information systems through a representation of the obstetrical and neonatal domain using an approach based on ontological realism. This thesis is an explanation in applied qualitative research involving a case study in which we investigate the semantic gap in medical data exchange and to demonstrate the possibilities of formal ontologies for closing this gap. In an empirical, exploratory phase, we studied the obstetric and neonatal domain and determined the methodology to build the target ontology. We built on this basis an ontology of the obstetric and neonatal domain and validated the ontology by investigating its ability to foster interoperability. The intention is to offer a simple and generalizable framework in which to demonstrate how ontology can be used to solve issues of semantic interoperability across information systems, through a proposal of ontology-driven data architecture to data integration.

Keywords: Biomedical ontology; formal ontology; ontological realism; obstetric and neonatal ontology; basic formal ontology; interoperability; electronic health record, ontology-driven data architecture.

LIST OF ABBREVIATIONS

ACOG	-	American Congress of Obstetricians and Gynecologists
BFO	-	Basic Formal Ontology
ChEBI	-	Chemical Entities of Biological Interest
CL	-	Cell Ontology
DOID	-	Human Disease Ontology
DOLCE	-	Descriptive Ontology for Linguistics and Cognitive Engineering
EDS	-	Electronic Discharge Summary
EHR	-	Electronic Health Record
EMR	-	Electronic Medical Record
EnvO	-	Environment Ontology
FMA	-	Foundational Model of Anatomy
GFO	-	General Formal Ontology
GO	-	Gene Ontology
IAO	-	Information Artifact Ontology
ICT	-	Information and Communications Technology
IRI	-	Internationalized Resource Identifier
IRS	-	Information Retrieval Systems
KO	-	Knowledge organization
KOS	-	Knowledge Organization Systems
NeOn	-	Network Ontology
OAE	-	Ontology of Adverse Events
OBI	-	Ontology for Biomedical Investigations
OBO Foundry	-	The Open Biological and Biomedical Ontologies
ODP	-	Ontology Design Pattern
OGMS	-	Ontology for General Medical Science
OMRSE	-	Ontology of Medically Related Social Entities
OntONeo	-	Ontology of Obstetric and Neonatal domain
ORSD	-	Ontology Requirements Specification Document
OWL	-	Ontology Web Language
PATO	-	Phenotypic quality
PHR	-	Personal Health Record
PO	-	Protein Ontology
RDF	-	Resource Description Framework
RO	-	Relation Ontology

SIO	-	Semanticscience Integrated Ontology
SUMO	-	Suggested Upper Merged Ontology
SUS	-	Sistema Único de Saúde (in Portuguese) or Brazilian Unified Health System
UFMG	-	Universidade Federal de Minas Gerais (in Portuguese) or Federal University of Minas Gerais
UFO	-	Unified Foundational Ontology
URI	-	Uniform Resource Identifier
VO	-	Vaccine Ontology
XML	-	eXtensible Markup Language
YAMATO	-	Yet Another More Advanced Top-level Ontology

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

PART I

Exploratory phase

1 INTRODUCTION

"Success is not final, failure is not fatal: it is the courage to continue that counts".
Winston S. Churchill

Various authors in different research fields have discussed the knowledge involved in organizational activities. The manner how organizations acquire, share, create, validate, and publish knowledge reflects directly on their competitiveness and their governance. Organizations, both in the public or private ambit, yearn for providing quality services to individuals and related entities. Therefore, providing the right information in the shortest possible time becomes a major challenge in a society that demands quick decisions from organizations, requiring also adaptation of organizations to both economic and legal variations.

Currently, the great challenge of organizations is the need of integration of their large volume of information with the information of their partners business, both internals and externals. The initiative around the interoperability among information systems is a worldwide trend for organizations. Indeed, it is no so difference in the health field. The amount of data produced in the medical practice and the biomedical research achieves huge volumes. Health institutions operate in environments in which generating thousands of health records of patient per second through either a laboratory test or image exams.

Patient's record, also called medical record, is a document used by health professionals to record patient data including both comprehensive and continuous care. This document became an important communication tool to health professionals and health institutions. In addition, with the advent of the *Information and Communications Technology* development, the paper-based medical record gives way to the *Electronic Health Record* (EHR) or *Electronic Medical Record*.

EHRs serve as repositories of information about the physical and mental state of patients, as well as the health state of the patient's families as these evolve over time. In order to foster continuity of care, health care providers (institutions and professionals) should use information documented both in EHR and hospital discharge summaries. However, analyzing the structure of the Brazilian Network Healthcare, we noted that a patient's healthcare could be provided within a plurality of healthcare institutions, probably located in different geographical locations (different cities or states) or political jurisdictions (federal, state or municipal), often involving a plurality of different EHR systems or other type of information system. Additionally, when we thinking about the continuity of care, we identified that healthcare providers (institutions and professionals) should use the information documented in both EHR and hospital discharge summaries to foster better care. Accordingly, we observed the need for data integration between the different information systems of the Brazilian Network

Healthcare. Nonetheless, such data integration involves data coming from different information systems that could adopt different terminologies or information models and consequently, this heterogeneity leads to semantic interoperability problems.

With the aim of minimizing interoperability gaps between health systems, the Brazilian Government by the decree number 2.073, defined the standards of interoperability for health information systems (Brasil, 2011b). Another Brazilian initiative comes from the State of Minas Gerais that created the Public Portal of Electronic Health Record Base (B-RES). It uses ISO 13606 standard as information model and intends to serve as a central base for healthcare information, in subsidizing both the planning and the improvement of public health policies at state level. However, the adoption of these standards is not a reality yet, and the data exchange between EHR systems remains poor.

Regarding to the continuity of care to women and her child, the Brazilian Ministry of Health established the Stork Network Program with the aim of to ensuring the continuity of care for women (prior, during and post pregnancy) and her child since birth until 24 months (Brasil, 2011a):

> *The Stork Network, established under the National Health System, is a network of care that aims to ensure women's entitlement to reproductive planning and attention humanized to pregnancy, to childbirth and to the postpartum period, and the child's entitlement to safe birth, and to growth and development healthy (Brasil, 2011a)[1].*

Hence, in order to foster a better continuity of care to women and her child, healthcare providers should gather data of their patients documented on during the care of prenatal, parturition, and postnatal stages becoming these data accessible to the healthcare professionals. Nevertheless, in Brazil, we detected a clear fragmentation of care information of the obstetrical and neonatal patients, which are the target of Stork Network, into various information systems described in Table 1. We highlighted the minimal semantics standardization among these systems.

Next, observing the operation of the Maternity Hospital from the Federal University of Minas Gerais, from now referred in this thesis just as *Maternity of UFMG*, we noticed that there are gaps in data exchange both internally between departments and externally with other institutions. Figure 1 presents a scheme of the information flow of the obstetrical and neonatal

[1] In Portuguese: “A Rede Cegonha, instituída no âmbito do Sistema Único de Saúde, consiste numa rede de cuidados que visa assegurar à mulher o direito ao planejamento reprodutivo e à atenção humanizada à gravidez, ao parto e ao puerpério, bem como à criança o direito ao nascimento seguro e ao crescimento e ao desenvolvimento saudáveis” (Brasil, 2011, Art. 1°).

patient through the different information systems used by the Maternity of UFMG, against the information systems referenced in Table 1.

Table 1 – Summary of information system of the Brazilian healthcare network.

Information system	Description
AGHU[2] system	The system used by the Hospitals of Federal Universities (federal school hospital) to register patients in a routine medical consultation, an exam, or a hospitalization.
SOUL MV® system	EHR system used by the Maternity of UFMG to register the findings of the care in prenatal and childbirth provided to women.
SISMater[3] system	Used by the Maternity of UFMG to group both clinical and administrative data on obstetric and neonatal care, and provides some indicators of maternal and neonatal care to help the monitoring the healthcare quality service offered to the population.
SISPRENATAL[4] system	The system used within the scope of the Brazilian Unified Health System (SUS[5]) to provide adequate follow-up of pregnant women of the Prenatal and Birth Humanization Program (PHPN) under the Stork Network Program.
SISAB[6] system	The system used in the scope of the SUS, to integrate information regarding the programs and strategies of the Brazilian Primary Care improving information management. This system is an element of the e-SUS AB strategy.
PEC[7] system	A system used to capture data during a healthcare encounter. It intended for the Basic Units of Health (UBS[8]) of SUS with computers and some degree of connectivity. This system is an element of the e-SUS AB strategy.
CDS[9] system	A transitional system that uses paper forms to meet the UBS that are not yet computerized or have internet connection. This system is an element of the e-SUS AB strategy.

Source: Prepared by the author based on Brasil (2013, 2015, 2016).

The Maternity of UFMG has two kinds of patients: the patient of the primary care and the obstetrical and neonatal patient (stork network patient). The AGHU system records the demographic data of both kinds of patient and the SOUL MV system records the clinical data of the stork network patient. The PEC system records clinical data of primary care patients. Thus, the B-RES should gather data from both PEC system and SOUL MV system,

[2] AGHU is an acronym to Application of Management for University Hospitals, in Portuguese "*Aplicativo de Gestão para Hospitais Universitários*".

[3] SISMATER is an acronym to Information System for Maternal and Neonatal Health, in Portuguese "*Sistema de Informação em Saúde Materna e Neonatal*".

[4] SISPRENATAL is an acronym to Pregnancy Monitoring System, in Portuguese "*Sistema de Acompanhamento da Gestante*".

[5] SUS is an acronym to Brazilian Unified Health System, in Portuguese "*Sistema Único de Saúde*".

[6] SISAB is an acronym to Information System for Health Primary Care, in Portuguese "*Sistema de Informação em Saúde para a Atenção Básica*".

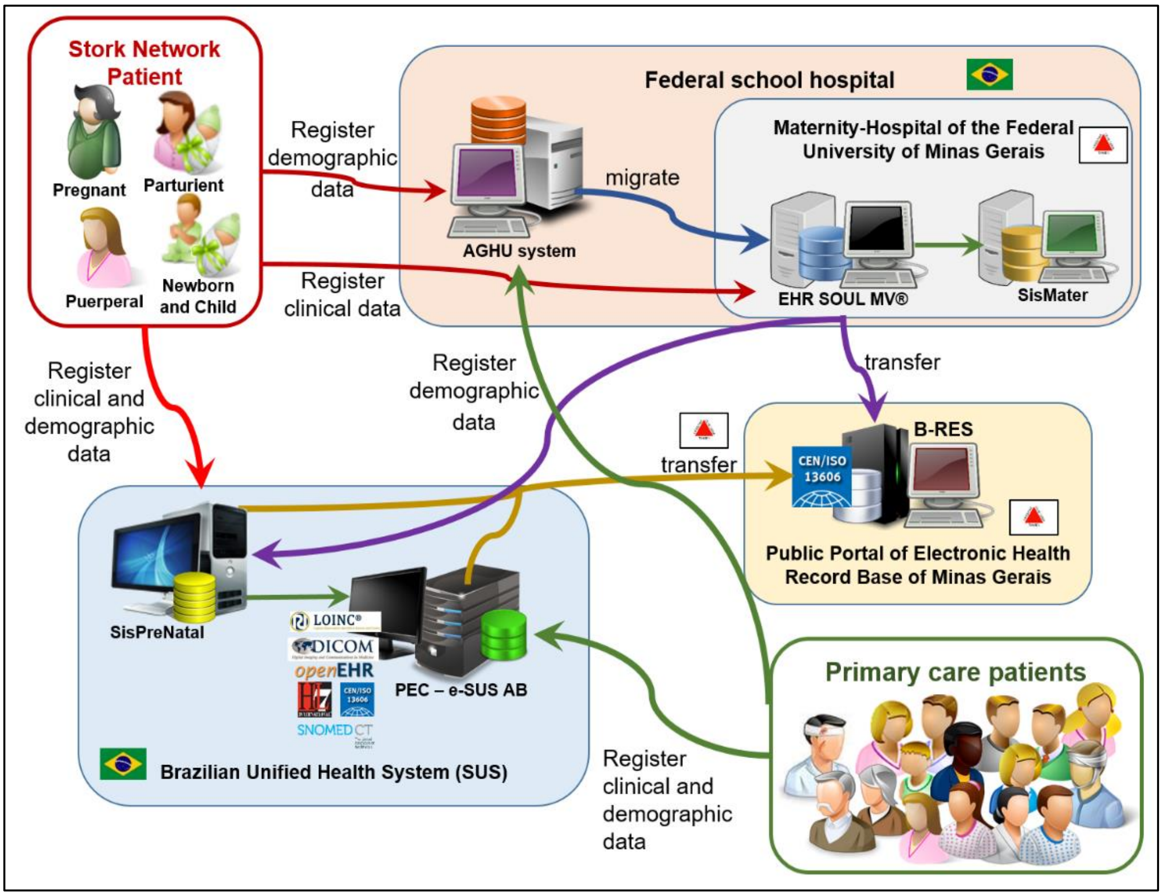
[7] PEC is an acronym to Citizen Electronic Health Record, in Portuguese "*Prontuário Eletrônico do Cidadão*".

[8] UBS is an acronym to Basic Units of Health, in Portuguese "*Unidade Básica de Saúde*".

[9] CDS is an acronym to Simplified Data Collection, in Portuguese "*Coleta de Dados Simplificado*".

and the SISPRENATAL system should receive both clinical and demographic data against stork network patient.

Figure 1: Information flow of the Maternity of UFMG through information systems.



Source: Prepared by the author

Given the complexity of this scenario, full access to clinical data, usually available in EHR system, is crucial to ensuring continuity of care when patients move from one healthcare facility to the next, established in different geographical locations or political jurisdictions. The main idea here is that the data exchange of obstetric and neonatal care is necessary to ensure continuity of care to women and newborns. Thus, the promotion of data exchange among obstetric and neonatal EHR systems depends on a semantic solution able to represent this knowledge domain without ambiguity.

In order to overcome the lack of semantic interoperability, several initiatives made use of solutions based on a formal ontology since its capability to represent knowledge without ambiguity. There is a recognition that domain ontologies can mitigate semantic interoperability issues among the terminological heterogeneity (Bittner; Donnelly; Winter, 2005, Gangemi; Fisseha*; et al.*, 2002a, Smith, 2003c).

Ontology as a research subject of Information Science assumes an important role as a tool of knowledge organization (Almeida; Barbosa, 2009, Søergel, 1999, Vickery, 1997). In general, Information Science field deals with problems around the needs of information and its applicability to society. Seminal authors of Information Science claim the interdisciplinarity of the field and its social function providing solutions to other areas (Borko, 1968, Hawkins, 2001, Saracevic, 1995, 1999). Thus, this thesis does not attempt to produce new theories in the Information Science field, but to apply existing theories to solve problems in other fields.

Assumptions and research problem:

Faced the presented synopsis regarding the fragmentation of the care information of the obstetric and neonatal patient into various information systems with minimal semantics standardization among them, some assumptions drive this research:

- *Assumption 1:* The recognition of the Information Science field as a solution provider to others scientific fields regarding information problems.
- *Assumption 2:* A perception of methods to knowledge representation and information retrieval as research subject of Information Science field.
- *Assumption 3:* The research on formal domain ontologies as a suitable solution to knowledge representation and information retrieval.
- *Assumption 4:* Formal domain ontologies mitigate semantic interoperability problems among the terminological heterogeneity.
- *Assumption 5:* There is a gap of semantically interoperable solutions for the exchange of EHR data of the patient undergoing on prenatal, childbirth and postnatal care.
- *Assumption 6:* There is a lack of formal representation of obstetric and neonatal knowledge.
- *Assumption 7:* There are several methodologies and approaches on ontology development, and there is no consensus on the most appropriated on solutions of data interoperability between information systems.

Based on these assumptions, this study aims to answer the following **research questions**:

What is the contribution of the realism-based formal ontologies to the semantic representation of a knowledge domain in order to foster data interoperability between information systems?

What the opportunities and challenges of the ontology building to the semantic representation of a knowledge domain in order to foster data interoperability between information systems?

Research goals:

This study is an initiative that seeks to enrich the discussion of the Information Science social role in which can help other scientific fields in reference to solutions related to representation, organization, and retrieval of information. Guided by the assumptions and questions above, the **general objective** of this study is *to demonstrate an alternative for semantic interoperability among information systems by using a representation of the obstetrical and neonatal domain of knowledge using ontological realism approach.*

In order to reach such general objective, we expected to achieve the following **specific objectives**:

1st. Identify the scope of knowledge of the obstetric and neonatal domain that needs representation.

2nd. Identify methodologies and approaches to build formal ontology compatible with semantic interoperability issues.

3rd. Analyze such methodologies and approaches in order to determine a suitable methodology to build the obstetrical and neonatal ontology.

4th. Build a domain ontology covering the knowledge scope of the obstetrical and neonatal domain.

5th. Define an ontology-driven architecture suitable for data retrieval compatibles semantically.

6th. Demonstrate the use of the ontology built to retrieval data of obstetrical and neonatal records considering such architecture.

The intention here is to offer a simple ontology-driven architectural framework, within which to be possible the use of ontology to solve semantic interoperability issues across information systems.

Struture of this document:

The remaining of this thesis is organized as follows.

Chapter 2 refers to essential background theories needed to give understanding to this thesis. The chapter begins presenting basic approaches in the field of Information Science related to this research (subsection 2.1) and then some fundamentals of interoperability (subsection 2.2). Next, we present the fundamentals of ontologies (subsection 2.3) discussing issues such as the interdisciplinarity of ontology topic as a research subject, basic concepts of ontology, and interoperability questions against ontologies. Subsection 2.4 reviews some main references on methodologies to build ontologies that served as a guide to this thesis development. To situate the reader of this thesis, subsection 2.5 presents the concept of the semantic web and some significant relate technologies. Afterward, subsection 2.6 examines the context of information and information systems in healthcare.

Chapter 3 reports the research methodology describing scientific research methodology issues regarding this thesis and presenting the methods used to answer the question and to reach the goals. Then, chapter 4 presents the results of the empirical research. Next, chapter 5, we present a discussion of our findings resuming triggers elements of this investigation such as research questions and objectives. Chapter 6 presents final remarks, including contributions highlights, limitations, publications already originated from this study, and recommendations for future work.

2 THEORETICAL BACKGROUND

"To be useful, knowledge must be organized"
(Søergel, 2009a)

A literature review is a basic step in a process to research development, which usually occurs in initial research phases but could happen during all research (Creswell, 2013, Kumar, 2011). Thus, this chapter presents an overview of the knowledge acquire during our literature review. This overview of knowledge support the development of this study and consequently is relevant to hold readers understanding of the thesis.

2.1 The Information Science approach of this thesis

It is fundamental to understand the relation of this research with the Information Science field. Some theoretical essentials Information Science seminal authors of that field such as Borko (1968), Saracevic (1995, 1999), and Hawkins (2001) to mention a few that help to explain this relation.

The field of Information Science explores the methods, processes, and features that involve information in order to understand the information cycle process. Information Science is a discipline in which both scientific researchers and practitioners deal with problems around the needs of information and its applicability to society. Aspects like interdisciplinarity across different science fields, strong relations with information technology, and involvement in the development of information society motivate the nature of Information Science. Information Science field is interdisciplinary and focused on theoretical, practical, and technological aspects of knowledge and information. There are interdisciplinary connections between Information Science and fields such as Librarianship, Computer Science, Cognitive Science, and Communication studies.

Reviewing these seminal authors it is possible highlighted the relation between Information Science studies and information cycle processes. Information cycle processes involve generation, collection, organization, representation, storage, retrieval, interpretation, transmission, dissemination, processing, and use of information. In addition, the need of information retrieval in any science field reinforces the basis of interdisciplinarity from Information Science field to others fields.

Hawkins (2001) and Hawkins; Larson and Caton (2003) proposed a taxonomy for Information Science distributed in 11 thematic categories. This investigation focuses on the thematic category "*organization of knowledge*" aiming the processing, retrieval and dissemination of information in the healthcare area, and to address the deficiency of organization and retrieval of information regarding obstetric and neonatal healthcare. The knowledge organization category has as subcategories controlled vocabularies, terminologies, taxonomies, and ontologies (Hawkins, 2001).

Seminal authors such as Saracevic (1995, 1999), Hawkins (2001), Hawkins; Larson and Caton (2003), highlighted the role of Information Science in dealing with questions related to information and knowledge to serve as a reference for other fields of research. Considering the claim of interdisciplinarity between Information Science field and other scientific fields, and the need of knowledge organization in different contexts in almost every scientific field, the goal of this thesis fits perfectly as a research subject in the Information Science field. This thesis research subject involves the application of Information Science theories, ontologies more specifically, into biomedical and healthcare fields, aiming to solve issues of organization and retrieval of information among data environments, which are not interoperable.

Initiatives around the knowledge organization and representation emerge with the evolution of society that yearns for tools that enable sharing, decoding, and use of knowledge generated, accumulated, and recorded by society itself. Knowledge Organization has become an important subject of study and practice in Information Science. Some seminal authors used in this thesis, who discuss and develop theories about Knowledge Organization (KO) are Søergel (1985, 2009a, 2009b), Dahlberg (1993, 1995, 2006), and Hjørland (1994, 2003, 2007, 2008).

Compiling the arguments of these mention authors, we realized that KO has scientific nature and aims to systematize knowledge units (or concepts) from the elements or features inherent to them, combining the application of these concepts and classes of concepts in order to allow the transfer of content or subjects. Knowledge Organization involves the performance of operational activities by libraries, archives, and databases. It also concerns with the nature and the quality of processes and knowledge organization systems, that is, building conceptual systems.

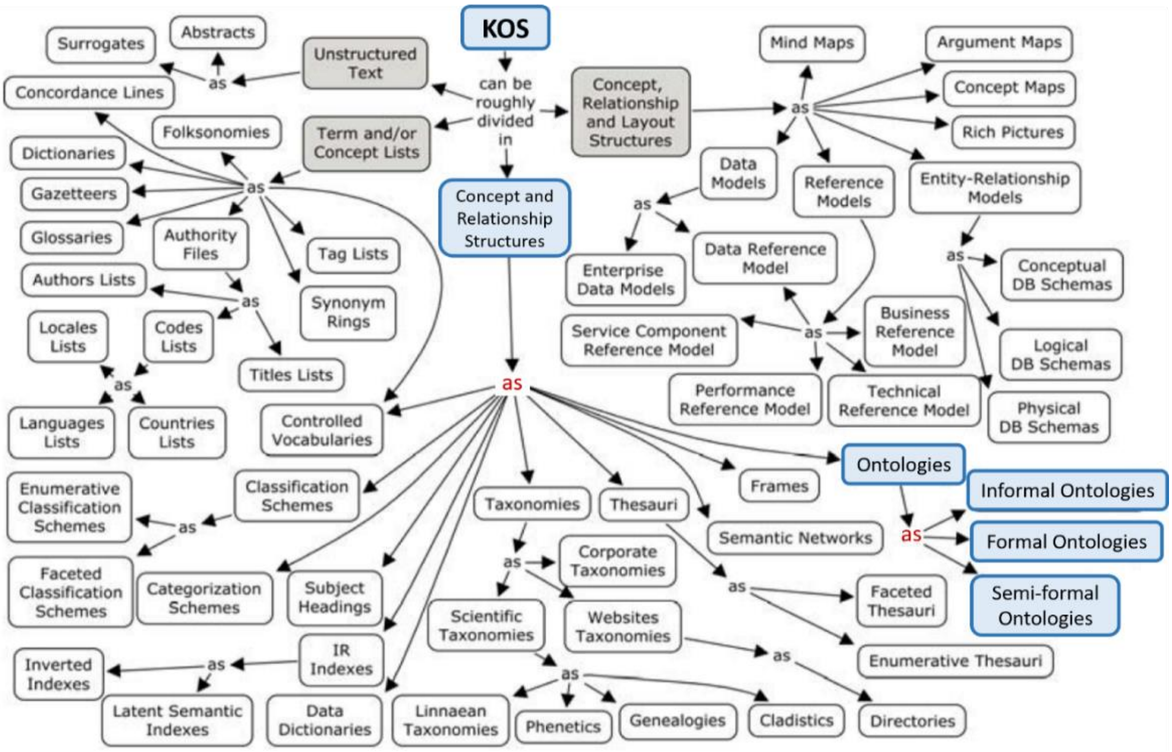
Søergel (2009a) considers a possible nonexistence of knowledge without knowledge organization. Knowledge must be organized to be used either by people or by computers. Knowledge is any representation of anything, and the organization refers to the placement of anything in a specified formation. Thus knowledge organization is the placement of knowledge according to a predefined structure.

Since we introduce a meaning for Knowledge Organization, now it is required to understand Knowledge Organization Systems (KOS) as mechanisms of knowledge representation. KOS covers any model or design created in order to enable knowledge organization and management (Hodge, 2000). KOS are structures created to organize knowledge and to foster its recovering and use. Some elements such as "concepts, categories, classes, relationships between them" are the focus of KOS (Søergel, 2009b).

> *Knowledge Organization Systems are used by people to find information and make sense of it; KOS must support people in their quest for meaning, they must present meaningful structures of concepts. KOS are also used by computer programs to reason about data; KOS must represent formal knowledge about concepts (Søergel, 2009b).*

Souza; Tudhope and Almeida (2010, 2012) presented various types of knowledge organization schemes with different level of complexity (Figure 2).

Figure 2: Ontologies as type of knowledge organization system.



Source: Adapted from Souza; Tudhope and Almeida (2010).

As showed in Figure 2, in which we highlight in blue color the ontology branch, these complexity level starting with free schemes as *Unstructured Texts*, following simple schemes that commonly do not have hierarchies as *Term and/or Concept Lists*. In incrementing the representation complexity by adding some relationship expressiveness, we can find the *Concept and Relationship Structures*, and finally, the most sophisticated schemes that involve *Concepts, Relationship, and Layout* (Souza; Tudhope; Almeida, 2010, 2012). According to Souza; Tudhope and Almeida (2010), "[…] *formal Ontologies allow the representation of all sorts of relationship types, depending on the expressiveness of the language used for representing them*".

Finally, different authors from Information Science consider the ontologies as a tool for knowledge organization, in other words, ontologies are a type of knowledge organization systems (Almeida, 2013, Fonseca, 2007, Søergel, 1999, Souza; Tudhope; Almeida, 2010, 2012, Vickery, 1997).

2.2 Interoperability essentials

Starting a discussion on interoperability, the data management body of knowledge defends that data integration and interoperability became relevant knowledge area in the last decade given the emergence of Big Data. Big Data solutions focus in integrate various types of data, structured data (stored in databases) and unstructured data (texts, audio, video, and streaming). However, data integration and interoperability depends on the data semantics or meaning, once Big Data deals with structured and unstructured data, and the data schema is not precisely defined (Dama, 2017).

> *"Data Integration and Interoperability (DII) describes processes related to the movement and consolidation of data within and between data stores, applications and organizations. Integration consolidates data into consistent forms, either physical or virtual. Data Interoperability is the ability for multiple systems to communicate (Dama, 2017, p. 269)."*

Presenting some basics concepts on interoperability, we start with the Institute of Electrical and Electronics Engineers (IEEE) Standards Glossary[10] defining "*interoperability is the capability of two or more tools to function together in the absence of additional endeavour to make its capability possible*". The Merriam-Webster Dictionary[11] defines interoperability as "*a capacitance in which system have to function with other systems or by using pieces or fragments of other systems*". Finally, the Online Dictionary for Library and Information Science (ODLIS)[12], defines the term interoperability as: "The capability of a computer hardware or software system to communicate and work effectively with another system in the exchange of data, usually a system of a different type, designed and produced by a different vendor".

Indeed, both in the literature of Information Science as in the Computer Science one can find examples of authors that consider interoperability, in a simple way, as the ability of a system has to share and exchange information and applications with another system. This subsection is based in the follows references: Arms (2000), Arms *et al.* (2002), Bishr (1998),

---

[10] Retrieved at 10/17/2016 from: <https://www.ieee.org/education_careers/education/standards/standards_glossary.html>

[11] Retrieved at 10/17/2016 from: <http://www.merriam-webster.com/dictionary/interoperability>

[12] Retrieved at 01/04/2018 from: <https://www.abc-clio.com/ODLIS/odlis_i.aspx>

Guy (2005), Harvey *et al.* (1999), Miller (2000), Ouksel and Sheth (1999), Sheth (1999), Ukoln (1999, 2005).

In short, interoperability is the ability of a system (computerized or otherwise) to communicate transparently (or as closely as possible) to another system (or not). Interoperability should be seen as an activity of organization in which are continually seeks to foster reuse mechanisms of existing resources in the organizational environment, either internal or external resources to the organization. One important goal of interoperability approaches consist in the development of useful solutions that enable information users to reach information, considering that each user has specific needs and abilities, and that they could be located in several organizations and places.

To understand each level of interoperability, one should examine the types of heterogeneity that might exist between different systems: information heterogeneity and system heterogeneity. The system heterogeneity still can be divided in the heterogeneity of information, information system, or platform. Table 2 presents the relation between each interoperability level and the heterogeneity type. Considering the comprehensiveness of the interoperability, the relations showed in Table 2, address the heterogeneity core that each type of interoperability should deal.

Table 2 - Levels of interoperability by types of heterogeneity

Type of Heterogeneity		**Level of interoperability**
General type	**Specific type**	
Information	Semantic Heterogeneity	Semantic
	Structural Representational Schematic	Structural
	Syntax Format	Syntactic
Information System	Digital Media Repository Management Systems Database Management Systems	System
Platform	Operating Systems Hardware	

Source: Adapted from Figure 1 of Sheth (1999).

Table 3 summarizes the explanation of each interoperability level. Syntactic interoperability refers to the adoption of different models or languages in a system. Structural interoperability considers the differences of data structure. Semantic interoperability means that systems are divergent in understanding or interpreting the same meaning against the information exchanged. According to Euzenat (2001), semantic interoperability is the ability to interpret knowledge imported from other languages at the semantic level, that is, to assign to each imported knowledge the correct interpretation or set of models. Semantic interoperability tries to guarantee that the interpretation or meaning of imported and transformed knowledge endures the same across languages.

Table 3 - Levels of interoperability

Interoperability level	Explanations
Technical interoperability	The ability of exchanging data between two or more different information systems and using these data to perform some function without the need of human intervention. It involves the adoption of communication, transport, storage patterns.
Semantic Interoperability	The capacity of two or more systems to exchange data between them and, understanding these data as the same meaning. The exchanged data is precisely interpretable by all information systems with the same meaning without human interference. It involves the adoption of controlled vocabularies, for instance, thesauri and ontologies.
Political-Human Interoperability	Involves the decision around how to share data, in which to make the data available, who can access the data shared, and what data to share. This considers organizations involved in data sharing as a whole, their employees or service providers, and the end users.
Intercommunity interoperability	It addresses access to data originated in several sources and communities that are often limited by organizational controls. It refers to data interaction between different domains and community actors of different nature, such as organizations, professionals, and end users.
Legal Interoperability	Involve an adherence to the legal requirements and implications to make the data free and widely available.
International interoperability	Involves cooperation on an international scale, in which the exchange involves a great diversity of standards and rules, as well as inherent problems of communication created by language barriers.
Organizational interoperability	It refers to the organizational context of each interacting organizations to be able to perform their tasks seamlessly together. This involves standardizing workflows, business goals, relations, and culture.

Source: Prepared by the author based on Miller (2000), Guy (2005) and Sheth (1999).

In order to foster interoperable services is desirable to define technical, content, and organizational agreements. Technical agreements establish technical issues to enable the messages exchanging, such as communication protocols, formats, and security systems. Content agreements seek to promote semantic interoperability by establishing patterns of knowledge representation and organization such as metadata. Ultimately, organizational agreements aim to reduce political differences among organizations by defining common standards and technologies for regulating activities around data such as access, maintenance, and authentication. A desirable interoperable environment involves the adoption of a single standard at each interoperability level by all participants.

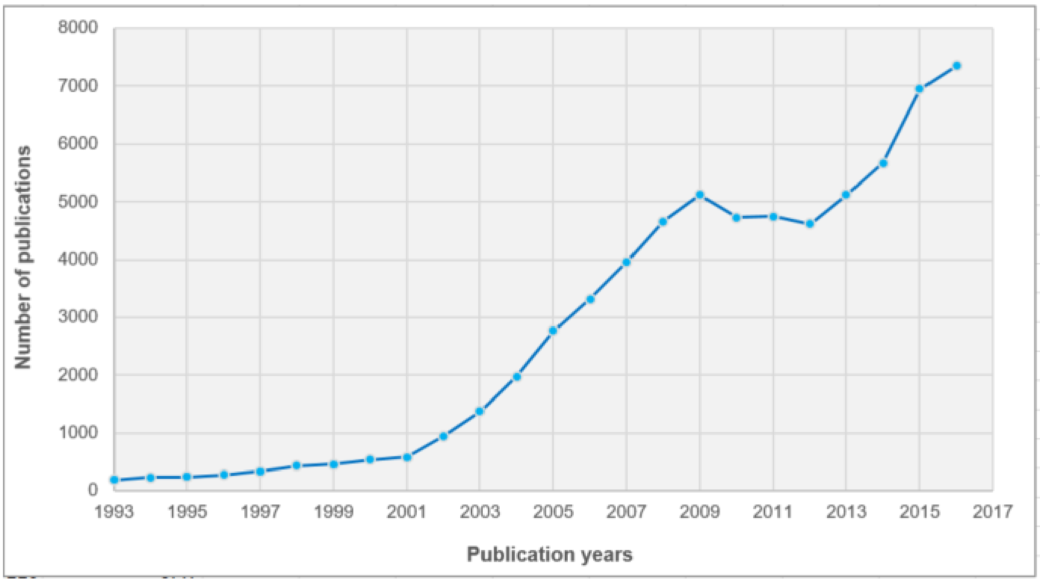
To conclude this subsection, we mention Smith and Welty (2001), whose emphasize the inconsistency in modeling during the early years of conceptual modeling and pointing such inconsistency as one of the causes for current interoperability problems between systems. These authors remark the adoption of ontology-based model to address the semantic interoperability problem.

2.3 Ontology fundamentals

In the last decades, the research on Ontology received much attention due to the possibilities it offers for the representation and organization of information. A bibliometric analysis of literature published during 1900-2012 from the Web of Science database concluded that ontology research has been growing, as well as the participation and collaboration in research (Qiaoli *et al.*, 2015). Smith mentions the growing of publications and conferences on ontology due to ontology application in fields related to Computer and Information Science (Smith, 2003c).

In fact, the research around the term "ontology" has been rising significantly since 2007. Figure 3 showed a result of a non-methodological search at the *Web of Science Database*, performed in the beginner of 2017 using "TS=ontolog*"[13] and a timespan from 1900 to 2016 as search parameter. Notice that a search at the *Web of Science Database* performs simultaneous searching for thirteen primary databases such as *Web of Science Core Collection Database*, *BIOSIS Citation Index*, *Current Contents Connect, Data Citation Index, Derwent Innovations Index, KCI-Korean Journal Database, MEDLINE, Russian Science Citation Index, SciELO Citation Index, Zoological Record*.

Figure 3: Number of publications on ontology by year of publication



Source: Prepared by the author.

[13] TS=ontolog* (TS mean Topic and indicate that the search will be performed in the "Topic fields" of all databases. "Topic fields" are Titles, Abstracts, Keywords, and Indexing fields. The use of asterisk character means that any group of characters could be searched. Thus, this search consider to find at "Topic fields" words like "ontology", "ontologies", etc.)

2.3.1 Ontology as a research topic in Information Science

Ontology has the origin in Philosophy as a study of the nature of being and the relations of existence. It concerns to the 'first philosophy' of Aristotle and his followers. A prominent philosopher in the literature on ontology, Barry Smith, published in 2003 his clear view of Ontology as a philosophical discipline that "[...] *is the science of what is, of the kinds and structures of objects, properties, events, processes, and relations in every area of reality*" (Smith, 1978, 2003b).

Although the research on ontology originates in Philosophy, rapidly, ontology took its place as a research theme in fields such as Computer Science, Information Science, Engineering, Geography, Medicine, among others. Research on ontologies pervades the use of ontology for knowledge representation, to allow computer reasoning and information integration, to foster natural language processing applications, to ensure semantical meaning, and so on (Almeida, 2013, Gruber, 1993b, 2008, Guarino; Giaretta, 1995, Hepp, 2008, Smith, 1978, 2003c, Søergel, 1999, Vickery, 1997).

There are two main arguments supporting the idea of interdisciplinary research involving the term '*ontology'*. The first argument considers Ontology as a philosophical discipline, which aims to understand the existing reality, its relations, and features. The second argument, shared by both computer science and Information Science, in which ontology is an information artifact able to represent some domain of knowledge, such a controlled vocabulary or conceptual model (Almeida, 2013 - Table 1). In this study, the discussion of Ontology (as a philosophical discipline) and ontology (as representational artifact) centers on Barry Smith works that are available in references such as Smith (1978, 1989, 1998a, 1998b, 2003b, 2003c, 2008b), Smith and Mulligan (1983), Smith and Welty (2001), and Arp; Smith and Spear (2015).

The coexistence of interdisciplinary approaches is a key feature of ontology studies, in which exists seven interpretations available for the term ontology (Guarino; Giaretta, 1995):

- A philosophical discipline.
- An informal conceptual system.
- A study related to formal semantics.
- A specification of a conceptualization.
- A representation of a conceptual system via logical theory.
- A vocabulary used by a logical theory.

- A specification (meta-level) of a logical theory.

In the literature of computer and Information Science, the first time that the term 'ontology' appears was in 1967. A work realized by S. H. Mealy on the foundations of data modeling, mentions a distinction of three different domains on data processing field, that are: i) the real world itself; ii) ideas about it existing in the minds of men; and iii) symbols on paper or some other storage medium. In this context, the term refers to a software artifact, written in a formal language with specific applications (Smith, 2003c).

Moreover, ontologies have attracted the interest of numerous researchers of Computer Science, with three main areas: databases, Software Engineering and Artificial Intelligence (Smith; Welty, 2001). In the 1980s, researches in Software Engineering applied ontologies with emphasis on the improvement of conceptual models during the process of developing information systems (Wand; Weber, 1990). In the beginning of 1990s, in order to enable knowledge representation, the research on ontologies as a software artifact grows, mainly in Artificial Intelligence field. In both fields, computer and Information Sciences, ontologies serve as a tool to create models and metamodel of information that aims to represent knowledge in a machine-readable way, or as a categorization model used to index information (Gruber, 1993a, 1995, 2008, Guarino, 1995, 1997b, Guarino; Giaretta, 1995).

One defends a possibility of ontologies use by both human beings and software agents, aiming common understanding about concepts and their relationships in the context of a knowledge domain (Noy; Mcguinness, 2001). In general, a possible use of ontology is to describe the meaning of the symbols used in a system, following a conceptualization of the world in the system context. Ontologies, in computer science, helps to solve problems of the conceptual modeling process, allowing system designers in developing better conceptual models. Additionally, analyzing the influence of the use of ontology in the information systems comes up against two dimensions: the temporal dimension with the use of the ontology in information system at the time of development; and the structural dimension with the use of the ontology as a component of an application (Guarino, 1998a).

In this context, an ontology can function both as a model and as a component of information systems, that is, there are two distinct views: "*ontologies of information systems*" and "*ontologies for information systems*". The first view, ontologies of information system, focuses on the development of ontologies to support the creation of conceptual modeling tools. This view implies in the use of ontology to model information systems at the conceptual level. The second view, ontologies for information system, refers to the application of ontologies to help the development of information systems oriented to ontologies. In this view, an ontology aims to create conceptual schemas and their interrelationships, and it is a component of the information systems (Fonseca, 2007). This differentiation of views is relevant once it guides the choice of the method of building ontologies and the degree of formalism the representation granularity the ontology can cover.

Even in the scope of the Information Science, it is possible to observe the research of ontologies to support basic principles of classification theory and categorization structures, besides the representation of existing knowledge in its entirety, in order to explain the reality in common sense of the society. Currently, several domains use ontologies for their knowledge structuring (Almeida, 2013, Vickery, 1997).

As observed in subsection 2.1, many Information Science authors consider the ontologies as a tool for knowledge organization, in other words, ontologies are a type of knowledge organization systems. Note that the research on ontologies contributes to the improvement of the formal representation of a knowledge domain in Information Science field. Notwithstanding, at the same time, Information Science field promotes a great contribution to the construction of ontologies adding its theories of conceptual analysis, semantic relations, and classification theories (Almeida, 2013, Fonseca, 2007, Søergel, 1999, Vickery, 1997).

Once verifying that the ontologies have been the object of studies in several fields of research, it is admissible a conclusion on the interdisciplinary possibilities of research involving the application of the ontologies. Although, as observed, ontologies do not represent anything new in Information Science, the interest of researchers increasingly arouses. Such interest refers to ontologies capability to contribute on vocabulary disambiguation and to provide structures needed to ensure the semantics understanding of terms. Thus, ontologies serve as solution to problems related to semantic interoperability.

2.3.2 Definition of ontology

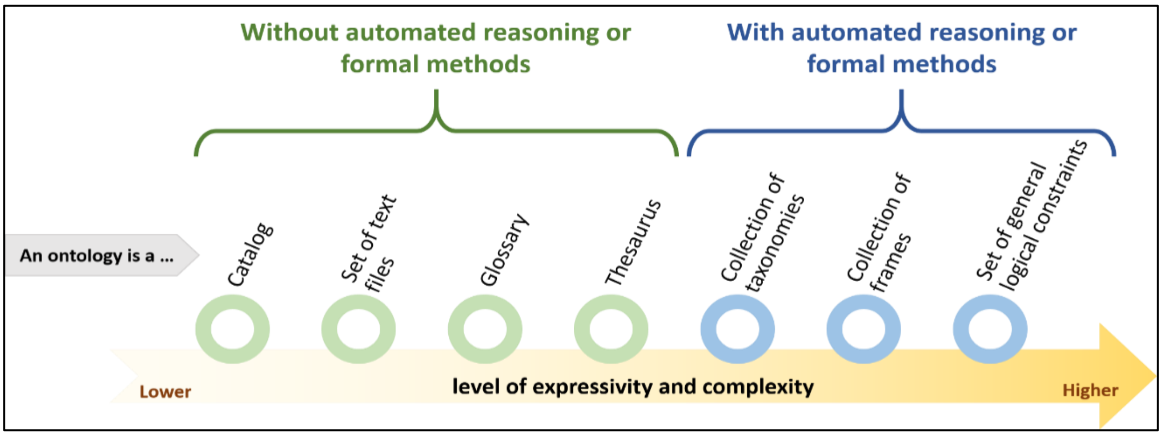
This subsection focuses on presenting the definition of the term ontology in the way that this thesis understands and uses it. First of all, this thesis assumes the following notation: '*ontology*', with lower initial letter, refers to an artifact; and '*Ontology'*, with initial capital letters, refers to a philosophical approach (Guarino; Giaretta, 1995).

On defining ontology, there are several approaches in which permeating different fields of research. Researchers such as Almeida (2013), Fonseca (2007), Søergel (1999, 2002), and Vickery (1997) from Information Science. Gruber (1993a, 1993b, 1995, 2008), Guarino (1995, 1997b, 1998a), Guarino and Giaretta (1995), Guarino; Oberle and Staab (2009), Noy and Mcguinness (2001), Uschold and Gruninger (1996), Smith and Welty (2001), and Hepp (2008) from Computer Science and Artificial Intelligence. Smith (1978, 1998a, 1998b, 2003b, 2003c, 2008b) from Philosophy and Biomedicine. The research field in which originated the definition of the term ontology leads such definition to suit to the context of the field itself.

In order to introduce and clarify these different employments of the term ontology, Figure 4 presented a synthesis by of Smith and Welty (2001) that depict several uses of the term ontology in the literature. According to this spectrum of Figure 4, what is understandable as ontology could mean a different type of artifact although, in general, they aim is to represent knowledge. The variation in these definitions usually happens considering the language applied to knowledge representation, for instance, natural/human language or logical/formal language. The more detailed the represented knowledge is, the more expressive is the artifact (or the ontology considering the spectrum above). In this case, the details could be a term or the name of the represented thing, descriptions of terms, relations between terms, a hierarchy of terms, restrictions regarding terms, and axioms of full first order, higher order, or modal logic. The more expressive the ontology can be the lower is the chance to find ambiguous definitions. Consequently, ontology is more apt to be used by automated reasoning tools (Smith; Welty, 2001).

Complementing this spectrum presented in Figure 4, and in order to help the understanding of ontologies as semantically committed knowledge organization system, we retrieve the study from Almeida; Souza and Fonseca (2011). This study critically evaluates different spectrums that organizing instruments of knowledge organization according to a scale reflecting the commitment of such instruments with the semantic.

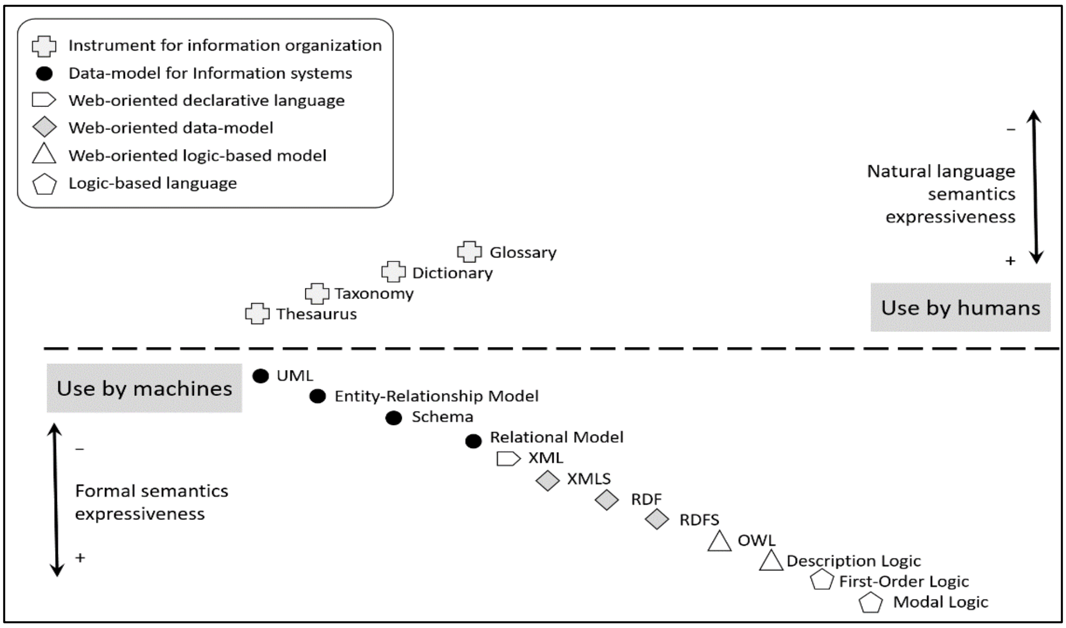
Figure 4: Spectrum of use to ontology term.



Source: Prepared by the author based on Smith and Welty (2001)

Almeida; Souza and Fonseca (2011) proposed a spectrum against a set of instruments of organization/representation of information considering semantic expressiveness to both humans and computers summarized according to the preceding Figure 5.

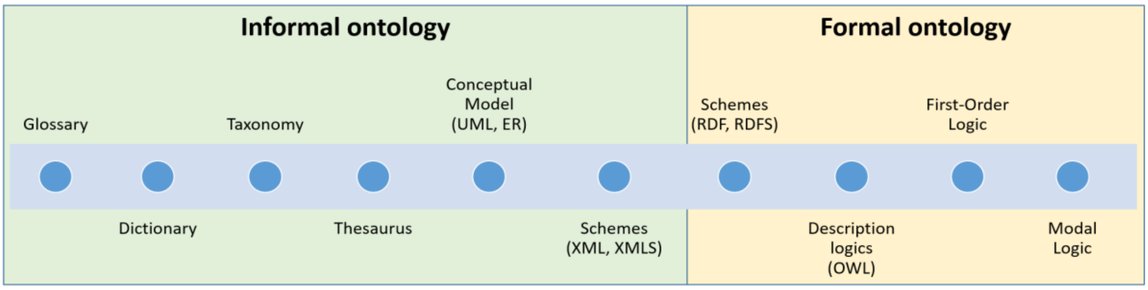
Figure 5: Spectrum of knowledge organization instruments by semantic expressiveness



Source: Prepared by the author based on Almeida; Souza and Fonseca (2011)

We notice that ontologies can be constructed at different levels of formality. Figure 6 presents several types of instruments, classified here as ontology types, formal or informal considering the arguments presented by Almeida; Souza and Fonseca (2011).

Figure 6: Spectrum of information representational instruments by level of formality



Source: Prepared by the author based on Almeida; Souza and Fonseca (2011).

Notwithstanding the foregoing, we mention Husserl conceived the distinction between "formal ontology" and “material or regional ontology". In the case of formal ontology, the produced theories are domain-neutral level, unlike the material ontology that fit a specific domain (Grenon, 2003c, Hennig, 2008). Complementing, Smith (1998b) reveals the existence of two kinds of ontologies:

i) a formal ontology that focuses on the study of existing structures among different scientific areas;

ii) the material ontology that studies the phenomena of a given domain. Nevertheless, there is a complexity to understand the reality of the world and express the truth given independent of domain.

Hence, analyzing the view of a specific domain is possible to create a material ontology to this specific domain. Thus, considering the common issues in a set of material ontologies, a formal ontology comes up given a common-sense view (Smith, 1998b).

Formal ontologies are different from material ontologies because the formal intends to cover any domains of objects without concerns to singular details of a scientific domain specifically. By the way, formal ontology aims to build a representation able to reflect a reality of the universe through the adoption of formal methods approaches, i.e. mathematical logic and formal logic (Almeida, 2013, Guarino, 1995, 1998a, Smith, 1978, 1989, 1998a).

One argues on distinct interpretations to the term ontology that sometimes creates confusion to identify what an ontology is. Therefore, in order to do a good ontology design, it should first distinguish the meaning of ontology in order to bear better decisions during the ontology development and to bypass common pitfalls (Arp; Smith; Spear, 2015, Guarino; Giaretta, 1995).

The most frequent definition of ontology found in literature of Computer Science and Artificial Intelligence fields is "*an ontology is a specification of conceptualization in the context of knowledge sharing*", in which an ontology should be also formal in order to be understandable by machines (Gruber, 1993a, 1995, 2008). However, a critical analysis pointed the notion of *'conceptualization'* in this definition is most fit to represent a state of things than a conceptualization, that is, its extensional interpretation. Indeed, a conceptualization must represent the intended meanings by means of an intentional interpretation. In the case of ontology, it treats the concept independent of its state in the world, that is, by its intentional interpretation (Guarino, 1997b, Guarino; Giaretta, 1995).

Hereafter, considering an ontology construction design, there is a view of ontologies as representational artifacts attempt to represent reality for the purpose of providing the common-sense understanding of the entities in reality that these ontologies represent (Arp; Smith; Spear, 2015). Additionally, Guarino (1998a) considers ontologies as engineering artifact composed by a vocabulary that represents some domain and assumptions about this domain. Thus, ontology is defined *"as a logical theory focused on describing a certain reality through a formal vocabulary maintaining its ontological commitment to the concepts of the described world*" (Guarino, 1998a).

Despite the existence of several definitions for what is understood as ontology, the definition more appropriate for the purpose of this thesis goes beyond the definition adopted by Arp; Smith and Spear (2015). Such definition takes account of an ontology is *"a* *representational artifact that nominates some universals and defined classes as terms or entities by its own taxonomy, besides determines possible relationships among these entities nominated*". This definition takes into account the philosophical realism in which the ontology represents entities of the reality and not concepts. This philosophical context considers that ontology serves to describe the world as it is and not as we think it is (Arp; Smith; Spear, 2015, Smith; Ceusters, 2010).

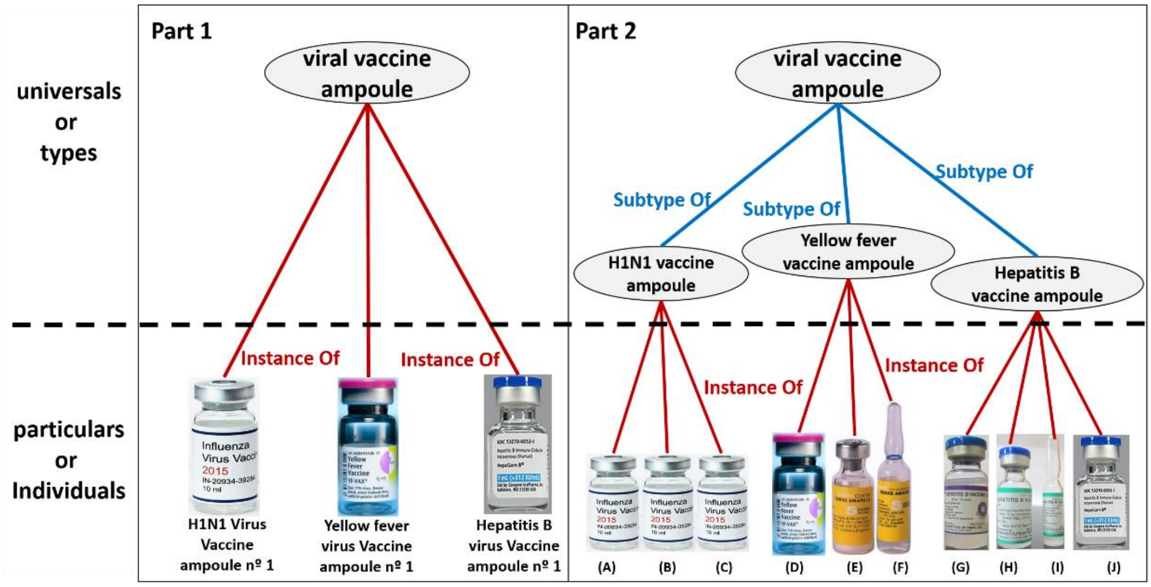
It seems to be that, in a wrong use of this ontology theory view, each person could have a specific view of how the world is and try to describe their own vision through an ontology, using an extensional interpretation of the world. In this sense, several ontologies about the same world could arise, in which each ontology expressing the specific vision of who defined it. Above all, there is only one ontology because there is only one true reality of the world, it means an intentional interpretation of the world. To solve the problem of the plurality of ontologies, Smith (1998b) clarifies using a terminological distinction between referent or reality-based ontology (R-ontology) and elicited or epistemological ontology (E-ontology). R-ontology corresponds to the philosophical approach, which concerns how the universe is organized, what is true of reality independent of the domain. In opposition, E-ontology addresses the task of conceptualizing assigned to a specific domain, i.e., how individual, group, language or science conceptualizes a particular domain (Smith, 1998b). We highlight that this thesis also focuses on the R-ontology terminological sense.

Universals and Particulars

In order to have a better understanding the definition of ontology assumed in this thesis, we describe in this section the distinction between *universals* and *particulars*. In the philosophical context originated in the work of Aristotle, the realism recognizes as core reference the existence of universals related to the existence of some instance or particular exists (Grenon, 2003b, Grenon; Smith, 2004, Lowe, 2006, Smith; Brogaard, 2000, Vezina, 2007).

Particulars or i*ndividuals* are single occurrences of something existing in reality, for example, *you* and *me*, each one of us is a single occurrence or individual of a *"homo sapiens". "This thesis"* is a single occurrence of a *"thesis"* as a general entity that mean, the term *"thesis"* refers to all thesis existing in the world, and *"this thesis"* or *"Fernanda's thesis"* refers to a single thesis. Hence, the terms *"thesis"* and *"homo sapiens"* refer to general entities or *universals.* Therefore, *universals* or *types* are general real entities or patterns existing in the world, which are not beliefs or thoughts. A *universal* exists exclusively if at least one *particular* of this universal exists. Suppose that until the existence of this thesis there was no other thesis in the world. This means that before the existence of this thesis there was none particular for the universal thesis. If the particular have not existed, consequently the universal does not exist (Arp; Smith; Spear, 2015, Grenon, 2003b, Grenon; Smith, 2004, Grenon; Smith; Goldberg, 2004, Smith; Ceusters, 2010, Smith *et al.*, 2006). Figure 7 presents some examples of what are *universals* and what are *particulars*.

Figure 7: Example of Universals and Particulars



Source: Prepared by the author.

Think about vaccines, more specifically "*viral vaccine ampoule*". Does "viral vaccine ampoule" exist? Yes, it exists. We know about the existence of vaccine against some virus, for instance, I already had taken a shot of the following vaccines: H1N1 virus, Hepatitis B virus, and Yellow fever virus. Hence, in the case of Figure 6 Part 1 (left side), "*viral vaccine ampoule*" is a universal (a generic entity) referring to the particular vaccines that I took. Moreover, "H1N1 virus vaccine ampoule nº 1", "Hepatitis B virus vaccine ampoule nº 1", and "Yellow fever virus vaccine ampoule nº 1" are the specific vaccine ampoules used for my vaccination, that is, they are particulars of the universal "*viral vaccine ampoule*".

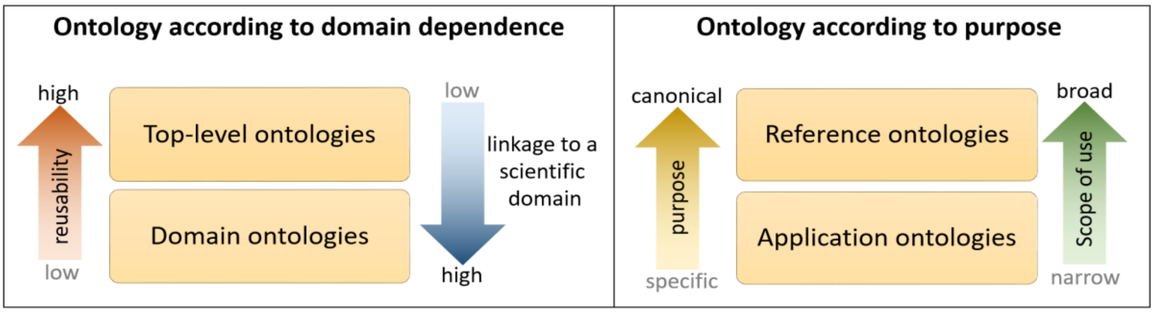
Notwithstanding, I am not the only person vaccinated against these viruses. Many other people should be vaccinated, consequently, these types of vaccines are produced in large quantities and stored in several ampoules. Then, according to Figure 7 Part 2 (right side), there are ten particulars to the universal "*viral vaccine ampoule"*, three ampoules of the H1N1 virus vaccine (ampoules A, B and C), three ampoules of the Hepatitis B virus vaccine (ampoules D, E, and F), and four ampoules of the Yellow fever virus vaccine (ampoules G, H, I and J).

Suppose you want to know about each type of viral vaccine depending on the virus that immunizes, rather than just knowing about the "viral vaccine ampoule" in general. New universals could be created and the ontology would present higher granularity or details level (Arp; Smith; Spear, 2015). Thus, as showed at Part 2 of Figure 7, the universal "viral vaccine ampoule" has three subclasses "H1N1 virus vaccine ampoule", "Hepatitis B virus vaccine ampoule", and "Yellow fever virus vaccine ampoule". All of them are universals and each one has they own particulars: "H1N1 virus vaccine ampoule" has ampoules A, B and C; "Hepatitis B virus vaccine ampoule" has ampoules D, E, and F; and the "Yellow fever virus vaccine ampoule" has ampoules G, H, I and J.

Ontology classifications approaches

In addition, the literature review shows that ontologies received some classifications. One divides kinds of ontology considering the level of domain dependence observing four kinds of ontologies, namely: top-level ontologies, domain ontologies, task ontologies, and application ontologies (Guarino, 1997a). In the same way, there is another classification that assumes the existence of domain ontologies, top-level ontologies, and application ontologies, and that introduces a new kind named reference ontology (Arp; Smith; Spear, 2015). Figure 8 shows a summary of these kinds of ontologies and each one has an explanation below.

Figure 8: Kinds of ontologies summary



Source: Prepared by the author.

Top-level ontologies, also named *upper ontologies*, are neutral domain ontologies, that is, represent general entities that are not specific to a domain, for instance, time, space, object, process (Arp; Smith; Spear, 2015, Guarino, 1997a, Jansen, 2009a). The most known top-level ontologies are:

- *Basic Formal Ontology* (BFO) (Arp; Smith; Spear, 2015, Smith *et al.*, 2015).
- *Descriptive Ontology for Linguisitics and Cognitive Engineering* (DOLCE) (Gangemi; Guarino*; et al.*, 2002).
- *Unified Foundational Ontology* (UFO) (Guizzardi; Wagner, 2005, 2010).

- *General Formal Ontology* (GFO) (Herre, 2010, Herre *et al.*, 2006).
- *Suggested Upper Merged Ontology* (SUMO) (Pease; Niles; Li).
- *Yet Another More Advanced Top-level Ontology* (YAMATO) (Mizoguchi, 2010).

If you want to take a look at a comparison between these upper level ontologies you would see (Mascardi; Cordì; Rosso, 2007).

Domain ontologies are ontologies with focus on domain vocabulary aim to represent entities of a specific scientific domain such as medicine or military. *Task ontologies* resemble domain ontologies but their focus is the functionality of a domain such as diagnostics or purchases (Guarino, 1997a). However, Arp; Smith and Spear (2015) did not do this distinction, since they argued the existence of the domain ontology with the focus on representing the entities or universals of a specific scientific domain. Usually, the construction of good domain and task ontologies ground under a top-level ontology (Arp; Smith; Spear, 2015, Guarino, 1997a). Some examples of domain ontologies built under BFO top-level ontology are (Arp; Smith; Spear, 2015):

- Gene Ontology (GO).
- Foundational Model of Anatomy (FMA).
- Environment Ontology (EnvO).
- Chemical Entities of Biological Interest (ChEBI).

Application ontologies are ontologies that represent entities specific from an application or purpose in an use case scope that aims to describe. Usually, their entities are specializations of the domain and task ontologies entities applied to this specific use case instead of a general purpose (Arp; Smith; Spear, 2015, Guarino, 1997a, Jansen, 2009b).

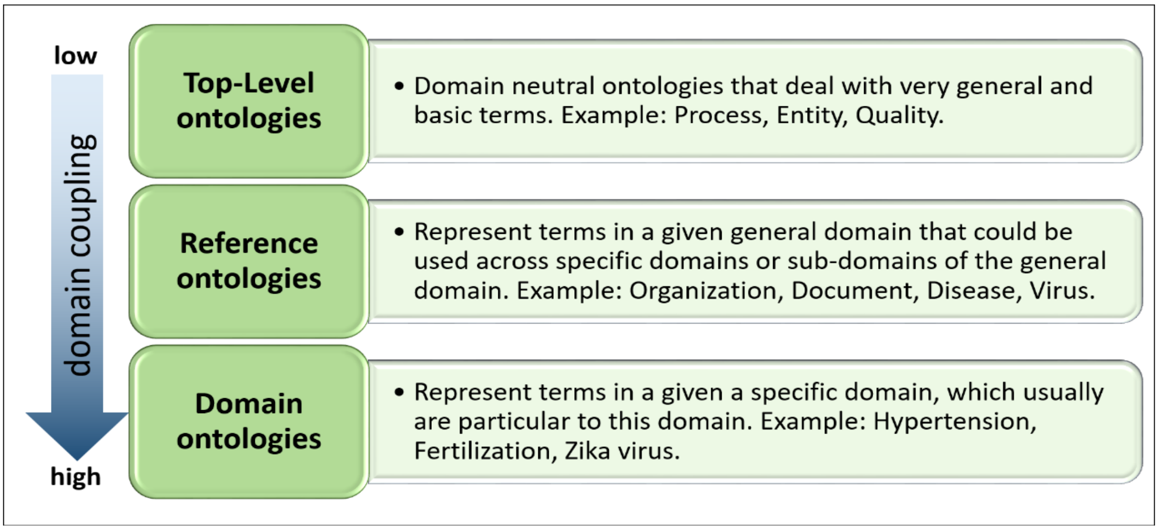
*Reference ontologies a*lso named *core ontologies* or *medium-level ontologies*, aims to describe a scientific domain as comprehensive and accurate as possible. These ontologies embed what is known about such a domain and thus constitute a reference of the domain knowledge (Arp; Smith; Spear, 2015, Jansen, 2009b). Some examples of this kind of ontology are (Arp; Smith; Spear, 2015):

- Foundational Model of Anatomy (FMA).
- Gene Ontology (GO).
- Cell Ontology (CL).
- Protein Ontology (PRO) (Arp; Smith; Spear, 2015).

We notice that according to these examples of domain and reference ontologies some of them such as Gene Ontology appear both as domain and reference ontology. This occurs because the kinds of ontology given by Arp; Smith and Spear (2015) considers this classification as two aspects: purpose and generality. According to ontology goal or purpose, the authors classify an ontology as an application or a reference ontology. Analyzing the generality of the subject or dependence of scientific domain that an ontology covers, the authors distinguish an ontology as a domain or a top-level ontology (Arp; Smith; Spear, 2015).

There are not consensuses on ontology classification. Each seminal author has his own way to classify ontologies. Sometimes, an author adopts the same classification label but with a different meaning. However, this thesis does not include a long discussion of ontology classification. Thus, compiling the presented previously ideas about ontology classification, we created a summary of the kinds of ontology, as shown in Figure 9.

Figure 9: Schema of ontology classification adopted by this thesis.



Source: Prepared by the author.

Since we established the definition of ontology and clarified the different types of ontologies adopted in this study, now, it is necessary to clarify what are the elements that are part of an ontology.

Components of ontology

We begin this discussion by mentioning Guarino and Welty that present taxonomy, concepts, properties of concepts, relations between concepts, and algebra as elements of an ontology (Guarino; Welty, 2000, Welty; Guarino, 2001). Gruber (1993a, 1995) suggests five components existing in an ontology to represent some knowledge: concepts, relations, axioms, instances, and functions. According to the definition of ontology given by Arp; Smith and Spear (2015), a taxonomy, universals, defined classes, and relations are the elements of an ontology.

Considering the Web Ontology Language (see subsection 2.5), an ontology has individuals, properties, and classes. Properties are binary relations or restrictions that can be used to establish relationships between individuals (object properties) or between individuals and data values (datatype properties) (W3c, 2004a).

In order to better understanding ontologies and the process of ontologies creation, Table 4 presents a summary of the basic elements of an ontology. This table is a compilation of the follows documents: Arp; Smith and Spear (2015), Gruber (1993a, 1995) Guarino (1997a), Noy and Mcguinness (2001), Uschold and Gruninger (2004); Guarino and Welty (2000), Welty and Guarino (2001).

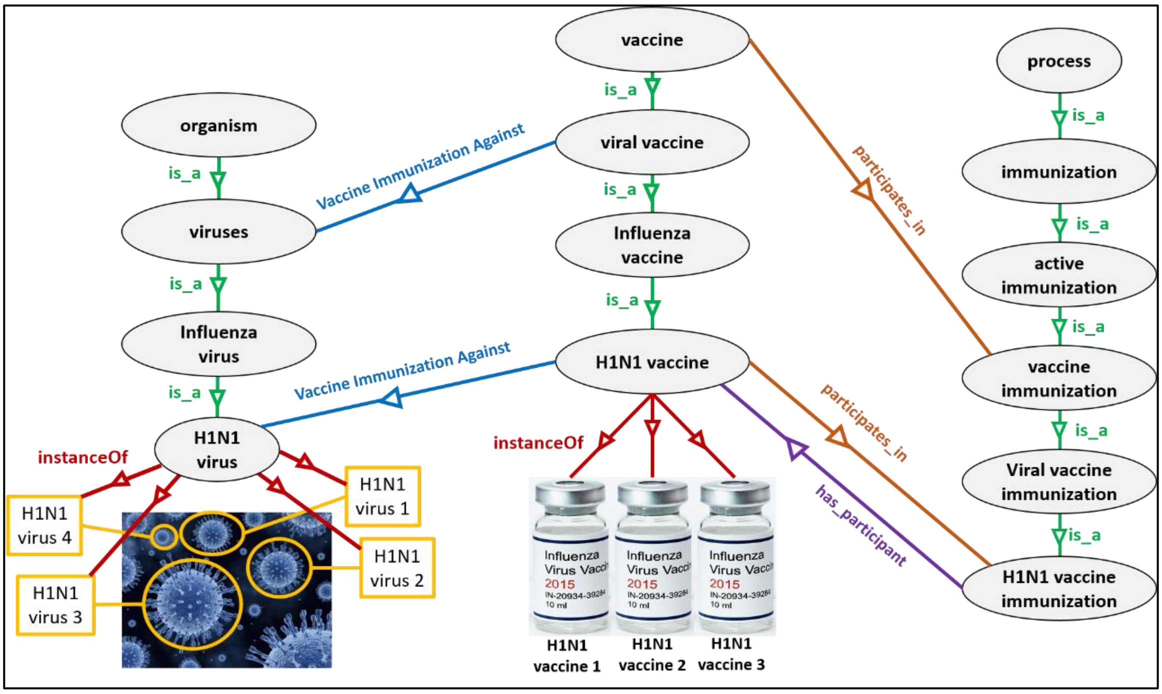
Table 4 - Summary of the basic ontology elements and its respective description

Element	**Description**
Entity	It is something that you want to represent on a particular domain. Anything that exists or has existed or will exist. It could be events, process, inanimate or living objects, qualities, and so on.
Classes	Represent entities of the domain. Elements that organize the entities of a domain in a taxonomy.
Class attributes	Relevant properties of the class that helps to describe it.
Relationship or relation (also known as property)	Describe the type of interaction between two classes, two instances or one class and one instance. Relationships express the nature of the connection between two concepts and may have cardinality.
Cardinality between classes or entities	A measure of the number of occurrences of one entity associated to a number of occurrences in another.
Instance or individual	It is used to represent a unity of objects specific to an entity.
Instance attributes	These are relevant properties that describe the instances of an entity.
Axiom	A statement or proposition represented in a logic pattern that is taken to be true. They restrict the interpretation and use of the classes involved in the ontology.
Rules	They impose conditions to the domain and enable the inference of values for attributes.
Constant	Kind of property that always has the same value. Rules or formulas usually use it to infer knowledge in the ontology.

Source: Prepared by the author.

According to this view, in Figure 10 below we exemplify the elements of an ontology. The oval shaded figures represent the classes, in total there are 14 classes, for instance, "*H1N1 virus*", "*H1N1 vaccine*" and "*H1N1 immunization*" classes. There are four instances of "*H1N1 virus*" class ("*H1N1 virus 1*", "*H1N1 virus 2*", "*H1N1 virus 3*", "*H1N1 virus 4*") and three instances of "*H1N1 vaccine*" class ("*H1N1 vaccine 1*", "*H1N1 vaccine 2*", "*H1N1 vaccine 3*"). In addition, there are some relations. The relation represented in red ("*instanceOf*") is a relation between a class and an occurrence of this class. The relations represented in green ("*is_a*" same as "*subClassOf*"), blue ("*Vaccine Immunization Against*"), purple (“*has_paticipant*”), and orange ("*participates_in*") are relations between classes. You could notice that the relations “*has_paticipant*” and "*participates_in*" are inverse relations.

Figure 10: Piece of ontology with basic elements of ontology.



Source: Prepared by the author.

Moreover, usually relations have a related axiom, for example, the relation “*instanceOf*” is an instantiation relation between *particular* and *universal*. Smith (2003a) and Smith and Rosse (2004) discuss in-depth the relation “*instanceOf*” arguing that is a primitive relation and has the following axiom related: “*(1) that it holds in every case between an instance and a class, in that order; and (2) that nothing can be both an instance and a class*” (Smith; Rosse, 2004, p. 445).

The formalization of this axiom in a first-order language conform the following expression:

$$e \; instanceOf \; e' \quad \Rightarrow \quad \forall e \forall e' \left(inst(e,e') \rightarrow p(e) \wedge u(e') \right)$$

Wherein ***e*** and ***e’*** are variables to instance and class respectively, and ***inst***, ***p*** and ***u*** are function to ***instance***, ***particular*** and ***universal*** respectively, the symbol ∧ means conjunction, ∀ is the universal quantifier, → is an implication.

Regarding the Figure 10, we could understand some assertions such as:

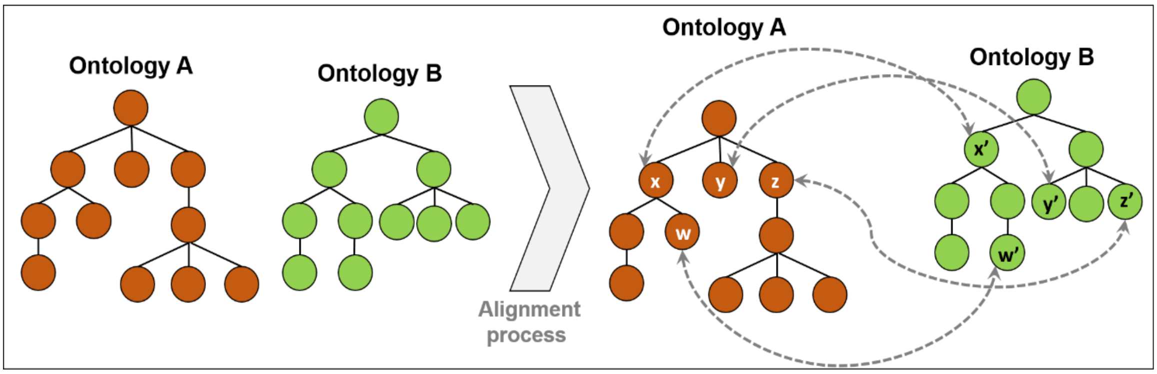
- “H1N1 virus” *is_a* “Influenza vírus”.
- “H1N1 virus 1” “*instanceOf*” “H1N1 virus”.

- “H1N1 vaccine” “*Vaccine immunization agaist*” “H1N1 virus”.
- “H1N1 vaccine” “participates_in” “H1N1 vaccine immunization”.

2.3.3 Interoperability between ontologies

Some scenarios proposed by methodologies to build an ontology predict needs to reuse existing ontologies. Issues related to the relationship between two or more ontologies consist in interoperability between ontologies. Different approaches arose to provide ontologies reuse and interoperability between ontologies such as alignment, mapping, combination, and integration.

Figure 11: Ontology Alignment Approach

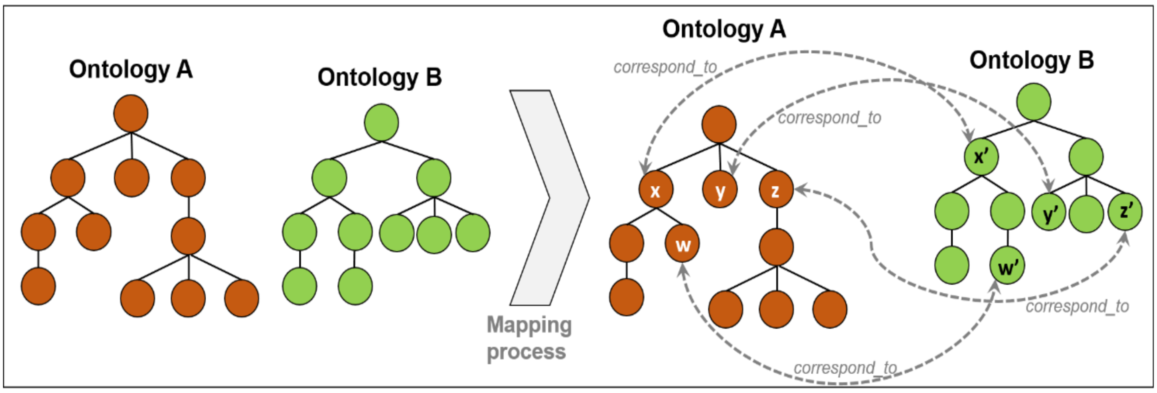


Source: Prepared by the author.

Ontology alignment (Figure 11) consists of finding, in two or more distinct ontologies, the link between ontology entities. This approach aims to determine links among similar terms of these ontologies, allowing them to reuse each other's information. Ontology alignment results in the original ontologies stay separated, connected by the links of correspondence created during the alignment process. This approach is usually applied when both original ontologies cover complementary domains (Ehrig, 2006, Noy; Musen, 1999, 2000, Stoutenburg, 2009).

An ontology mapping approach aims to define a formal structure, which includes the correspondence among two or more ontologies with the aim to establish expressions that connect concepts of the original ontologies (Figure 12). Such a mapping can be used to transfer data instances, integration schema, combination schema, and similar tasks. The mapping approach is different of the alignment approach insofar as the former considers that mapping connections are based on formal expressions (Choi; Song; Han, 2006, Ehrig; Sure, 2004, Kalfoglou; Schorlemmer, 2003, Noy; Musen, 1999, 2000)

Figure 12: Ontology Mapping Approach

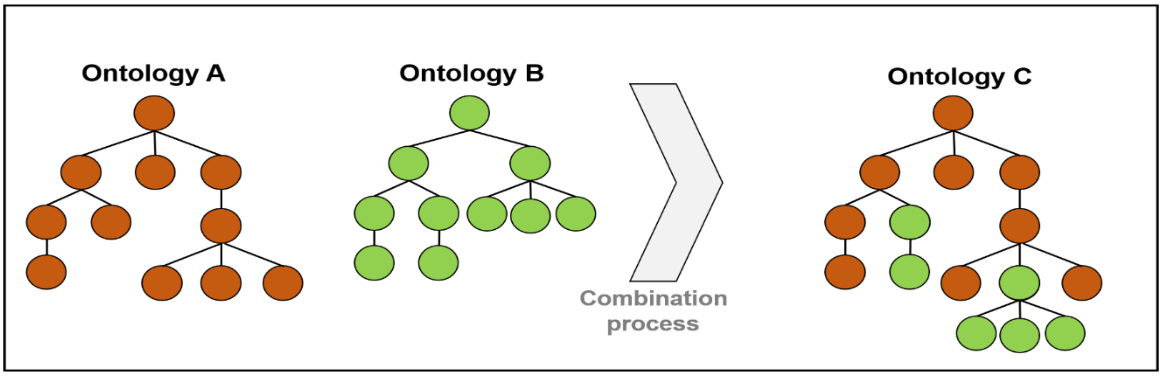


Source: Prepared by the author.

Considering two ontologies (e.g. A and B), according to Ehrig (2006), the aligning of ontologies happens by the creation of correspondence between an element (concept, relation, or instance) from ontology A to an element of the ontology B, "*which has the same intended meaning* (Ehrig, 2006, p. 19)" (Figure 12).

In the ontology combination approach, the original ontologies A and B are combined into a single ontology with all their terms together, generating a third ontology C, as shown in Figure 13. This approach is usually applied when the original ontologies refer to a similar domain or when they has some type of overlap (Ehrig, 2006, Noy; Musen, 1999, 2000, Stoutenburg, 2009).

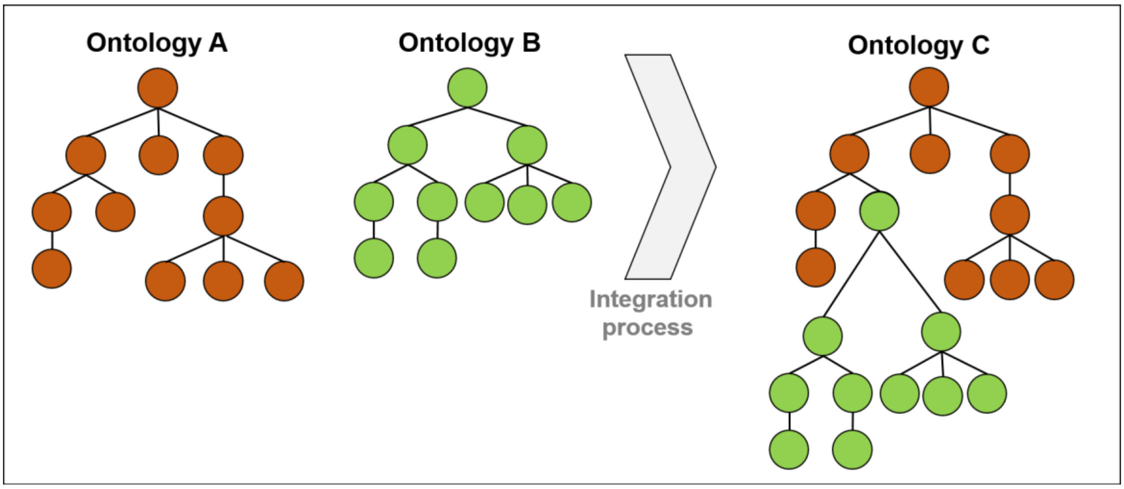
Figure 13: Ontology Combination Approach



Source: Prepared by the author.

The approach to ontologies integration aims to generate a single ontology from the assembling, extension, specialization, or adaptation of other ontologies that not always deal with the same subject (Figure 14). The difference between this approach and the combination approach is the possibility of identifying the parts that were created from the original ontologies. In addition, while in the combination approach the original ontologies deal with the same domain, in the integration approach the ontologies can be about distinct subjects (Calvanese; De Giacomo; Lenzerini, 2001, Pinto; Gómez-Pérez; Martins, 1999, Pinto; Martins, 2001).

Figure 14: Ontology Integration Approach



Source: Prepared by the author.

These four approaches to ontologies interoperability, which generate different results, are concerned with the definition of similar terms in the original ontologies. During this research, we realize the importance of understanding these four approaches to define the ontology reuse strategy.

2.3.4 Ontologies providing semantic interoperability

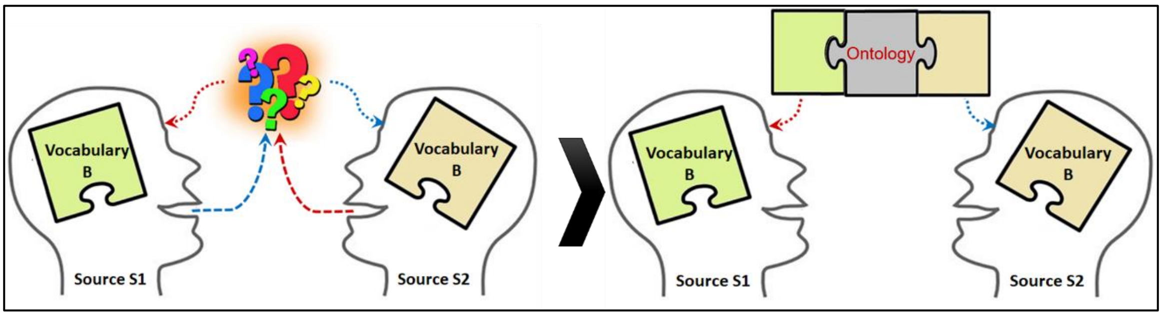
During the literature review, we found several documents approaching the relationship between ontologies and semantic interoperability, such as Bittner; Donnelly and Winter (2005), Euzenat *et al.* (2002), Gangemi; Fisseha*; et al.* (2002a), Liyanage; Krause and De Lusignan (2015), Obrst *et al.* (2006), Orgun *et al.* (2008), Ouksel and Sheth (1999); and Obrst (2003). There is a consensus of the ontology capability to improve the semantic interoperability among information systems. These authors discuss the role of ontologies supporting semantic interoperability in different areas of knowledge.

All discussion starts with the understanding between agents (sender and receiver) in a communication process. The mathematical theory of communication presented by Weaver (1949, 1953) helps to understand the communication process, both between systems or people. Communication process happens from information exchange between agents, a message sender agent, and a message receiver agent. During communication, a language expresses the message exchanged. A language syntax consists in a set of organized symbols and a set of rules to define how to organize these symbols. Additionally, to syntax, a language also involves the semantic aspect that is the meaning of the symbols itself and the set of symbols organized in a context (Weaver, 1949, 1953).

Thus, in order to have an effective communication both sender and receiver should have the ability to interpret the language used to express the message communicated. Communication problems caused by semantic incompatibility, for instance, the existence of different symbols having the same meaning or the same symbol with different meanings are known as polysemy. When polysemy occurs, a recursive interaction between two human agents solves the communication issues, in natural language, between the sender and the receiver. However, when agents are computational systems that act without human intervention this is not possible. Different from humans, computational systems by itself do not have the native ability to interpret the symbols throughout the communication (Weaver, 1949, 1953).

Since semantic interoperability in information systems is the communication capacity among two or more systems, the target system should receive and understand the message sent by the origin system with the same meaning or semantic. Exemplifying, consider two systems or information sources S1 and S2, and each one has its own language (A and B respectively). During a communication process between these two information sources, a semantic problem happens: S1 does not understand language B, and S2 does not understand language A (Figure 15). In order to enable a semantic understanding between S1 and S2, it is necessary to define a language that is more expressive than languages A and B themselves.

Figure 15: A scheme of the ontology role in semantic interoperability.



Source: Prepared by the author.

This common language must explicitly express the meaning of terms and avoid ambiguities inherent to natural language (Jasper; Uschold, 1999). At this point, formal ontologies can be used with advantages. A formal ontology can integrate system S1 and S2 by offer a semantic pattern able to link languages A and B.

A solution to semantic interoperability using ontologies considers the capability of ontologies to specify a semantic link between different terminologies or vocabularies, assuming the role of "*lingua franca*". The mappings between specific terminologies or languages of the systems involved in communication are indirect, that is, a mapping of each terminology to or from the reference terminology. Since formal ontologies also specified the semantics of the specific terminologies, mappings can then be computed automatically (Bechhofer; Goble, 2001).

Moreover, when a formal ontology specifies the semantics of vocabularies underlying information systems, terms and relationships between vocabulary terms appear as variables, constants, and predicates of a formal logical language. Logical axioms are added to the ontology, in order to express relations between entities represented therein. Axioms are a means of restricting the meaning of the terms: certain interpretations are accepted and others rejected. In this way, the semantics of each term is defined explicitly and without ambiguity (Guarino, 1998a, 1998b). It is worth remembering that realism-based formal ontologies specify the semantic of terms that represent entities of the reality existing in different domains of knowledge (Arp; Smith; Spear, 2015, Smith; Ceusters, 2010).

This approach of the mappings generation by the systems involved in the communication process, in others words, automatic translation between the semantics of the terminologies involved, is an essential solution to semantic web success. Noticed that the formal ontologies provide efficiency in the exchange of information between systems, by promoting the semantic interoperability between them.

Thus, a detailed analysis of the use of ontologies to support interoperability can be found in Bittner; Donnelly and Winter (2005), Mena *et al.* (1998), Wiederhold and Jannink (1999), and Gangemi; Fisseha*; et al.* (2002b).

2.3.5 Understanding the Basic Formal Ontology

Basic Formal Ontology (BFO) is a top-level ontology that follows philosophical theories from Aristotle and Edmund Husserl. BFO development started in 2002, at the Institute for Formal Ontology and Medical Information Science (IFOMIS). BFO is a small (35 entities currently) formal and domain-neutral ontology designed to represent general entities that do not belong to specific domains (Arp; Smith, 2008, Smith *et al.*, 2015). The foundational theories that sustain BFO were originated mainly from researches of both Barry Smith and Pierre Grenon, and some important references are Arp and Smith (2008), Arp; Smith and Spear (2015), Bittner and Smith (2003), Grenon (2003a, 2003d), Grenon and Smith (2004), Grenon; Smith and Goldberg (2004), Smith (2004, 2012), Smith and Grenon (2004).

BFO deals with the concepts of universals and particulars according described at subsection 2.3.2. According to BFO, everything that exists is a type of *entity* (Arp; Smith, 2008, Smith *et al.*, 2015). In addition, based on Grenon and Smith (2004), the BFO entities are distinguished in *continuant* and *occurrent* based on “[…] *the existence mode in time of entities existing in the world* (Grenon; Smith, 2004, p. 139)”.

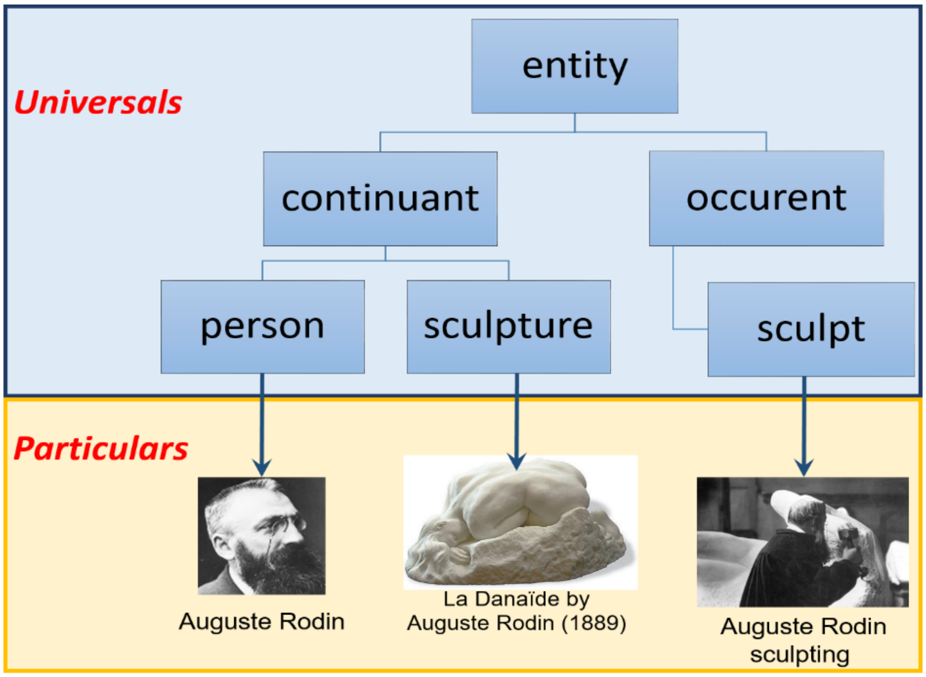
> *“BFO endorses first of all a view according to which there are entities that have continuous existence and a capacity to endure (persist self-identically) through time even while undergoing different sorts of changes (Grenon; Smith, 2004, p. 139)”.*
> *“In addition, however, BFO endorses a view according to which the world contains occurrents, more familiarly referred to as processes, events, activities, changes. (Grenon; Smith, 2004, p. 140)”.*

Continuants are entities of the existence that persists over time, that is, entities able to existing independently of us, may pass through changes, but maintaining their identity preserved. Cells, the Moon or Earth, a vaccine ampoule, a person and a sculpture, to mention a few examples of universals of BFO type *continuant*. In contrast with continuants, *occurrents* are entities in which the existence manifests in time, but not all the time. Occurents reflect processes, events, activities, changes, or happenings. For instance, the orbiting of planets, a beginning of football game, adulthood or childhood of an individual, a baby's childbirth, a horse race (Arp; Smith; Spear, 2015, Grenon; Smith, 2004, Grenon; Smith; Goldberg, 2004, Smith; Kumar; Bittner, 2005).

In order to provide a better understanding of the BFO dichotomy of entities as continuants or occurents, we created the Figure 16. As example of particulars from the *continuant* subtype *person,* we could mention me, you and the sculptor *Auguste Rodin* (Figure 16). In the case of the subtype *sculpture*, the sculpture *La Danaïde* from *Auguste Rodin* (Figure 16) or *Pietà* from *Michelangelo* are suitable examples.

Think about Pietà, the famous Michelangelo sculpture, which is on displayed in St. Peter's Basilica at Vatican, finished by the artist at the end of the 15th century. Unfortunately, in 1972 the sculpture suffered vandalism that damaged its nose, one of its eyelids and the left arm of Our Lady. Although the sculpture today is different from the original sculpture due to the damage suffered, the Pietà today is the same one as in 1972 and is the same one that Michelangelo sculpted. Therefore, the identity of the sculpture persisted despite it having undergone physical changes throughout its existence. This example illustrates the theory behind the continuants.

Figure 16: The central axis of the BFO ontology.



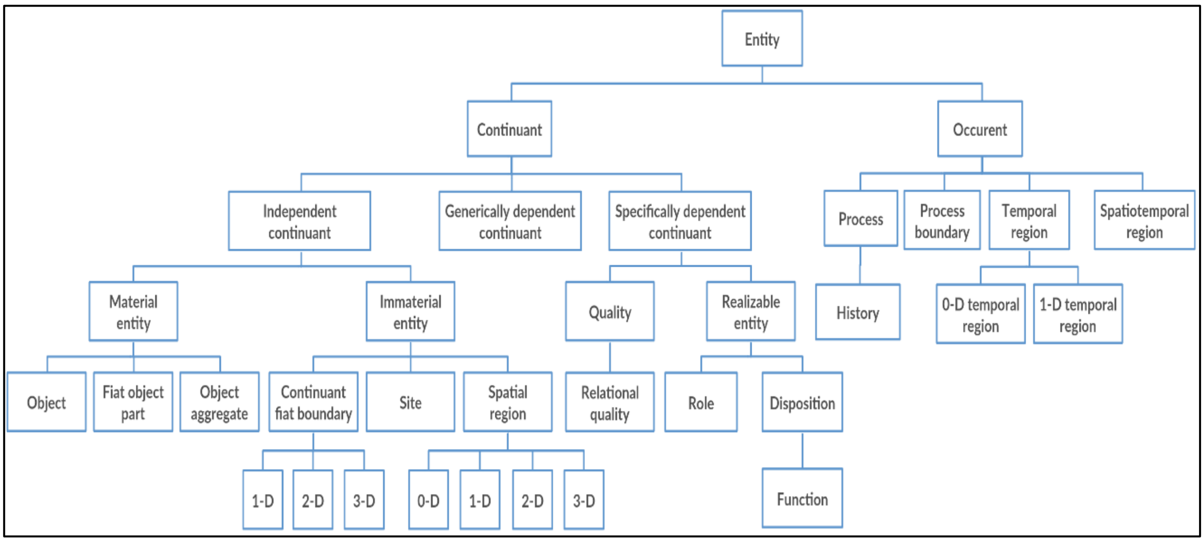
Source: Prepared by the author.

As examples of particulars from the occurrent subtype *sculpt*, imagine the action of *Auguste Rodin* sculpting one of his sculptures (Figure 16), or the sculpting process of *Michelangelo* to create *Pietà.* Another example of an occurent is a football match, for example, the FIFA World Cup final match, which the result determines the champion team. This football match happens each four years, which means that it is not a persistent entity, but rather an event that happens from time to time.

Another example is the childhood and adolescence of an individual, which are stages of the individual's life delimited by certain ages that the individual attains. For instance, today I am living my adulthood, but in the past of my life, I lived my childhood and adolescence. Childhood, adolescence, and adulthood are stages of individual's life, which happen in a specific time and do not persists all individual's life.

This two type of entities constitutes two views, or perspectives, of the same reality which although distinct complement each other (Arp; Smith; Spear, 2015). For example, it is possible examines an individual by itself (continuant) or an individual's life stages (occurent). This distinction between continuants and occurrents determine the central axis of BFO ontology, and each of these types subdivides into others, as shown in Figure 17. For further details on each subdivision of continuants and occurrences, we suggest reading chapters 5 and 6 of Arp; Smith and Spear (2015) and section 3 of Smith *et al.* (2015).

Figure 17: The hierarchy of BFO entities



Source: Based on Smith *et al.* (2015)

In addition, the interaction between two BFO elements of the ontology happens by means of ontological relationships specified in Relation Ontology (RO). RO relations are binary relations that determine the interaction between two elements of the ontology. There are three types of primitive binary relations: the class-class relation, the instance-class relation, and the instance-instance relationship (Arp; Smith; Spear, 2015, Smith *et al.*, 2015, Smith *et al.*, 2007b, Smith *et al.*, 2005).

Finally, despite the existence of several good top-level ontologies, our choice of BFO to ground the development of our ontology is based on its large acceptance and use in medical and biological domains (Arp; Smith; Spear, 2015, Bittner; Smith, 2004, Smith; Kumar; Bittner, 2005). We suggest to read the "*Basic Formal Ontology 2.0: Specification and User's Guide*" in order to understand each type of BFO, and in this thesis is the reference Smith *et al.* (2015).

2.4 References on ontology building

In ontology building, there are guidelines orienting the construction of good ontologies. These guidelines suggest that good ontologies need to be formal, use explicit specifications, and suitability to the domain in representation (Schulz *et al.*, 2012). In following, we present a brief explanation of each one of these three principles:

- *Formal*: Defining the meaning of terms without ambiguity and the specification of the relations between terms using logical axioms, allowing automatization for both recovering and reasoning.

- *Use explicit specifications*: Enabling humans to reason and understand the domain represented by the ontology.
- *Need to be adequate to the target domain on representation*: The ontology built must represent the reality and the knowledge present in the domain desired.

Evaluating the methodologies for developing ontologies, it was clear in a systematic review of literature (see appendix 2 for details), the existence of several methodologies, and that there is no consensus on which methodology is most appropriate. The most known methodologies and methods from the mid-90s until to the mid-2016 are:

- Enterprise Ontology (Uschold; Gruninger, 1996, Uschold; King, 1995).
- Toronto Virtual Enterprise (Grüninger; Fox, 1995).
- Methontology (Fernández-López; Gómez-Pérez; Juristo, 1997).
- 101 Method (Noy; Mcguinness, 2001).
- CYC Method (Reed; Lenat, 2002).
- DILIGENTE (Pinto; Staab; Tempich, 2004).
- On-To-Knowledge Methodology (Sure; Staab; Studer, 2004).
- NeOn Methodology (Suárez-Figueroa, 2010b).
- Systematic Approach for Building Ontologies (SABiO) (Falbo, 2014).
- Up for ONtology (UPON) (De Nicola; Missikoff; Navigli, 2005, 2009).
- The methodology of ontological realism (Arp; Smith; Spear, 2015, Smith; Ceusters, 2010).

The focus of this thesis is not to discuss the methodologies or present the detail of each one. Insofar as some researchers already did it. They identify, evaluate, and compare different methodologies such as the following works: Gómez-Pérez; Fernández-López and Corcho (2004), Hafner and Carole (1997), Jones; Bench-Capon and Visser (1998), Suárez-Figueroa (2010b); and Silva; Souza and Almeida (2011), which are available in English, and Mendonça (2015), Silva (2014), which are available in Portuguese.

In this research, we choose the framework of ontological realism (Arp; Smith; Spear, 2015, Smith; Ceusters, 2010) as our main orientation mainly in semantic aspects. In addition, we use the Neon Methodology (Suárez-Figueroa, 2010a) to employ some best practice of ontology engineering process.

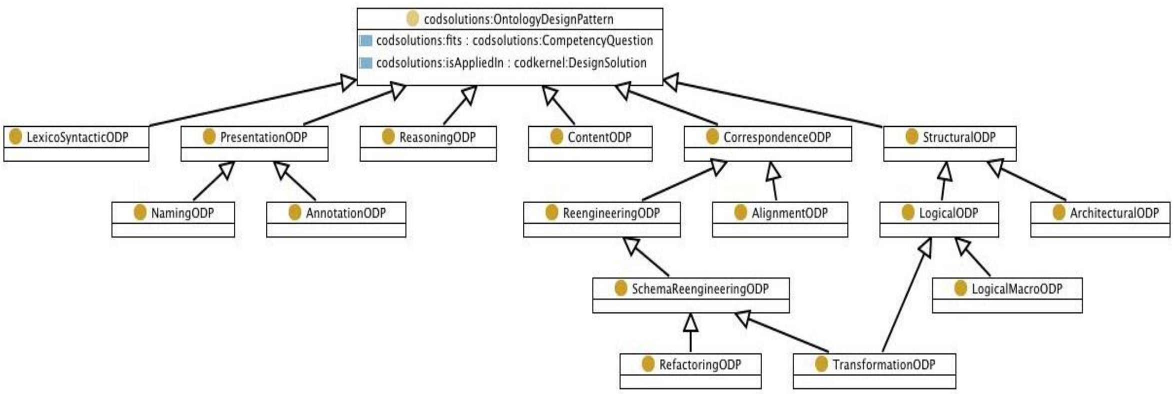
In this section, we present some important references on ontology building. First, we present some general guidelines on ontology building, namely, ontology design patterns and OBO Foundry initiative. Second, we present some references on ontologies evaluation and pitfalls in ontology building. Then, we present an overview of the two methodologies that we consider suitable to our investigation, namely, the methodology of ontological realism and NeOn methodology. Finally, we present a new methodology to develop our ontology in which was conceived from a harmonization of the methodology of ontological realism and NeOn methodology.

2.4.1 Ontology Design Patterns

Design pattern idea has emerged in Software Engineering as a standardized way to solve specific problems of object-oriented modeling. Design pattern reflects the experience, knowledge, and ideas of software developers based on their solutions used in various situations (Gangemi *et al.*, 2007). In the case of ontologies, the term Ontology Design Patterns (ODPs) introduced by Gangemi (2005) and Blomqvist and Sandkuhl (2005), is a set of references practices on Ontology Engineering aiming to turn ontology development easy and with better quality.

Currently, there is a web portal[14] maintained by the Association for Ontology Design & Patterns (ODPA) in which focused on to join all ODPs in a single place to help ontologists to find it (Gangemi; Presutti, 2009, Odpa, 2017b). There are six families of ODPs types according to their intended use, showed in Figure 18 and described in Table 5. The complete ODP catalogue is published in (Odpa, 2017a).

Figure 18: Ontology Design Pattern types



Source: Extracted from (Odpa, 2017b).

[14] Ontology Design Patterns portal is available at OntologyDesignPatterns.org.

Table 5 - Ontology Design Pattern types

Family of ODPs	Subcategory	Description
Structural ODPs	Logical ODPs	Are compositions of logical constructs that solve a problem of expressivity
	Architectural ODPs	Are constructs that characterize the Structure of the ontology as a whole.
Correspondence ODPs	Re-engineering ODPs	Set of rules aiming to transform a non-ontological resource into an ontology resource.
	Alignment ODPs	Templates that set the correspondences between two existing ontologies.
Content ODPs		Concentrate on ontology content and not in logical issues. Focus the design problems solutions addressing the content of competency questions (requirements).
Reasoning ODPs		Use of Logical ODPs to get reasoning results.
Presentation ODPs	Naming ODPs	Names conventions to support usability and readability of ontologies.
	Annotation ODPs	Set of annotation properties or annotation property schemas to increase the understandability of ontologies and their elements.
Lexico-Syntactic ODPs		Organized linguistic structures able to provide conclusions of significance in the context used.

Source: By this thesis researcher based on Gangemi and Presutti (2009), Odpa (2017b).

Finally, a contribution to ODPs from the research in the domain of biological knowledge is the Ontology Design Patterns Public Catalog (ODP-PC)[15] detailed in Aranguren; Antezana*; et al.* (2008), Aranguren; Rector*; et al.* (2008). In order to understand ODP in details, we suggest to read the follow references: Blomqvist and Sandkuhl (2005), Gangemi (2005), Gangemi and Presutti (2009) and Odp-Pc (2017).

2.4.2 OBO Foundry: The Open Biological and Biomedical Ontologies

Considering the idea of creating non-redundant ontologies by minimizing the ambiguities of terms and definitions among ontologies, in addition to promoting the integration of biomedical ontologies, the initiative known as OBO Foundry[16] emerges. OBO Foundry was created by a community of ontologists committed to creating a kind of guide for the development of interoperable biomedical ontologies. Originally, the acronym OBO referred to "*The Open Biomedical Ontologies*" later became "*The Open Biological and Biomedical Ontologies*" (Smith *et al.*, 2007a).

Some premises of OBO Foundry ontologies are the use of BFO as top-level ontology and the adoption of a common set of relations currently existing in the Relations Ontology (Smith *et al.*, 2007a). According to OBO Foundry community:

[15] Ontology Design Patterns Public Catalog is available at http://www.gong.manchester.ac.uk/odp/html/

[16] OBO Foundry portal is available in http://www.obofoundry.org/.

"The mission of the OBO Foundry is to develop a family of interoperable ontologies that are both logically well-formed and scientifically accurate (Obo_Foundry, 2017a)".

OBO Foundry provides a set of principles considered good practices in ontology building and to be part of OBO Foundry library an ontology should follow these principles. Currently, there are thirteen principles, which are described below (Obo_Foundry, 2017a):

Principle 1[17] - Open: The ontology must be open to the ontology community in order to be freely reused by any ontologists. This principle implies that the ontology could not have any constraint against its distribution. However, this does not imply that the ontology credits against one class or relation should be omitted. On the contrary, you should mention the source ontology (Obo_Foundry, 2017a).

"The ontology MUST be openly available to be used by all without any constraint other than (a) its origin must be acknowledged and (b) it is not to be altered and subsequently redistributed in altered form under the original name or with the same identifiers (Obo_Foundry, 2017a)."

Principle 2 - Common Format: The ontology should be implemented and published according to a common formal language in order to maximize reuse and people access. Some formal languages accepted by OBO Foundry are OWL or OWL2 (RDF/XML, OWL2-XML or OWL2-Manchester Syntaxes) and OBO Format (Obo_Foundry, 2017a). To details of OWL and OWL2 see (W3c, 2004b, 2012b, 2012c)

Principle 3 - URI/Identifier Space: The ontology elements (class, relation, annotation property, and object property) must have a single unique identifier space inside OBO Foundry repository. This single identifier is usually implemented by an IRI (Internationalized Resource Identifier). The IRI is composed of an ontology prefix follows by a local identifier. This local identifier usually is a sequence of numbers or mnemonics, which are not meaningful to humans (Obo_Foundry, 2017a).

Principle 4 – Versioning: It is imperative that the ontology maintains a mechanism of control version. Each ontology version should be identified by a label such as version date (Obo_Foundry, 2017a).

[17] OBO Foundry did not establish numbers to mention each one of these principles. This thesis establishes such numbers, randomly, in order to make some references in advanced sections.

Principle 5 – Scope: This principle is not closed and seeks to avoid overlaps of terms and scope between ontologies. Its idea focuses on a clear definition of the ontology's scope. In addition, the content covered by the ontology should adhere its scope properly. In addition, instead of creating a class that already existing in an OBO Foundry ontology, It should be preferred the reuse (Obo_Foundry, 2017a). Note that currently exists tools capable to search for a term in the OBO Foundry repository of ontologies, for instance, the Ontobee server (Ong *et al.*, 2017, Ontobee, 2017, Xiang *et al.*, 2011).

Principle 6 - Textual Definitions: The ontology terms should be defined or described with textual definitions. In addition, formal definitions should be included. According to OBO Foundry web page, it is acceptable a lack of definitions, however, it is not clear a percentage of terms without definition is acceptable (Obo_Foundry, 2017a).

Principle 7 – Relations: This principle claims the adoption of unambiguous relations such as the Relation Ontology (RO). However, there is an observation at this principle that it needs some adjusts. OBO Foundry recommends the establishment of a plan for the future definition of undefined terms (Obo_Foundry, 2017a).

Principle 8 – Documentation: This principle claims a good documentation to the ontology. OBO Foundry suggested some documentations such as developer documentation that could include: whose are the contributed authors, licensing terms, rights of use, version control, release process, changes between versions, development methodology, what are the ontologies in reuse, and deprecation policy (Obo_Foundry, 2017a).

Principle 9 - Documented Plurality of Users: The ontology should mention its use by a plurality of independent users (people or organizations) and to show evidence of non-ontologists whose use the ontology (Obo_Foundry, 2017a).

Principle 10 - Commitment To Collaboration: The ontology development should allow a collaborative development to prevent duplicate work, enlarge interoperability between ontologies, and to guarantee ontology's cover meets the users' needs (Obo_Foundry, 2017a). It could be used version control system (VCS) tools such as Git repository that allows the distributed version control, source code management (SCM), opening of a pull request, and tracking requests (Chacon; Straub, 2014).

Principle 11 - Locus of Authority: This principle establishes that just one person should perform the communications between the community and the ontologists, and mediate discussions related to OBO Foundry issues and maintenance. A centralized communication aims to guaranty answers to all questions (Obo_Foundry, 2017a).

Principle 12 - Naming Conventions: Implies in the adoption of a set of common naming conventions in order to increase the understanding and meanings represented by the ontology (Obo_Foundry, 2017a). Details of the conventions adopted by OBO are available in Schober *et al.* (2009).

Principle 13 – Maintenance: Refers to the evolution of knowledge covered by the ontology. The ontology should evolve to reflect the scientific domain of knowledge in which is part of its scope. The ontology maintainer should be committed to monitoring changes in knowledge and still respond in a timely manner to the necessary changes required by the community (less than 3 months) (Obo_Foundry, 2017a).

Currently, OBO Foundry has a list of a hundred fifty-eight ontologies, wherein eight are foundational ontologies (status Foundry), a hundred nine ontologies have the status library, and forty-one ontologies are without specification of status (Ontobee, 2017). In order to register an ontology on the OBO Foundry, one should follow the guide available at OBO Foundry webpage (Obo_Foundry, 2017b).

2.4.3 Evaluation of ontologies

Ontology evaluation consists of technically investigating the developed ontology. The ontology evaluation process aims to determine if the ontology was suitably built, as well what ontology elements were determined correctly, incorrectly, and what domain elements such ontology is not covering. This process should analyze criteria such as consistency, completeness, conciseness, expandability, and sensitiveness (Gangemi *et al.*, 2005, Gómez-Pérez, 1999, 2004, Gómez-Pérez, 2001).

Some of the key questions for the evaluation of ontologies are (Almeida, 2006):

> *"What are the mechanisms of interaction with ontologies?*
> *What is the formalism of knowledge representation employed?*
> *Is the ontology well documented?*
> *(Almeida, 2006, p. 129)"*

In addition, the evaluation of ontologies demands the emphasis in concepts that compose the ontology itself (Gómez-Pérez, 1999):

- *To check the structure or architecture of the ontology*: are the terms definitions built according to the design criteria?
- *To check the syntax for the definitions*: are there keywords what would be wrong with that? or syntactically incorrect structures in the definitions?
- *To check the content of the definitions*: what does the ontology define? What does it define incorrectly? What can be inferred and what cannot?

According to Suárez-Figueroa (2010b), the process of ontology evaluation checks the technical quality of the developed ontology compare to ontology requirements. This process is composed of two activities, namely, ontology verification and ontology validation. These activities seek to identify, respectively, whether the developed ontology is the correct ontology and whether it is produced in the right way. In the ontology validation activity, the meaning of the definitions is compared against the domain of reality that the ontology has been developed to represent. In the ontology verification activity, the created ontology is compared against the requirements specification document elaborated in the requirements specification phase in order to verify whether the ontology is built according to the specification (Suárez-Figueroa, 2010b).

Gómez-Pérez (2004) discusses on the following ontology quality criteria against the ontology evaluation process: consistency, completeness, conciseness, expandability, sensitiveness. In ontology verification activity, one investigates the consistency of the ontology, its documentation throughout the life cycle and the software environment associated with it. In ontology validation activity, one investigates whether the software and the documentation correspond to the planned information system (Almeida, 2006).

A framework for ontology evaluation presented by Vrandečić (2009) suggested a set of ontology quality criteria and ontology aspects that can be followed in order to evaluate an ontology. Such ontology quality criteria are: accuracy, adaptability, clarity, completeness, computational efficiency, conciseness, consistency and organizational fitness. And the set of aspects are vocabulary, syntax, structure, semantics, representation, and context.

In addition, Gangemi divides evaluation methods into three main categories: structural, functional and usability-based (Gangemi *et al.*, 2006):

- Structural evaluation that focus on the evaluation of both formal syntax and semantics of the ontology.
- Functional evaluations in which evaluate ontologies in their context of use, that is, try do find out whether the ontology meets the intended conceptualization.

- Usability-based evaluations that focus on how the ontology is communicated or how it conveys the intended conceptualization.

The ontology evaluation process is a complex task. First, because of its applicability in different ontology engineering scenarios, and secondly because of the large number of existing evaluation approaches and metrics. Thus, a good way to validate ontologies is to analyze whether the ontology complies with best ontology modeling practices (Poveda-Villalón; Suárez-Figueroa; Gómez-Pérez, 2012b).

2.4.4 Pitfalls in Building Ontologies

Ontology development is a complex activity involving methodologies, tools, domain knowledge, and a set of consolidated practices to produce good ontologies. Several existing methodologies try to help ontologists during the ontology development by fitting their candidate ontologies to a set of good practices and consequently improving the ontology quality. However, sometimes ontologists fail to follow these good practices or do not have technical knowledge in modeling, then generating anomalies in the ontology built (Poveda-Villalón; Suárez-Figueroa; Gómez-Pérez, 2010a, 2010b, Schulz *et al.*, 2012, Vigo *et al.*, 2014).

Considering this context, Poveda-Villalón; Suárez-Figueroa and Gómez-Pérez (2010a, 2010b) identified a set of common worst practices in ontology development and proposed a catalogue to help ontologists to avoid these so-called pitfalls. According to Poveda-Villalón; Gómez-Pérez and Suárez-Figueroa (2014) and Poveda-Villalón (2016) the catalogue consists of 41 pitfalls. Each one with its respective definition presented in Appendix 1.

In addition, Poveda-Villalón; Suárez-Figueroa and Gómez-Pérez (2010b) defined two different classifications to pitfalls listed in their catalogue, improved later by Poveda-Villalón (2016). The first classification considers ontology quality dimensions defined by Gangemi *et al.* (2006), namely: structural, functional and usability-profiling. Later, Poveda-Villalón (2016) provided an adaptation to the dimensions of Gangemi *et al.* (2006), explained in follows.

- <u>*Structural dimension:*</u> refers to syntax and formal semantics of an ontology Gangemi *et al.* (2006). Under this dimension, Poveda-Villalón; Suárez-Figueroa and Gómez-Pérez (2010b) and Poveda-Villalón (2016) analyze the pitfalls in four aspects:
    - *Modeling decisions:* analyze the correct use of ontology implementation languages (searching mistakes) and search possible improvements in ontology modeling aspects.

- *No inference:* verify the ontology capability to provide inferences on knowledge represented.
- *Wrong inference:* verify if the inference provided by ontology can result in wrong or invalid knowledge.
- *Ontology language:* check ontology compliance with the ontology language specification and the syntax used in ontology formalization.

- *Functional dimension:* Refers to the ontology function in a given application context (Gangemi *et al.*, 2006). In addition, Poveda-Villalón; Suárez-Figueroa and Gómez-Pérez (2010b) and Poveda-Villalón (2016) analyze three functional aspects:
    - *Real world modeling or common sense:* evaluate adherence of ontology coverage to knowledge domain represented.
    - *Requirement completeness:* check whether the ontology realizes the ontology's requirements specified in initial phases of development.
    - *Application context:* analyze ontology adherence for a specific use case in its application context.

- *Usability-profiling dimension:* Involve the communication context of an ontology (Gangemi *et al.*, 2006), which means, the set of additional information provide by an ontology in order to help it to be understood. Thus, the three aspects analyzed in this dimension by Poveda-Villalón; Suárez-Figueroa and Gómez-Pérez (2010b) and Poveda-Villalón (2016) are:
    - *Ontology understanding:* assess every information available on the ontology that provides a better understanding of the ontology.
    - *Ontology clarity:* verify the comprehensibility (or understandability) capacity given by ontology elements properties to ontology users.
    - *Ontology metadata:* verify the existence of ontology metadata (information about the ontology).

In Figure 19 we present the classification of the pitfalls importance levels by ontology quality dimensions given by Poveda-Villalón (2016). The importance levels are highlighted by colors, in which the black highlights pitfalls are the *critical* pitfalls, the blue highlights pitfalls are *important* pitfalls and the red highlights pitfalls are *minor* pitfalls.

According to the classification presented in Figure 19, we notice that the pitfalls affect the ontology in different importance levels, and each level demand a different action (Poveda-Villalón, 2016, p. 116):

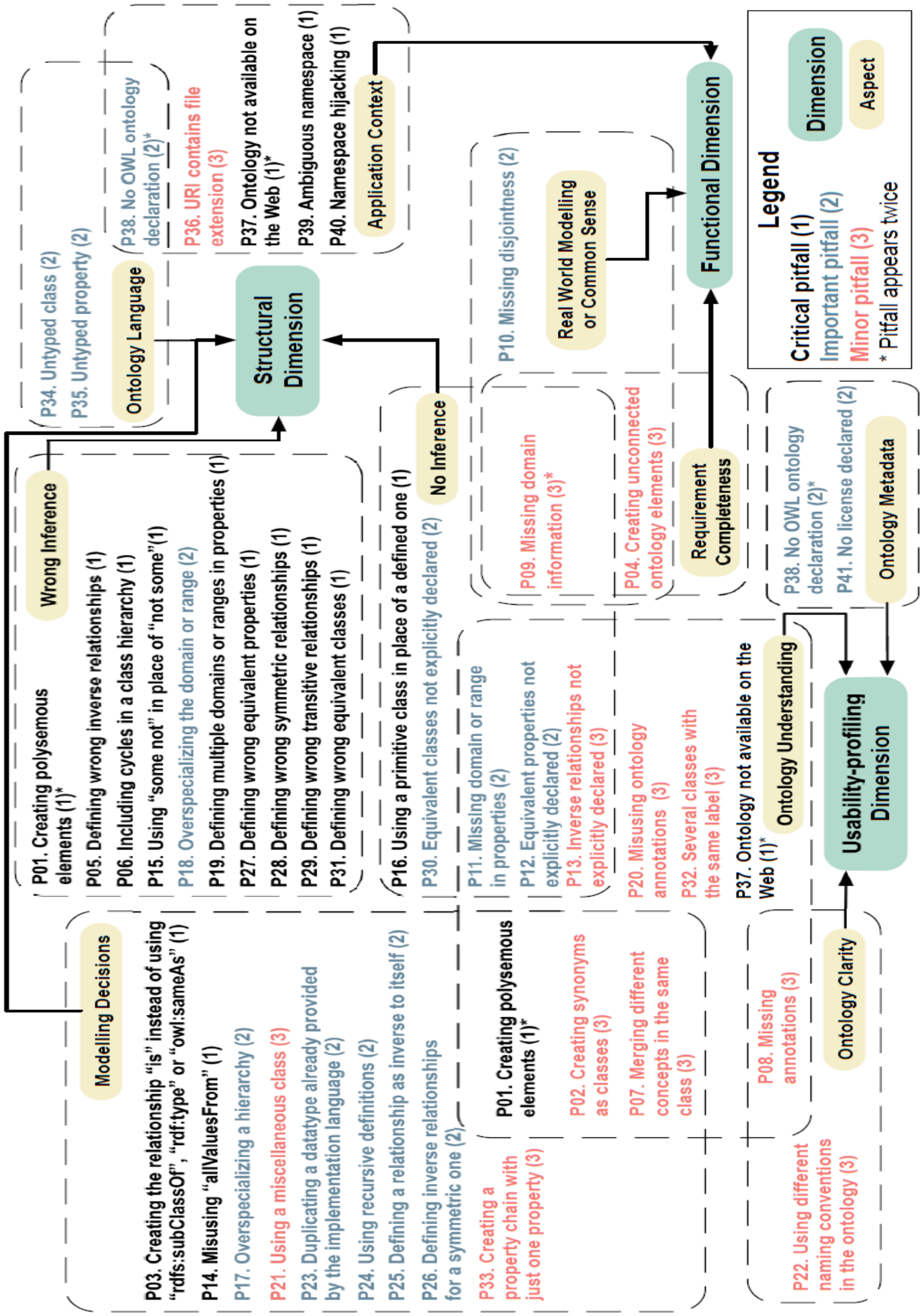
- *Critical* pitfalls that affect the consistency, reasoning, and applicability of the ontology, consequently it mandatory to correct.
- *Important* pitfalls that affect some ontology characteristics but not ontology function as a whole, thus, it would be important to correct this kind of pitfall but not mandatory.
- *Minor* pitfalls that do not affect the ontology characteristics and its functioning as a whole, but turns the ontology less understandable. Thus, it is not required to correct, but it would make ontology more rich and understandable.

The second classification proposed by Poveda-Villalón; Suárez-Figueroa and Gómez-Pérez (2010b) follows three criteria of ontology evaluation proposed by Gómez-Pérez (2004), namely: consistency, completeness, and conciseness. Figure 20 shows the classification of 20 pitfalls according to these criteria. At that time of classification, there were only 24 pitfalls.

- *Consistency:* Checks whether the ontology affords contradictory conclusions from valid data entry.
- *Completeness:* Evaluate the ontology coverage among the scope of knowledge to be covered.
- *Conciseness:* It indicates how concise the ontology is, that is, whether the representations handled by the ontology are both useful and precise in the scope of its application.

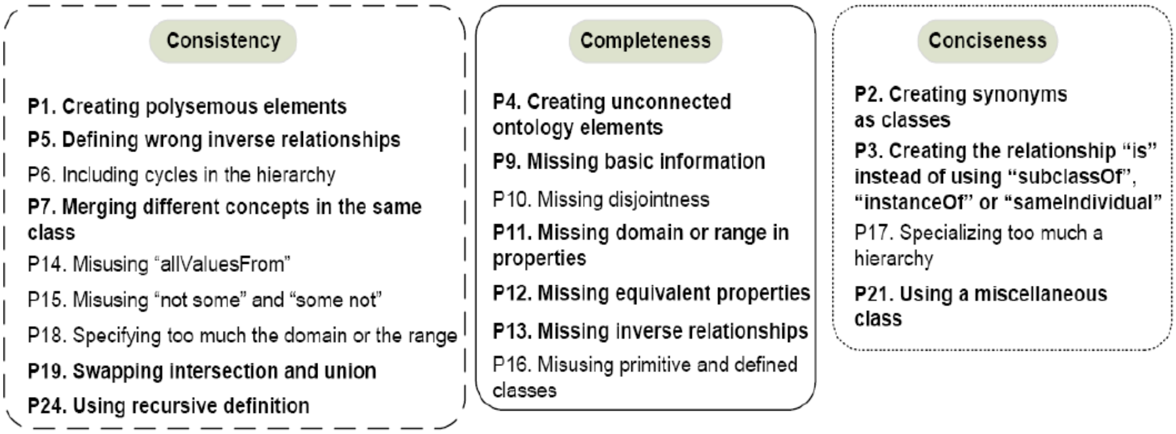
According to Poveda-Villalón (2016), each pitfall correlates to at least one aspect of the first classification (Figure 19), and could even correlate to more than one. Nonetheless, in the case of the second classification, there are pitfalls that cannot match these three criteria because they do not approach the ontology content (Poveda-Villalón; Suárez-Figueroa; Gómez-Pérez, 2010b).

Figure 19: Classification of ontology pitfalls by ontology quality dimensions and its aspects.



Source: Extracted from Poveda-Villalón (2016)

Figure 20: Classification of the pitfalls according to the evaluation criteria.



Source: Extracted from Poveda-Villalón; Suárez-Figueroa and Gómez-Pérez (2010b).

Finally, a great contribution of Poveda-Villalón research is an online tool, called *Ontology Pitfall Scanner*[18], for ontology evaluation that checks an ontology against the list of pitfalls. In order to get details on *Ontology Pitfall Scanner* tool, we recommend to read the following references: Poveda-Villalón; Gómez-Pérez and Suárez-Figueroa (2014), Poveda-Villalón and Suárez-Figueroa (2012), Poveda-Villalón *et al.* (2009), Poveda-Villalón; Suárez-Figueroa and Gómez-Pérez (2012a, 2012b), Suárez-Figueroa; Kamel and Poveda-Villalón (2013).

2.4.5 The methodology of ontological realism

The methodology of ontological realism supports the development of ontologies providing solid philosophical theoretical background. The driving principle of ontological realism recognizes ontologies as artifacts that represent the reality of a given scientific domain (Arp; Smith; Spear, 2015, Smith; Ceusters, 2010). According to Smith and Ceusters (2010), this methodology has been tested since 2002 to help the ontology building in several knowledge domains. This claim drives some thoughts of the ontological realism methodology as a consolidated and mature methodology.

The methodology of ontological realism can be summarized as the view that a suitable ontology for a domain should be constructed not to represent existing data or models of the domain, but rather to represent (the terminological part of) the relevant established science. Ontological realism defends that a good ontology is result of a building process that follows the eight principles listed below (Arp; Smith; Spear, 2015, Smith; Ceusters, 2010).

---

[18] "*Ontology Pitfall Scanner!*" or OOPS!, and it is available in http://oops.linkeddata.es/.

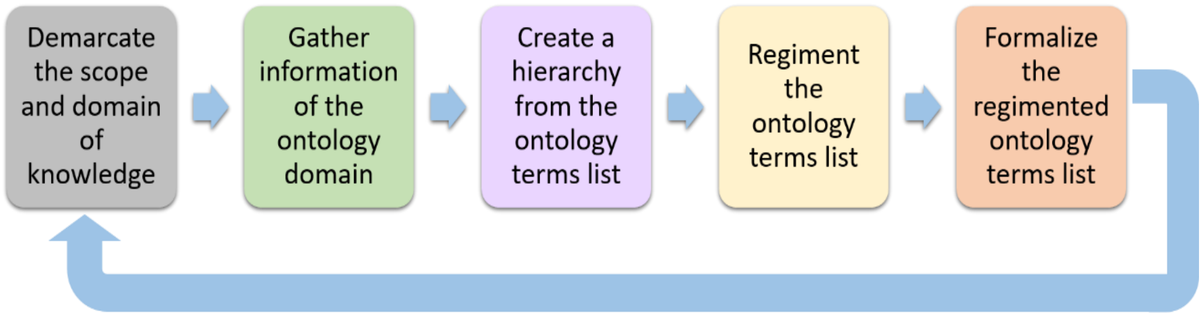
- *Realism*: Implies that an ontology represents entities of reality and not concepts or mental representations from people.
- *Perspectivalism:* It considers the complexity of reality as a set of different perspectives, in which an explanation of reality could follow different theories or reality views, and any of these representations can be correct. In other words, an ontology should pursue a modularization, in which each module follows a single perspective of reality, and involves people skilled conform the perspective of reality that it represents.
- *Fallibilism:* It means that theories regarding a reality could fail and are subject to correction. In this sense, as part of ontology design, a maintenance of an ontology control version helps by providing a mechanism in which users are able to indicate errors and gaps in the ontology.
- *Adequatism:* As opposed to reductionism, it assumes that any perspective of reality has the same importance, and one should not minimize a perspective of reality in trying to explain all other perspectives. A good ontology should be built together with others ontologies in an interoperable way, forming a structural basis of entities and relations between the entities at different levels of granularity.
- *Reuse:* An ontology should privilege reuse of other existing ontology resources concerning the domain of ontology in development. Ontology resources previously existing are candidates to reuse or regarded as a reference to the ontology in development.
- *Balance between utility and realism representations:* Representations of reality, according the theory of realism, sustains a common reference about the domain of ontology in development, even though they mean different concepts of applications in use. Sometimes, when the concepts of applications prevail over the concept of reality, to maintain ontologies of close domains interoperable requires considerably higher effort.
- *Ontology design is an Open-Ended process:* Construction of a domain ontology is an open process in which an ontology should be expandable and changeable since knowledge is in constant evolution. Given that an ontology expresses knowledge under a scientific domain, if the knowledge changes, increases or improves the ontology should reflect this. This implies that the process should

provide maintainability, updating, correction, evaluation of the ontology in focus in addition to providing the ability of integrating related ontologies.

- _Follow the idea of Low-Hanging Fruit_: The construction of ontology follows a top-down representation, starting with the identification and definition of the most common entities and their relations, and move forward to more complex entities within the domain.

According to the guidelines of the in methodology of ontological realism, the design of a domain ontology consists of five steps (Arp; Smith; Spear, 2015, Smith; Ceusters, 2010). Figure 21 demonstrated the workflow of the methodology of ontological realism with its steps performed in an iterative and incremental cycle (Arp; Smith; Spear, 2015).

Figure 21: Outline of the steps to build a domain ontology according ontological realism.



Source: By thesis' researcher based on Arp; Smith and Spear (2015).

Another important aspect is the top-down characteristic of the life cycle process. It starts defining the most general terms and in each iteration down to the term least general. Then, at the end of each iteration, the ontology receive an increment of terms and details (Arp; Smith; Spear, 2015). In following we present details of the steps of Figure 21 according to the main references of ontological realism methodology, namely, Smith and Ceusters (2010) and Arp; Smith and Spear (2015).

The development process starts by demarcating the scope of an ontology and the nature of knowledge that one intends to model. At the end of this step the three following questions need to be provisional answered (Arp; Smith; Spear, 2015):

> *"What part of reality is this ontology an ontology of?*
> *What are the domain universals and relations that need to be represented?*
> *What are the appropriate domain-specific terms that should be used in representing these universals and relations?*
> *What levels of granularity of entities are salient for the ontology? (Arp; Smith; Spear, 2015, p. 55)"*

Besides that, it is important to identify existing ontological resources and artifacts whose scope meets the requirements of the ontology under construction. Such artifacts and resources can compose a list of candidates for reuse (Arp; Smith; Spear, 2015, p. 51).

In short, the expected artifacts generated as result of this step are:

- The knowledge domain in which the ontology will cover.
- Ontology scope and iteration scope, according to step comprehensiveness.
- Levels of granularity that the ontology entities can represent.
- A list of ontological resources candidates for reuse.
- A list of non-ontological resources candidates for reuse, such as textbooks, reports and documents, terminologies, etc.

The second step consists in gathering information of the ontology domain, concentrates on knowledge acquisition. The guidance here is to get the most common terms used on textbooks and documents relating to the domain in focus. In addition, one should seek relevant terms on the existing ontologies related to the concerned domain. At the end of this step, it is expected a list with general terms about the knowledge of the domain ontology. Besides that, this step also aims to demarcate a list of entities and relations between these entities to be represented (Arp; Smith; Spear, 2015, p. 51).

During the execution of this step, ontologists could (Arp; Smith; Spear, 2015):

- Identify general terms available in textbooks, documents, reports, and, terms used by domain experts.
- Analyze terms list to remove redundancies and identify synonyms.
- Evaluate the content of the ontologies candidate for reuse and define the potential reusable ontological resources such as entities, relations, and properties.

Moreover, based on the experience of the thesis researcher, ontologists could:

- Identify possible relations between the general terms identified.
- Maintain a documentation of decisions made such as:
  - Prioritizing of main terms and their synonyms.
  - Choosing for an ontology instead of another when more than one ontology defines the same term necessary.

- Choosing for using of a specific standard or ontology design pattern.

The third step is the creation of a hierarchy with the ontology terms list. It consists in organizing the list of terms in an *is_a* hierarchy in identifying the most general term down to the least general. Besides that, for each term, it is important to determine a preliminary definition of the term (Arp; Smith; Spear, 2015). During this step performance, the ontologist should:

- Identify the root node candidates and non-root terms.
- Identify common characteristics between these terms in order to put the terms with same features together.
- Identify any relevant terms in neighboring ontologies to reuse and to guide the organization.

Arp; Smith and Spear (2015) claim that taxonomies contribute for organizing the hierarchy. The principles of taxonomies are:

- Organize every ontology around a hierarchy axis, using the *is_a* relation to connect classes.
- Assure completeness of the *is_a* hierarchy certifying that each term is connected to the root term of the hierarchy by way of consecutive *is_a* relations between terms and their parents.
- Ensure mono-hierarchy in which each term in the *is_a* hierarchy has just one parent, avoiding multiple inheritance.
- Respect the open-world assumption regarding ontologies, that is, ontologies should be resilient, allowing its extension and correction.
- Represent the reality itself, and not what is known about the reality.

Then, the fourth step, regimenting the ontology terms list, aims to organize the hierarchy of terms according a set of rules in order to guarantee coherence, defining these terms through human understandability definitions (natural language definitions), and determining significant terms from related ontologies. The regimentation is iterative, with constantly performed improvements over versions of the terms hierarchy and definitions (Arp; Smith; Spear, 2015).

According to Arp; Smith and Spear (2015) this regimentation should address terminological and definitional aspects explained bellow, besides the aspect of location in the *is_a* hierarchy already cited and explained in the previous step. The principles of terminology refer to a set of best practices and rules to be observed in the construction of the terminology used by the ontology. Notice that the word terminology here refers to human language or parlance, and not to the terminology subject of Information Science field. On terminological issues, given the *is_a* hierarchy of general terms list, Arp; Smith and Spear (2015) suggested:

- Ensure the exclusion of redundant terms.
- Ensure the identification of synonyms, mainly for terms commonly used by humans. You may identify also possible antonymous of these terms.
- Keep track of synonyms for terms already in the terminology list.
- Check for existing ontology content suitable to your ontology project and evaluate this content as reusable in your ontology.
- Create a list of terms under the knowledge domain scope, checking textbooks and relevant documents, but prioritizing terms used by influential groups of scientists.
- Ensure maximal conformity with the terminological usage of influential groups of domain scientists.
- Identify areas of disciplinary overlap in which terminological usage is not consistent.
- Above all, try to reuse as many existing resources as possible.

Still, on terminological questions, Arp; Smith and Spear (2015) suggest the use of the following formatting rules for terms:

- Prioritize single nouns and noun phrases to create the list of words.
- Use singular words, avoiding plural and collective words.
- Avoid acronyms and abbreviations and mass terms, usually, acronyms and abbreviations are alternatives terms and should be treated as such. However, exception applied in the case of an acronym or abbreviation that becomes more popular in the language than the pure term or expanded form, for instance, "DNA" or "AIDS".
- Ensure distinction between universals and instances.
    - Use lowercase italic format in words (nouns) referring to universals (types or classes). For example, use "*human being*" or "*homo sapiens*" to indicate universal of human beings.

    - Use initial capital letters and italic format in words (nouns) referring to names of particulars (instances). For example, use "*Fernanda*" to indicate a particular of the universal "*homo sapiens*".

- Ensure unicity of terms in such way that a unique identified term has just one meaning.
- Ensure unicity of relational expressions.
- Avoid mass nouns transforming it into count nouns. For instance, use "*portion of blood*" instead of "*blood*".

Principles of definitions concern how is the term conceptualized and described in the ontology. Since an ontology is a semantic artifact, during regimentation of the ontology terms, ontologists should define each term in consonance with its meanings. Definitions should express the necessary and sufficient conditions to be the entity in which the term intended to represent. Recommendations of Arp; Smith and Spear (2015) to ontologists under definitions issues are:

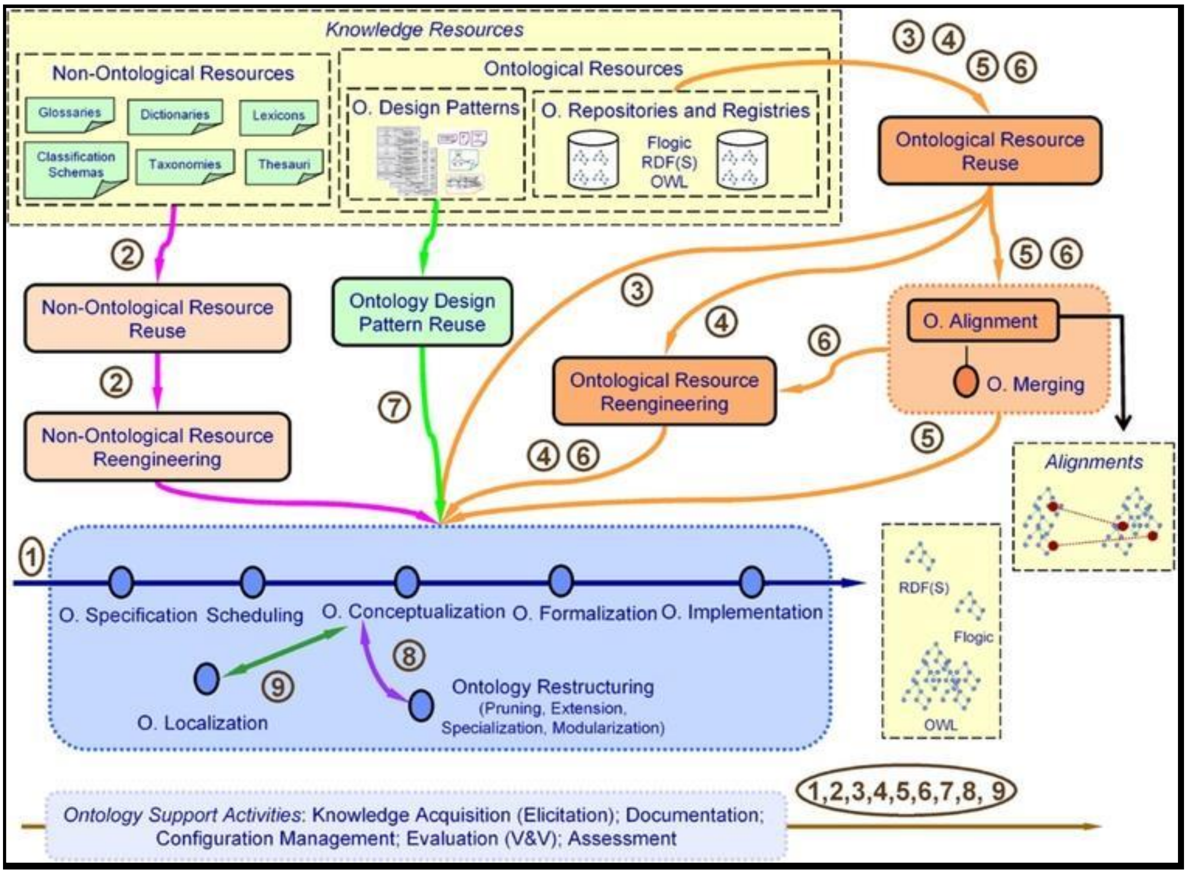
- Determine a definition for all non-root terms.
- Use simpler and intelligible terms to compose the term definition.
- Define terms using the essential features of a thing is, and without these features, the thing would not be the type of thing that it is.
- Define terms in a top-down method, that is, from the most general terms to specific terms.
- Define terms in a human understandability way, preferably using Aristotelian definitions style (read Arp; Smith and Spear (2015, p. 69) to details).
- Avoid circularity in defining terms, which means, a term definition should not mention the term itself.
- Ensure to not create terms for universals by logical combination.
- Avoid the creation of negative terms (negative universals), for instance, there is a universal "*rabbit*" but not a "*non-rabbit*" universal. A "*non-rabbit*" could be any other animal species. However, sometimes a universal admits the prefix “*non-*” not in terms of a negative term, but as a definition of the feature of a universal, for example, "*non-smoker*" as "*a human being who does not smoke*".
- Ensure an unpackable capacity for terms definitions, i.e., the replacement of a term by its definition, and vice versa, does not changing the meaning expressed.

Finally, the last step, formalizing the regimented ontology terms list, refers to encode the ontology or representation of terms and relationship between terms using a formal language, such as *Ontology Web Language* (OWL), readable by computers (Arp; Smith; Spear, 2015). At the end of this step, each iteration delivery a piece of ontology that should be merged to the previously piece. For more details on the methodology of ontological realism and each of your steps, we suggest Arp; Smith and Spear (2015). This reference so far is the most complete, including examples of real cases experienced by the authors.

2.4.6 NeOn Methodology for Building Ontology

The NeOn is a methodology for building ontology that supports the collaborative development as well as the reuse and the dynamic evolution of networked ontologies in distributed environments. NeOn is an acronym to Network Ontology. This methodology provides guidance for all key aspects of the ontology engineering process: collaborative development, reuse of ontological and non-ontological resources, and both the evolution and the maintenance of networked ontologies (Suárez-Figueroa, 2010b, Suárez-Figueroa; Gómez-Pérez; Fernández-López, 2012, 2015).

Figure 22: Ontology development scenarios proposed by NeOn Methodology



Source: Extracted from Suárez-Figueroa (2010b)

The NeOn Methodology framework is a scenario-based methodology in which there is a set of nine scenarios for building ontologies (Figure 22) emphasizing the reuse of ontological and non-ontological resources (Suárez-Figueroa, 2010b, Suárez-Figueroa; Gómez-Pérez; Fernández-López, 2012, 2015).

According to Figure 22, the arrows with associated numbers between 1 and 9 indicate each one of the nine scenarios. The colored rectangles with rounded edges and the colored circles (both blue and orange) are the processes or activities performed in each scenario. The NeOn Glossary defines each one of these processes or activities. The dotted rectangles at the top of the figure represent knowledge resources available to reuse. The dotted rectangles at the right bottom side of the figure represent outputs of the scenarios execution (Suárez-Figueroa; Gómez-Pérez; Fernández-López, 2012).

The idea of NeOn in distinguishing nine scenarios makes the methodology flexible and adaptable to the specific needs of each ontology project (Suárez-Figueroa; Gómez-Pérez; Fernández-López, 2012). These nine scenarios are described below:

- *Scenario 1 - From specification to implementation*: Consists in the development of an ontology from the scratch, that is, without using anything that already exists, assuming that there is anything to be reused. In this scenario, the ontology requirements specification activity is the first activity performed and guides the follows activities. This scenario is the base scenario and is present in every others scenarios (Suárez-Figueroa; Gómez-Pérez; Fernández-López, 2012, 2015).

- *Scenario 2 - Reusing and re-engineering non-ontological resources (NORs)*: A scenario that involves reuse of non-ontological resource related to the knowledge domain of the ontology in building, such as glossaries, taxonomies, thesauri and so on. First, ontologists should select the non-ontological resources available to reuse. Next, ontologists perform a re-engineering these non-ontological resources converting them into an ontological resource. Finally, the resulting ontological resource serves as input to perform the scenario 1 (Suárez-Figueroa; Gómez-Pérez; Fernández-López, 2012, 2015).

- *Scenario 3 - Reusing ontological resources*: This scenario involves reuse of an ontological resource related to the knowledge domain of the ontology under construction. In order to build the new ontology, the reuse of ontological resources could be total or partial. First, the ontologists should select the ontological resources available to reuse and then provide it as input for Scenario 1. The basic difference between scenarios 3 and 1 is the fact of the ontology is not build without using anything that already exists (Suárez-Figueroa; Gómez-Pérez; Fernández-López, 2012, 2015).

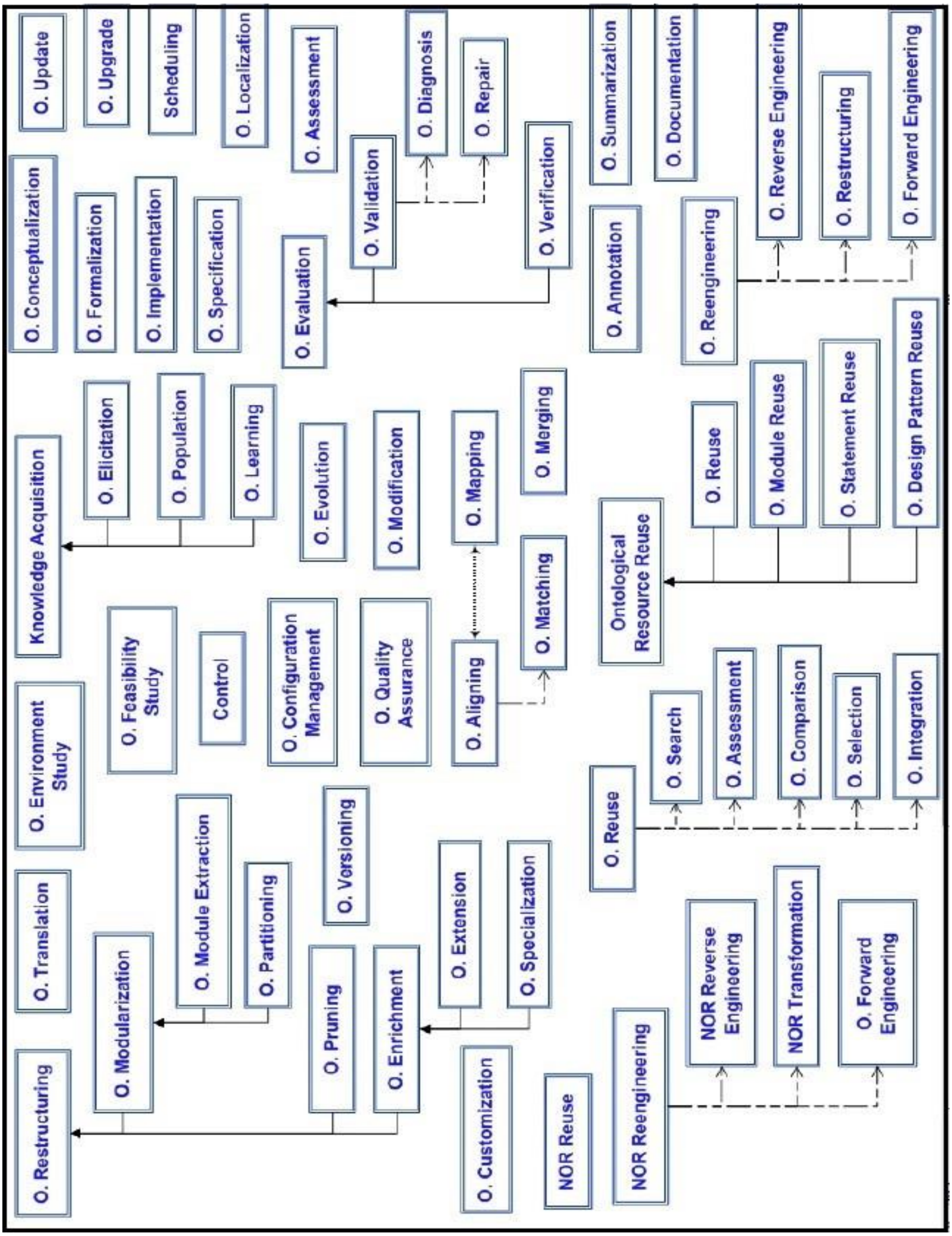
- *Scenario 4 - Reusing and re-engineering ontological resources*: This scenario involves reuse of ontological resource that is not ready to be used in the new ontology since it needs a re-engineer to become suitable. After the selection of ontological resource, the ontologist should perform a re-engineering of such resource preparing it as an input to the Scenario 1 (Suárez-Figueroa; Gómez-Pérez; Fernández-López, 2012, 2015).
- *Scenario 5 - Reusing and merging ontological resources*: This scenario involves both to reuse of an ontological resource and to combine of more than one ontological resource in order to create another one. Following the same idea of scenario 3, ontologists select all ontological resources available to reuse and then succeed the merging of them to combine these resources in just one. Finally, the resulting resource becomes an input for Scenario 1 (Suárez-Figueroa; Gómez-Pérez; Fernández-López, 2012, 2015).
- *Scenario 6 - Reusing, merging, and re-engineering ontological resources*: This scenario makes sense when the ontological resource resulting from the merging process of scenario 5 is not suitable for reusing and needs some adaptation or re-engineering. Thus, after the merging, the ontologist should performs the re-engineered process in order to create a suitable input to scenario 1 (Suárez-Figueroa; Gómez-Pérez; Fernández-López, 2012, 2015).
- *Scenario 7 - Reusing ontology design patterns*: This scenario involves the reuse of ontology design patterns. In this scenario, the ontologist should select the ontology design pattern suitable to be reuse in the target ontology. Usually, the selected patterns compose the requirement document and used during the conceptualization and implementation activities (Suárez-Figueroa; Gómez-Pérez; Fernández-López, 2012, 2015).
- *Scenario 8 - Restructuring ontological resources*: This scenario deals with situations in which the conceptual model of the ontology requires rearrangements to cover the requirements of the ontology. It usually involves ontological resources generated in multiple scenarios and that should integrate and compose the target ontology (Suárez-Figueroa; Gómez-Pérez; Fernández-López, 2012, 2015).

- *Scenario 9 - Localizing ontological resources*: This last scenario fits in situations that the ontology should attend multiple languages. Thus, the ontology in development demands terms and definitions written in multiple languages. In this case, the recommendation is to use at least English and translate to other required languages (Suárez-Figueroa; Gómez-Pérez; Fernández-López, 2012, 2015). For more details of this scenario see Espinoza; Montiel-Ponsoda and Gómez-Pérez (2009) and Montiel-Ponsoda and Espinoza (2010).

In addition to the scenarios, NeOn Methodology provides an extensive, detailed, and rational list of fifty-nine processes and activities (showed in Figure 23). NeOn also provides methodological guidelines for some processes and activities and the "*NeOn Glossary of Processes and Activities*" to standardize the terminology and the understanding of these activities and processes (Suárez-Figueroa, 2010b). It is not the focus of this thesis to describe the entire NeON methodology and both its activities and processes. For details on NeOn activities and processes see Suárez-Figueroa (2010b).

NeON methodology proposed a document template to record the ontology requirements according to Figure 24, and suggested this specification document as result of the ontology requirements specification activity. This template includes specific section to the purpose, the scope, the implementation language of the ontology network, the target group, the intended uses of the ontology network, as well as the set of requirements that the ontology network should fulfil, mainly in the form of competency questions (CQs).

Figure 23: Processes and activities of NeOn Methodology.



Source: Extracted from Suárez-Figueroa (2010b)

Figure 24: Ontology requirements specification document template of NeON methodology.

Ontology Requirements Specification Document Template	
1	**Purpose**
	The main general goal of the ontology. In other words, the main function or role that the ontology should have.
2	**Scope**
	The general coverage and the degree of detail that the ontology should have.
3	**Implementation Language**
	The formal language that the ontology should have.
4	**Intended End-Users**
	The intended end-users expected for the ontology.
5	**Intended Uses**
	The intended uses expected for the ontology.
6	**Ontology Requirements**
	a. Non-Functional Requirements
	The general requirements or aspects that the ontology should fulfil, including optionally priorities for each requirement.
	b. Functional Requirements: Groups of Competency Questions
	The content specific requirements that the ontology should fulfil, in the form of groups of competency questions and their answers, including optionally priorities for each group and for each competency question.
7	**Pre-Glossary of Terms**
	a. Terms from Competency Questions
	The list of terms included in the competency questions and their frequencies.
	b. Terms from Answers
	The list of terms included in the answers and their frequencies.
	c. Objects
	The list of objects included in the competency questions and in their answers.

Source: Retrieval from Suárez-Figueroa (2010b, p. 115)

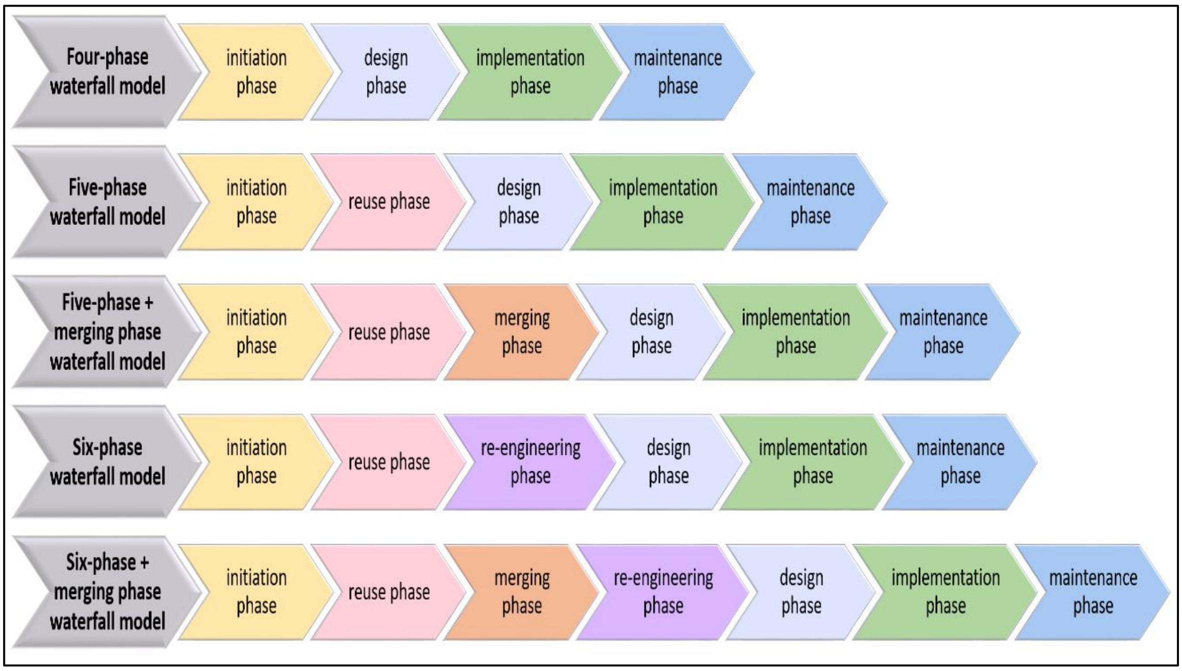
Finally, based on software engineering, NeOn methodology also defines a set of life cycle models applied to the development of ontologies grounded in the Waterfall and Iterative-incremental Life-Cycle Models. In an ontology development project, life cycle model is a way to arrange the processes and activities in phases, as well as the sequence to perform these phases (Suárez-Figueroa, 2010b, Suárez-Figueroa; Gómez-Pérez; Fernández-López, 2012, 2015).

In a "*Waterfall ontology network life-cycle model*", the development stages flow sequentially. A stage starts just when the previously stage finish. In the case of the "Iterative-Incremental ontology network life cycle model", the development of ontology is divided into small projects in which the scope is a part of the general scope. Each iteration or small project should be treated as an individual project. At the end of each iteration, a piece of ontology is created and added to the previously piece, in which was delivered in the previously iteration (Suárez-Figueroa, 2010b, Suárez-Figueroa; Gómez-Pérez; Fernández-López, 2012, 2015).

The waterfall model is still subdivided into five types conforming the needed stages to perform (Suárez-Figueroa, 2010b). Figure 25 illustrated the phases and the execution sequence also explained below.

- *Four-phase waterfall model*: A waterfall model that involves four-phases or four-stages to complete a cycle. These phases are initiation, design, implementation, and maintenance phases, performed in the same sequence as listed and demonstrated at Figure 25.
- *Five-phase waterfall model*: It is a four-phase waterfall model that includes the reuse phase (identify reusable ontological resources) between initiation and design phase.
- *Five-phase + merging phase waterfall model*: It is a five-phase waterfall model that includes the merging phase between reuse and design phase. Merging phase aims to create an ontological resource by joining the ontological resource found in the reuse phase.
- *Six-phase waterfall model*: It is a five-phase waterfall model that includes the re-engineering phase between reuse and design phase. Note that there is not the merging phase here. The re-engineering phase aims the creation of new ontological resources combining ontological and non-ontological resources.
- *Six-phase + merging phase waterfall model:* It is a six-phase waterfall model that includes the merging phase between reuse and re-engineering phase.

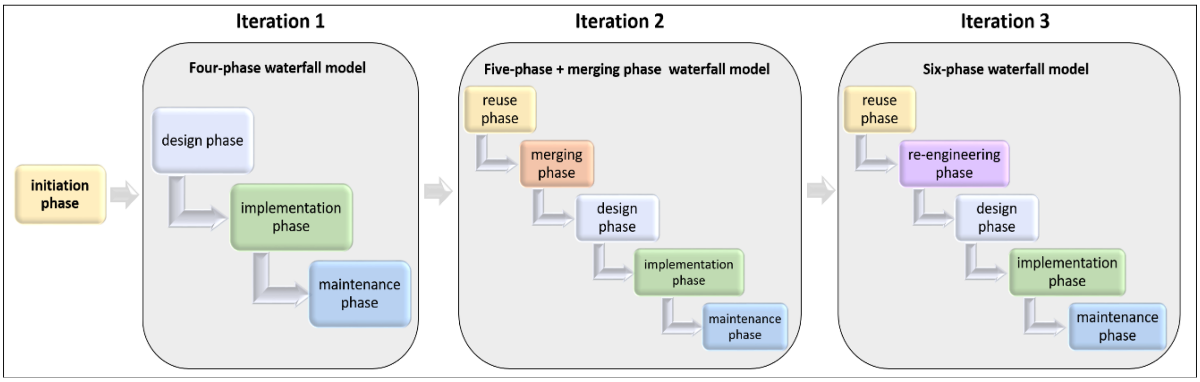
Figure 25: Types of Waterfall life-cycle models according NeOn Methodology.



Source: By thesis' researcher based on figures 32 to 36 of Suárez-Figueroa (2010b).

In the case of the life cycle iterative-incremental, each iteration could assume one different type of waterfall model due to the particularity of the iteration. Thus, in an ontology building project that follows an iterative-incremental life cycle model, it could be performed several different waterfall models until the project conclusion. Figure 26 presents an scheme of an iterative-incremental ontology development.

Figure 26: The Iterative-incremental life-cycle model according to NeOn Methodology.



Source: By thesis' researcher based on figure 37 of Suárez-Figueroa (2010b).

Finally, the NeOn Methodology further maps the scenarios with possible processes and activities that can be performed (Suárez-Figueroa, 2010b, p. 98); related scenarios and lifecycle models (Suárez-Figueroa, 2010b, p. 109-110); and created a correspondence between processes and activities and phases of the life cycle model of the ontology network (Suárez-Figueroa, 2010b, p. 129).

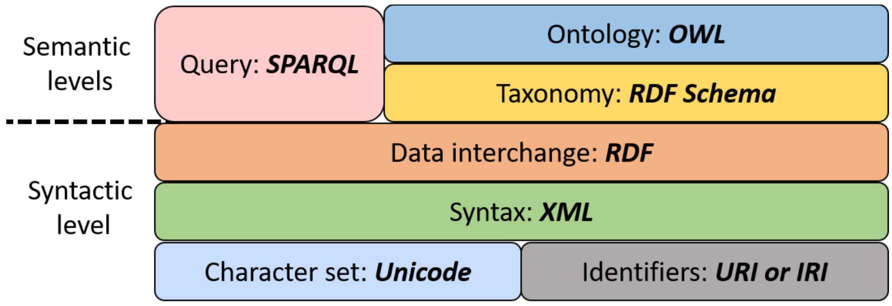
The NeOn Methodology is a result of a widely PhD research work conceived from best practices of three other methodologies: Methontology, On-To-Knowledge and Diligent (Suárez-Figueroa, 2010b). For more details on NeOn Methodology, we recommend the lecture of the follows materials Suárez-Figueroa (2010b), Suárez-Figueroa; Gómez-Pérez and Fernández-López (2012), and Suárez-Figueroa; Gómez-Pérez and Fernández-López (2015).

2.5 Semantic web technologies

"*The Semantic Web is a web of data*" (W3c, 2013c). *Semantic Web* is extension of the traditional Web including to web content structures of meaning in a clear and well-defined way, allowing computers to interact with each other by exchanging information. The main idea of Semantic Web is the enrichment of the traditional Web enabling and improving web interaction in such way that both human beings and computers can understand the content (Berners-Lee; Hendler; Lassila, 2001, Shadbolt; Berners-Lee; Hall, 2006). Semantic Web gives people the ability to create their own data repositories on the Web, develop vocabularies, and formulate rules to interact with data (W3c, 2013c).

The Semantic Web Architecture has two distinct levels, namely, semantic and syntactic levels. Each level has layers that incrementally introduce expressive primitives (Figure 27) (Berners-Lee; Hendler; Lassila, 2001, Harth; Janik; Staab, 2011).

Figure 27: Semantic web architecture



Source: Based on Berners-Lee; Hendler and Lassila (2001) and Harth; Janik and Staab (2011)

The syntactic level incorporates consolidated technologies from hypertext web. In the first layer, Unicode refers to the standard of encoding international character sets used to express and manipulate text in multiple human languages. The Unicode standard gives an exclusive number for each character, independently of platform, program, or language (Unicode_Consortium, 2017). Data identifiers follow the Uniform Resource Identifier (URI[19])

[19] See details at http://www.ietf.org/rfc/rfc2396.txt

pattern or Internationalized Resource Identifier (IRI[20]) pattern. The purpose of a URI is clearly to specify a unique identifier to represent a resource on the web of data, as well as the relationships between them. IRI is a variant of URI that allows usage of a wide range of characters. In general terms, a difference between URI and IRI is that URIs work in ASCII characters, while IRIs work using Unicode typeset.

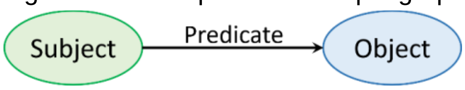
The second and third layers correspond to the structure layer that is responsible for structuring the data and defining its meaning. The structure layer is commonly expressed using the eXtensible Markup Language (XML) and Resource Description Framework.

XML is a markup language used for data exchange. XML together with XML namespaces and XML schema definitions ensure the use of a common syntax in the Semantic Web. Sometimes, XML document should follow the rules defined in an XML Schema Definition (XSD). An XML document is a data model, a format in which any data can be represented, transferred, and stored. While an XSD describes the structure of an XML document (W3c, 2000, 2006, 2016).

The Resource Description Framework (RDF), developed by the World Wide Web Consortium (W3C), is an infrastructure that allows the codification, the interchange, and the reuse of structured metadata. This infrastructure fosters metadata interoperability by designing mechanisms to support common semantic, syntax, and structure conventions. RDF contain resources that promote data merge, allowing structured and semi-structured data combination, exposition, and sharing across different applications (W3c, 2004c). RDF as a data model has two key data structures: RDF graphs and RDF datasets (W3c, 2014a).

An RDF dataset holds a set of RDF graphs consisting of one default graph and zero or more named graphs (W3c, 2014a, 2014e). An RDF graph is a set of triples pattern defined by triad subject-predicate-object. The graph nodes are the subjects and objects of triples, and the edges connecting two nodes are predicates (property) of the triple. A subject is an IRI or a blank node, a predicate is an IRI, and an object is an IRI, a literal or a blank node (W3c, 2004c, 2014a). Figure 28 shows an example of a simple RDF graph with only two nodes and one edge.

Figure 28: Example of RDF triple graph



Source: Prepared by the author.

[20] See details at http://www.ietf.org/rfc/rfc3987.txt

The primary representation language of the Semantic Web is the RDF language. There are multiple syntactic representations for the RDF model, some more suitable for machine processing, others more readable for people. An example is the RDF/XML notation standardized by W3C. To have more details on RDF[21], we recommend to read references W3c (2004c), (W3c, 2014a, 2014b, 2014c, 2014d, 2014e), Miller (1998), and Pan (2009).

In the semantic level, the first layer deal with the taxonomy of data. At this layer, we have RDF Vocabulary Description Language known as RDF Schema (RDFS). RDFS extends the basic RDF vocabulary providing a "*data-modeling vocabulary for RDF data (W3c, 2014f)*". RDFS can be used to describe classes and properties reflecting statements about data resources. RDFS introduces a layer that specifies some characteristics adding semantics to data defined in RDF. The elements to be used in an RDF graph can be defined using RDFS[22] (W3c, 2014f).

Next layer is the so called ontology layer that provides language support for creation of ontologies. Web Ontology Language (OWL) is an ontology language for the semantic web with formally defined meaning. OWL is a language that extends RDF and RDFS using an XML-based syntax. OWL adds more details to the vocabulary used to describe properties and classes, along with a formal semantics. Thus, the primary goal of OWL is to bring the expressive and reasoning power of description logic to the Semantic Web. In relation to the expressiveness of knowledge, there are three incremental sublanguages of OWL[23] (W3c, 2004b, 2012b, 2012c):

- *OWL Lite*: is the simplest OWL language used for taxonomies and simple constraints. OWL Lite is a language with extremely poor expressiveness, although it is simple to understand and easy to implement in automatic procedures.

- *OWL DL*: support description logic capabilities and includes all constructs of the OWL language. OWL DL is able to compose satisfactory expressive features without losing computational completeness and decidability.

- *OWL Full*: designated to ensure maximum expressiveness and syntactic freedom of RDF but without computational guarantees. OWL Full contains the full OWL vocabulary but does not impose any syntactic constrains.

[21] The resources in the RDF vocabulary have URIrefs http://www.w3.org/1999/02/22-rdf-syntax-ns#

[22] The resources in the RDF Schema vocabulary have URIrefs http://www.w3.org/2000/01/rdf-schema#

[23] The resources in the OWL have URIrefs http://www.w3.org/2002/07/owl#

Each of these OWL sublanguages is an extension of its own predecessor, both in relation to what can be expressed and in relation to what can be concluded from reasoning.

Finally, the last layer of Semantic Web architecture is the query layer. The *Simple Protocol and RDF Query Language* (SPARQL) is a SQL-like language for querying RDF data. One can make an analogy between SPARQL and the SQL query language for relational databases, considering that SPARQL has a syntax (Figure 29) suitable for querying data represented as a set of RDF triples (Ducharme, 2011, Feigenbaum, 2009, W3c, 2008, 2013a, 2013b).

Figure 29: Basic elements of SPARQL SELECT query

```
# Prefix declaration (shortcuts of URIs - optional)
PREFIX x: <…>
PREFIX y: <…>
…

# Query result clause and parameters to be found
SELECT …

# Define which sources of data will be consulted (optional)
FROM <…> OR FROM NAMED <…>

# Query pattern to be located in the sources of data
WHERE {
    …
}

# Query modifiers (optional)
GROUP BY …
HAVING …
ORDER BY …
LIMIT …
OFFSET …
VALUES …
```

Source: Prepared by the author.

SPARQL allows querying RDF datasets that follows the RDF specification of the W3C (W3c, 2008). There are four types of SPARQL Queries (Ducharme, 2011):

- SELECT query: Project out the list of values of variables and expressions added to a query pattern.
- CONSTRUCT query: Returns an RDF graph constructed by substituting variables in the query pattern.
- ASK query: Returns a Boolean value (true or false) indicating whether there are any matches against the query pattern.

- DESCRIBE query: Returns an RDF graph describing the resources matched by the given variables.

SPARQL also permits INSERT, UPDATE and DELETE operations against RDF datasets. SPARQL is a robust language allowing the semantic web user to access the huge amount of data available on the web of data. For more details about the SPARQL language, we recommend to read Ducharme (2011), Feigenbaum (2009), and W3c (2008, 2013a, 2013b).

2.6 Context of information and information systems in healthcare

Medical researchers and health professionals are constantly dealing with a growing volume of information originated from various sources of biomedical information and available in a variety of formats such as text, images, video, and sound stream. These variables generate a great challenge for medical practice because they make complex disciplines such as organization, representation, and information retrieval (Freitas; Schulz; Moraes, 2009). In addition, the large volume of medical information produced daily, it is necessary to manage information available in multiples sources and formats. We understand as a source of information any document, data, or record capable of providing to information users, information that when accessed are able to respond to their needs. There are a variety of sources of information in medicine, such as glossaries of terms, controlled vocabularies, terminologies, and ontologies (Oliveira; Almeida; Quintela, 2013). Mcglynn *et al.* (1998, p.7) lists some information sources for integrated information systems:

- Administrative files: Consist of administrative data created by the interaction between a patient and healthcare provider during a care.
- Clinical information: Information that reflects patient health state available in medical records, examination results, and other clinical data sources.
- Enrollment files: Consist of data available in employer or government database about people enrolled.
- Surveys: Data originated from research related to health issues such as patient health status, health risks, patient behavior, and satisfaction surveys.

Health information should be seen as an important decision-making tool, supporting health action planning activities, as well as implementation and evaluation activities. Information collected in healthcare units allows to analyze of the socioeconomic, demographic, and epidemiological reality of patients. In fact, *"better information means safer health care (Healthit.Gov, 2017, on-line)"*.

Given the large amount of information available to manage healthcare information resources, healthcare institutions adopt health information systems. In short, one defines health information systems (HISs) as information systems capable to process data, information, and knowledge in healthcare. HIS intend to provide a high-quality and efficient patient care (Haux, 2006). HISs are systems that gather, store, manage, or share information regarding the individuals' health, clinical care, or any activity performed in health-related organizations (Sirintrapun; Artz, 2016).

Adoption of information systems in the healthcare area gained strength in the middle of the last decade as verified by Cho; Mathiassen and Nilsson (2008). Use of healthcare information systems offers potential to improve patient safety and, as consequence, to improve the quality of care provided, as observed by Helms; Moore and Ahmadi (2008). Nevertheless, these authors also observe the existence of weaknesses in the use of these systems, for instance, the lack of system integration (Helms; Moore; Ahmadi, 2008). In addition, use of information systems in the health area allows the decentralization of information and increases the collaborative work since it enables the sharing of information regarding patients' health. Healthcare professionals can use the information shared by these systems to improve patient care.

According to Sirintrapun and Artz (2016), there are four categories of HIS: foundational systems, financial systems, departmental systems, and electronic medical records. Here, we focus on electronic medical record sometimes called electronic health record, but with some specificities.

2.6.1 Electronic Medical Record and Electronic Health Record

To understand electronic medical record (EMR) and electronic health record (EHR), first, it is necessary to review the document that originated it, the patient's medical record, or medical record. An important contribution from Galvão and Ricarte (2012, p. 2) is their study on understanding the definition of medical record. These authors contextualize the definition from a set of related concepts, namely: understanding of the meaning of health considering the World Health Organization (WHO) approaches; observation of the healthcare professional role; the principle of integral health assistance that assure to individuals a health care service at all levels of assistance; and finally the right to healthcare information.

Therefore, Galvão and Ricarte (2012, p. 6) defining medical record as a collection of information regarding the health status of a patient, accessible to the patient himself, which must be stored and shared with the security of the authorized users to access. The patient's medical record, also known as patients' chart, gathers essential information of patients' health status helping to guarantee the continuity of the treatments provided to the patient (Massad; Marin; Neto, 2003). According to the Brazilian Federal Council of Medicine (CFM), a medical record is a unique document consisting of a set of information, signs, and images registered, originated from facts, events, and situations of both patient's health and the care provided. A medical record has a legal nature, confidential and scientific, that allows the communication between members of the multiprofessional team and the continuity of the assistance provided to the individual (Cfm, 2002).

Romm and Putnam (1981) refer to medical record as source or repository of patient care information. In addition, Browning (2006, p. 1723-1724) argues that medical records are a set of information patient-related including written notes, graphs, test results, x-rays, and other data of patient's health status. Medical records help healthcare practitioners to follow up patient's health status and to keep legal requirements compliance.

Originally, medical records were in paper and usually maintained at the institution that generated these records. The fact of each institution maintaining its own record resulted in duplication of information. Both the format and content of these records varied according to the institution, making it difficult to exchange information. The sharing of these records between institutions was poor and difficult. The large volume of paper documents also generated problems for their retrieval and required large physical spaces for storage (Galvão; Ricarte, 2012, Massad; Marin; Neto, 2003).

Several studies demonstrate the disadvantages and limitations of paper medical records, for instance, Roukema *et al.* (2006), Stausberg *et al.* (2003); Tufo and Speidel (1971); and Cfm and Sbis (2012). The Brazilian Federal Council of Medicine listed some of these limitations (Cfm; Sbis, 2012):

- Unavailability of medical records to more than one professional at the same time.
- Low mobility to travel between physically distinct health care institutions with flexibility and agile time.
- Subject to illegibility of handwriting.
- Frequent loss of information due to mislay or damage of the paper.
- Difficulty to collective research.
- Lack of standardization of information.

- High cost for storage and archiving.

Thus, given all difficulties to use medical records on paper and significant advances and availability of information and communication technologies, which promote solutions for information storage, retrieval, and exchange, emerge the need for the Electronic Health Record (EHR) (Bezerra, 2009, Pinto, 2007). In short, EHR is a medical record in electronic format or supported by information and communication technologies (Häyrinen; Saranto; Nykänen, 2008, Onc, 2017b, Rector; Nowlan; Kay, 1991, Reis; Correia; Pereira, 2011). Table 6 clarifies three related concepts, Electronic Medical Records (EMR), Electronic Health Records, and Personal Health Records (PHR), created from Cfm and Sbis (2012), Ludwick and Doucette (2009) and Onc (2017c).

Table 6 - Differences between EMR, EHR, and PHR.

Term	Description
Electronic Medical Records	A medical record available in electronic format, containing patient's clinical data gathered during a healthcare encounter in a specific healthcare facility. EMR is available just for this specific healthcare facility and usually are not able to electronic sharing between facilities.
Electronic Health Records	EHRs are real-time patient's care records that provided by multiples healthcare facilities involved in a patient's care, and with access available to authorized users. An EHR both joins and shares information of patient's care between several health care providers.
Personal Health Records	PHRs are similar to EHRs but intended for manipulation by patients to create, maintain, access, and manage their health data.

Source: Prepared by the author.

EHR should gather information about physical, mental, and social stage of patients during a medical care (Cfm; Sbis, 2012, Galvão; Ricarte, 2012, Häyrinen; Saranto; Nykänen, 2008, Massad; Marin; Neto, 2003, Onc, 2017a, 2017b). Such information involves:

- Patient's demographic data such as name, address, birth date, gender, race, ethnicity, and so on.
- Emergency contact information.
- Symptoms reported.
- Patient's vital signs and basic health information as blood type and weight.
- Patient's medical history and patient's family medical history.
- Disease diagnoses and allergies.
- Medications taken and immunization with dates.
- Treatment plans and surgeries.

- Radiology images and laboratory test with results.

According to the literature, there are several identified benefits and advantages for the adoption of EHR, highlighted next (Cfm; Sbis, 2012, Galvão; Ricarte, 2012, Jamoom *et al.*, 2012, Massad; Marin; Neto, 2003, Onc, 2017b, Reis; Correia; Pereira, 2011, Tufo; Speidel, 1971):

- Independence of physical location.
- Reduction of paperwork.
- Reduction of information storage costs.
- Reduction of information redundancy (duplication).
- Possibility of information sharing among professionals and healthcare institutions.
- Distributed access to an individual's health information, improving information on the patient's past life.
- Improvement on actions of coordination of care.
- Improvement of patient participation in their own healthcare.
- Possibility of access to records by more than one professional at the same time.
- Reduction of information retrieval costs.
- Facilitation in information retrieval.
- Minimization of information loss when comparable to paper durability and plausibility to mislay.
- Improvement of information security.
- Improvement in the access control to information.
- Greater agility in solving health problems and improving the effectiveness of care due to information availability.
- Enabling of continuous data collection, mainly by health devices.
- Improvement in the decision-making processes by providing access to updated healthcare information.
- Improvement of the readability of information available in medical records, avoiding handwriting confusion.

- Narrowing the communication channel between health care providers and patients.
- Improvement of the integration of patient information available in different formats, such as texts, images, multimedia, and others.
- Improvement of the adhesion to clinical and care protocols, standardizing the information in the medical record.
- Reduction of health operation costs avoiding requests for exams in duplication or unnecessary tests.

Notwithstanding these vast list of benefits and advantages, some studies highlight EHRs adoption disadvantages (Häyrinen; Saranto; Nykänen, 2008, Ludwick; Doucette, 2009, Menachemi; Collum, 2011, Plantier *et al.*, 2017, Poissant *et al.*, 2005, Wang *et al.*, 2003):

- High costs of acquisition, licensing, implementation process, and maintenance.
- Long time required to implement.
- Staff anxiety during implementation periods.
- Staff anxiety due to alterations in work process already known.
- Lack of standardization of data required and formats.
- Unsuitability of EHRs to the real set of data needed by the professionals involved.
- Lack of fit between the EHR data and routine clinical practice.
- Long time spent by clinicians to document data in EHR instead to provide direct care to patients.
- Cost and time of learning to use EHR systems.
- Lack of patients' confidence in the privacy and security of access to their information.
- Instable technologic environment causing temporary unavailability.

Complementing, Sirintrapun and Artz (2016), use the term EMR in the sense of EHR as explained in Table 6. These authors mention as advantages brought by EMRs:

- Integrative virtual work environment.
- Accessibility and availability of information.

- Availability of messaging and alerts.
- Patient care safety.
- Error reduction.
- Information capture and management.

As disadvantages, Sirintrapun and Artz (2016) highlight:

- High cost.
- Training needs.
- User resistance to use.
- Change in the workflow.
- Poor results visualization.
- Technical issues.

EHR is the most common communication tool used among health professionals in healthcare facilities. It focuses on the integration of clinical and administrative information of patients, gathering information and making it available for use by healthcare professionals, hospitals, insurance companies, clinics, laboratories and other health entities. In addition, the EHR should be able to provide interoperability between the health information systems enabling data exchange between them. EHR also must meet the following pre-requisites: standardized structure and agreement on terminology, clear rules for communication, archiving, security and privacy (Galvão; Ricarte, 2012, Massad; Marin; Neto, 2003).

Data content of EHR could be organized in chronological order; by type of problems, diseases or subjects; by the source of the information source, or by the combination of these three elements (Reis; Correia; Pereira, 2011).

2.6.2 Electronic discharge summaries

Regarding information related to hospital discharge, one of the narratives that compose the EHR is the electronic discharge summary (EDS), which should have information about all events performed during a patient care hospitalization (Ferreira *et al.*, 2014, Galvão; Ricarte, 2012, Häyrinen; Saranto; Nykänen, 2008, Massad; Marin; Neto, 2003, Pinto, 2007).

EDS endeavors to subsidize the healthcare providers with summarized data on patient's hospitalization. It is an important element to establish the communication between the hospital and primary care and to guarantee a better continuity of care following patient's discharge. An EDS contains information about patient's hospitalization that guides the patient's medical follow-up, usually comprising (Acsqhc, 2010, Callen; Alderton; Mcintosh, 2008, De Souza, 2012, Ferreira *et al.*, 2014, Hiqa, 2013, Lapelle *et al.*, 2006, Paterson; Allega, 1999, Queiroga, 2014):

- Patient's admission motives, admission date.
- Patient's origin in the case of referral or transfer of the patient from other healthcare facility or medical specialty.
- Results of any examination performed.
- Examination finding.
- All given diagnosis.
- Short description of any treatment administrated.
- Complete relation of medication administrated, including patient's response, complications, and adverse events when observed it.
- Discharge date and patient's conditions at discharge.
- Prescription of medications to use after the patient's hospitalization discharge.
- Description of treatment follow-up needed after discharge.

Admittedly, EDS is an essential document for the continuity of patient care. Together with the EDS registration and storage in the healthcare institution, this document must be delivered to the patient upon discharge from hospital. This enables the patient to bring it with him/her in healthcare follow-up after discharge (Acsqhc, 2010, Greer *et al.*, 2016, Klück; Guimarães, 1999, Lapelle *et al.*, 2006, Raval; Marchiori; Arnold, 2003).

Efficacy of the hospital discharge summary has a direct relation to the reduction of risk to patient's health and improvement in patient's care. It includes factors such as the timeliness of receiving the EDS, legibility of the information registered in EDS, and completeness of this information (Acsqhc, 2010, 2011a, 2011b, Hiqa, 2013).

The Brazilian Ministry of Health, through ordinance no. 1459/2011 (Brasil, 2011a), establishes the Stork Network Program (known in Portuguese as "*Rede Cegonha*") under scope of the Brazilian Unified Health System (SUS). Stork Network Program invests in the construction of a care network focused on guaranteeing to women the entitlement to reproductive planning, and humanized care for pregnancy, childbirth and puerperium, and ensure children entitlement to have safe birth and healthy growth and development (Brasil, 2011a, Brasil, 2012).

The Stork Network concentrates in four general components (Brasil, 2011a):

- Prenatal follow-up.
- Intrapartum: Childbirth and delivery.
- Postpartum: Puerperium (mother side) and integral attention to children's health.
- Logistic system: sanitary transport and regulation.

As regards the healthcare of both the woman and the newborn, there are many medical specialties involved in the care performed and different phases of care: prenatal, intrapartum and postnatal. In addition, the structure of the healthcare network of SUS allows healthcare to be carried out in multiple healthcare institutions, and perhaps located in different geographic regions (cities and states) and managed by different levels of government or political jurisdictions (municipal, state, federal) (Brasil, 2006, 2007, Brasil; Saúde; Estratégicas, 2004, Marques *et al.*, 2015).

Therefore, reflecting on obstetric and neonatal care, a major challenge for Brazilian healthcare institutions is to ensure adequate continuity of care for women and the newborn. In order to ensure maternal and neonatal health, it is fundamental to maintain data of care for both the woman and her newborn since the delivery (Brasil, 2006). In this sense, EHRs related to care in the prenatal, labor, delivery, puerperal and newborn phases are necessary to ensure continuity of care for both infant and mother. Consequently, the obstetric discharge summary is a strategic document to guarantee the continuation of maternal and child care after hospital discharge (Ferreira *et al.*, 2014).

Computerization of health institutions, as well as the capability of health care data storage, emerge the need for semantic interoperability between health information systems. Since different health information systems adopt several terminologies and semantic representation of medical data, arise the need to establish semantic standards to allow the exchange of information (Freitas; Schulz; Moraes, 2009).

2.6.3 Biomedical ontologies and terminologies

The field of health possesses a large amount and diversity of information originated from various sources such as glossaries of terms, controlled vocabularies, terminologies, and ontologies. In this thesis, a source of information is a document, data or record capable of providing users of information services, information that when accessed are able to respond to their needs. Rosenbloom *et al.* (2006) mention that EHR systems can benefit from the use of interface terminologies since such terminologies "*provide the translation from clinicians' own natural language expressions into the more structured representations (Rosenbloom et al., 2006, p. 277)*".

In order to manage this diversity of information resources, health institutions adopt several different information systems that in general need to exchange information among themselves (Freitas; Schulz; Moraes, 2009). Thus, we quoted Schulz and López-García (2015) that describe different types of knowledge organization systems in the medical field, which seeks the development and improvement of information and data resources more than any other knowledge area, they are summarized at Table 7.

Table 7 - Types of knowledge organization systems in the medical field

Type	Description
Thesaurus	Describe existing specialized terms and uses semantic relationships such as synonymy. Example: MeSH.
Classifications	An ordered category of terms grouped according to its similarities. Example: International Classification of Diseases (ICD 10.
Nomenclatures	They are standardizations of terminologies. Example: SNOMED.
Formal ontologies	Classification systems that describe classes of domain objects using formal logic. They usually link a terminology in the strict sense, as a bridge to the primary ontological representation units independent of the language. Example: OBO foundry ontologies.

Source: Based on Schulz and López-García (2015)

In terms of clinical terminologies, Rector (1999) presented an interesting discussion of the challenges to developing such terminologies. Rosenbloom *et al.* (2006) observed the emergence of various terminologies for the representation of clinical knowledge, but none of them was chosen as a universal standard. According to these authors, some standards-based organizations point the adoption of some of them or just part of these terminologies. Elkin *et al.* (2002) establish a guideline providing a set of principles to follow when developing and maintaining controlled health vocabulary. Their define controlled health vocabulary as:

> *"A terminology intended for clinical use. This implies enough content and structure to provide a representation capable of encoding comparable data, at a granularity consistent with that generated by the practice within the domain being represented, within the purpose and scope of the terminology Elkin et al. (2002, p. 185)".*

According to Ceusters; Smith and Flanagan (2003), biomedical ontologies, terminologies, and controlled vocabularies configure a solution popularly adopted in information systems in order to represent medical information. In addition, Elkin and Tuttle

(2012) remark about the reputation of terminologies in healthcare as a tool of knowledge representation and its importance helping the systems interoperability. These authors discuss the demand on healthcare terminologies as unambiguous and nonredundant healthcare knowledge representations since "*the goal of healthcare terminologies was and is to aggregate patient descriptions by meaning*" (Elkin; Tuttle, 2012). Both Ceusters; Smith and Flanagan (2003) and Elkin and Tuttle (2012) point to a large number of medical terminologies as a mere list of terms organized hierarchically. Table 8 summarizes the four categories taxonomical vocabularies cited by Massad; Marin and Neto (2003).

Table 8 - Taxonomical vocabulary categories: usage and purpose.

Vocabulary	Category	Purpose	Formalism	Example
Intentional compositional vocabulary	Terminology	Facts recovery	Logical description	OpenGalen
Intentional combinatorial vocabulary	Nomenclature	Case recovery	Semantic Grammar	SNOMED
Extensional disjunctive vocabulary	Classification	Statistical evaluation, partitioning of real objects	Hierarchical and deterministic decision trees	ICD
Extensional associative vocabulary	Thesauri	Literature recovery	Frame type approach	MeSH

Source: Based on Massad; Marin and Neto (2003)

Moreover, ontologies make it possible to represent knowledge in a way that is closer to reality, which is fundamental for a biomedicine (Freitas; Schulz; Moraes, 2009). Ontologies have been widely used in representation and organization of specialized knowledge in various fields of biomedicine. Specifically, ontologies are used in the standardization of medical vocabularies as an alternative to medical terminologies, since greater expressiveness for the representation of knowledge by using logical formalisms (Ceusters; Smith; Flanagan, 2003, Smith, 2008a, Smith; Brochhausen, 2010).

To a large extent, ontologies have been adopted in the biomedical field to deal with the massive bodies of information and knowledge make available each day (Rector; Rogers, 2006, Rosse; Mejino, 2003). On the other hand, as the number of ontologies and terminologies grows, due to the lack of standardization, ambiguities and inconsistencies between the existing terms in them also grow. In effect, the consistency maintenance becomes complex (Simon; Fielding; Smith, 2004). Then, Simon; Fielding and Smith (2004) suggest the use of the formal ontology BFO in order to standardize the ontology development minimizing ambiguities of reality understanding. Bittner and Smith (2004) also defends the adoption of top-level ontology such as BFO to normalize a domain ontology since this top-level ontology supplies a set of fundamental class (domain neutral), and a list of the relations and axioms, in order to guarantee a semantic across ontologies. Both the work of Bittner and Smith (2004) and Grenon; Smith and Goldberg (2004) demonstrated the use of BFO in the building of biomedical ontologies.

Second Smith (2008a), date from the 1930s the first initiatives on standardization of biomedical terminologies. These pioneer efforts aim to allow a clear understanding of specialized terms in a global scale. Later, some initiatives to interoperate biomedical terminologies arose. The first initiative is International Classification of Disease (ICD), currently the tenth edition. Some approaches by means of controlled vocabularies are the Medical Subject Headings (MeSH) (Lowe; Barnett, 1994), the Systematized Nomenclature of Medicine (SNOMED) (Stearns *et al.*, 2001), and the Medical Dictionary for Regulatory Activities (MedDRA) (Brown; Wood; Wood, 1999, Fescharek *et al.*, 2004). Following by the creation of a comprehensive platform called the Unified Medical Language System (UMLS) (Bodenreider, 2004), in a tentative to integrate many controlled vocabularies. Another initiative, the Logical Observation Identifiers Names and Codes (LOINC) (Mcdonald *et al.*, 2004) is a database and universal standard created in order to identify medical laboratory observations. Following the theory of BFO and Ontological realism, a group of ontologies is available at OBO Foundry consortium (Smith *et al.*, 2007a). There are for example the Gene Ontology (GO), and the Foundational Model of Anatomy (FMA).

As previously seen, ontologies and terminologies can be applied to establish semantic patterns on a domain of knowledge. For more details on biomedical terminologies see the work of (Freitas; Schulz; Moraes, 2009) that presents the state of the art of terminologies and ontologies applied to Biology and Medicine.

2.6.4 Interoperability in Healthcare Information Systems

In subsection 2.2, we discussed concepts about interoperability. We considered that a semantic interoperability solution involves the adoption of approaches capable of ensuring uniform interpretations (meanings of terms) among systems, regardless the KOS employed, for example terminologies, thesauri, or ontologies. In section 2.3.4, we present in a greater detail the application of ontologies for solution of interoperability issues among information systems.

In Brazil, concerns about interoperability among health information systems become evident with the decree nº 2.073/2011 that regulates the use of interoperability and health information standards. One of the recommendations of this decree is precisely to adopt ontologies and terminologies to deal with interoperability issues on health information systems (Brasil, 2011b).

As mentioned earlier in this work, health information systems can be different in respect of the meanings of the same term, which limits the exchange of information and communication among systems. This finding can be identified in any information systems, regardless of the domain of knowledge that it covers. Section 3 discuss the application of solutions for interoperability among information systems in several fields of research.

Indeed, since the early 1990s, ontologies have been widely applied in the fields of medicine and biomedicine as a way of structuring the large volume of data available. Then, these fields have covered the research on the interoperability of SISs using ontologies (SIMON; FIELDING; SMITH, 2004), as it is evident by the several international initiatives produced with this technology.

According to REIS et al (2011), the adoption of solutions for health information systems integration that allow the exchange and sharing of information between different health institutions become increasingly necessary, insofar as the individual or patient may be assisted in multiple healthcare institutions. However, the heterogeneity of semantic patterns that different health information systems adopts makes this integration a hard task.

For the feasibility of semantic interoperability among health information systems handling RES, the ISO/TR 20,514 standard determines that a specific RES must meet four prerequisites (NEIRA et al, 2011), namely:

- A standardized reference model.
- A standardized service interface model to provide interoperability among healthcare services and others.
- Standardized conceptual models for specific knowledge domains.
- Standardized terminologies integrated with controlled vocabularies.

A proposal for the obstetrics discharge summary was developed with the aim of the standardization and consequently transmission of information through the OpenEHR reference model (FERREIRA et al, 2014). This model is determined as the default in Brazil by the decree nº 2,073 / 2011 (BRAZIL, 2011b).

In Brazil, integral healthcare services have been planned since the 1988 Constitution, but the concern with interoperability of information systems only came to the scene in 2011 with the decree no 2,073. Regarding interoperability for definition of RES, the decree nº 2,073/2011 demands the use of the OpenEHR[24] reference model, and also establishes other standards, for example:

- HL7[25] standard for the integration of both results and requests of exams.
- SNOMED-CT[26] terminology for coding clinical terms and mapping of national and international terminologies used in the country.
- Private Health Insurance and Plans Information Exchange Standard (TISS[27]) for interoperability with supplementary healthcare systems.
- ISO 13,606-2 standard for interoperability of knowledge models.

The decree also establishes also the use of international medical terminologies and some Brazilian terminologies/standards to enable semantic interoperability among health information systems (BRASIL, 2011b), such as:

- International Classification of Diseases (ICD).
- International Classification of Primary Care (ICPC-2).
- The Brazilian Unified Health Terminology Supplement (TUSS[28]).
- Hierarchical Brazilian Classification of Medical Procedures (CBHPM[29])(Amb, 2016).

The complexity of the medical activity, the vast amount of information produced, the health care activities distributed by multiple healthcare facilities, as well as the need of assuring the continuity of healthcare to individuals, require the integration and exchange of information among institutions and their information systems.

---

[24] Details at: https://www.openehr.org/

[25] Details at: http://www.hl7.org/implement/standards/

[26] Details at: https://www.snomed.org/snomed-ct

[27] A Brazilian standard. Details at: http://www.ans.gov.br/the-sector/information-exchange-standard

[28] In Portuguese "*Terminologia Unificada da Saúde Suplementar* (TUSS)".

[29] In Portuguese "*Classificação Brasileira Hierarquizada de Procedimentos Médicos* (CBHPM)".

PART II

Empirical phase

3 RESEARCH METHODOLOGY

"We do not know a truth without knowing its cause".
(Aristotle, Nicomacheian Ethics, I.1).

A scientific research methodology is the set of systematic and rational activities that allows achieving an objective, tracing the way to follow, detecting errors, and assisting decisions. To achieve the scientific research objectives, a researcher must use a set of intellectual and technical procedures known as scientific methods to find answers to research problems (Collis; Hussey, 2014, Gil, 2008, Hegde, 2015, Kumar, 2011, Marconi; Lakatos, 2003, 2007). Thereby, this current chapter presents the scientific research methodology for this study, including the methodological steps taken to achieve this research goal.

Particularly, establishing the suitable research paradigm to this study became a challenge because its research goals differ from traditional or natural research by does not focus on building a new Information Science theory, but rather apply a theory to solve problems. Research paradigms determine methods and procedures of research, they help to conduct how to make decisions and perform the research and denotes a conceptual framework guiding how to examine the research scenario and what to do to find a solution (Mackenzie; Knipe, 2006).

We notice an alignment of this research with the ideas of Gibbons *et al.* (1994) that suggest that investigations should seek the construction of knowledge applicable to the organizations. It is natural to think of knowledge production coming exclusively from classical scientific methods as in the positivist and anti-positivist paradigms. However, the production of knowledge also happens by applying previously existing scientific knowledge applied to solving real problems in organizations. (Gibbons *et al.*, 1994)

Therefore, the focus of this study is the application of a valid theory of Information Science - which is the use of ontologies as a knowledge organization tool – as a viable alternative to solve problems related to interoperability, organization, and recovery of information into the biomedical and healthcare organizations, more specifically knowledge of the obstetric and neonatal domain.

3.1 Scientific research methodology

We delimited the research regarding its typology on natural science. The research classification determines the theoretical framework and identifies possible approaches from the empirical point of view, seeking to confront the theoretical view with the data of reality. In order to do so, one sought some references in the literature on natural science research methodology such as Creswell (2009, 2013), Collis and Hussey (2014), Kumar (2011), Neuman (2014), Mason (2002), and (Yin, 2009). There are different, but complementary views on approaches to classifying a study according to the references studied. Given the relevance of defining some approaches, this study considers the problem approach, research nature, objectives of a study, and technical procedures.

Regarding *the problem approach,* this research is **qualitative** as it seeks to understand both the phenomenon investigated (low integration of data between health systems) and the proposed solution (construction of a single semantic vocabulary interoperable). Moreover, all learned knowledge is not quantifiable

Considering *the research nature or application perspective*, this study is **applied research**, because first generate knowledge of practical application of the ontologies oriented to the solution of semantic interoperability problems. Secondly, it generates results of direct practical application to several organizations. The ontology of the obstetric and neonatal domain is suitable in the context of both the University Hospital of the Federal University of Minas Gerais and in any other organization or business that needs a vocabulary with the knowledge covered by it.

Concerning *the objectives of a study*, this is an **exploratory research**, owing to provide greater familiarity with both a problem and solutions. This study seeks to improve the discussion of the Information Science social role helping any scientific field in reference to solutions related to representation, organization, and retrieval of information.

Finally, regarding *the methods and technical procedures* performed, this study is a **case study** in such a way that involved a deepening of knowledge about a phenomenon within its real-life context (Yin, 2009). This case study investigates the gap of data exchange in the medical field and ontology as an interoperability solution, aiming to describe and explain the phenomenon, and proposing an able alternative to solve phenomenon related problems.

3.2 Demarcation of case study

A research journey involves planning of how the research will be performed to achieve the defined objective. All study starts with the definition of a research problem. Deciding what to investigate will guide the type and design of the research (Kumar, 2011). Previously the identification of the research problem, the research participated in some meetings at University Hospital of UFMG, in which observed that participants mention on gaps of data exchange among the hospital information systems (Figure 30).

Figure 30: Researcher's observation of problem issues.

Observations

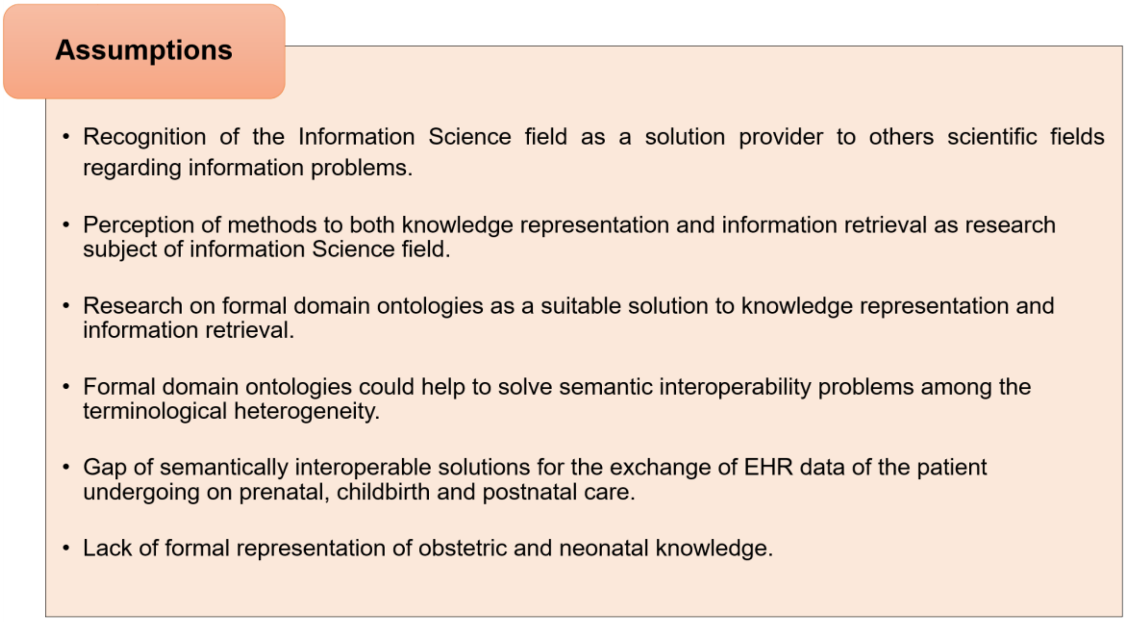
- Complexity of the structure of the Brazilian healthcare network under the scope of the Brazilian Unified Health System that allows healthcare to be carried out in multiple healthcare institutions, in which perhaps located in different geographic locations and managed by different political jurisdictions.
- Necessity to maintain health care data, for both the woman and her newborn, related to care in the prenatal, labor, delivery, puerperal and newborn phases, in order to help ensure adequate continuity of care for women and the newborn.
- Recognition of the medical record, in paper or electronic, as a document used by health professionals to record patient data gathered during a care.
- Recognition of the medical record, in paper or electronic, as an important communication tool both to health professionals than health institutions.
- Fragmentation of the care information of the obstetric and neonatal patient into various information systems with minimal semantics standardization among them.

Source: Prepared by the author.

This perception of the researcher led to an in-depth investigation into the issue. Thus, the researcher performed a preliminary bibliographical review, seeking out to understand knowledge in the biomedical field and to discern the subject around the problem observed during the previous meetings. This preliminary literature review helps understanding highlights issues of some possible research problem. We searched bibliographic sources, both primary and secondary sources that could guide the research and even justify the proposed solution to the problem in question. Based on studies analyzed, it highlighted some assumptions (Figure 31).

Then, to establish the research problem, we also employed some procedures of unstructured interviews and observations to gathering primary information of the environment visited (Maternity Hospital of UFMG). These procedures clarified the reason why the institution needs to exchange information between different systems, as well as helped to identify both the architecture of existing systems and what information needs to exchange (see Figure 1). Hence, this case study established the research question and general goal showed in chapter 1.

Figure 31: Researcher’s assumptions of problem issues.



Source: Prepared by the author.

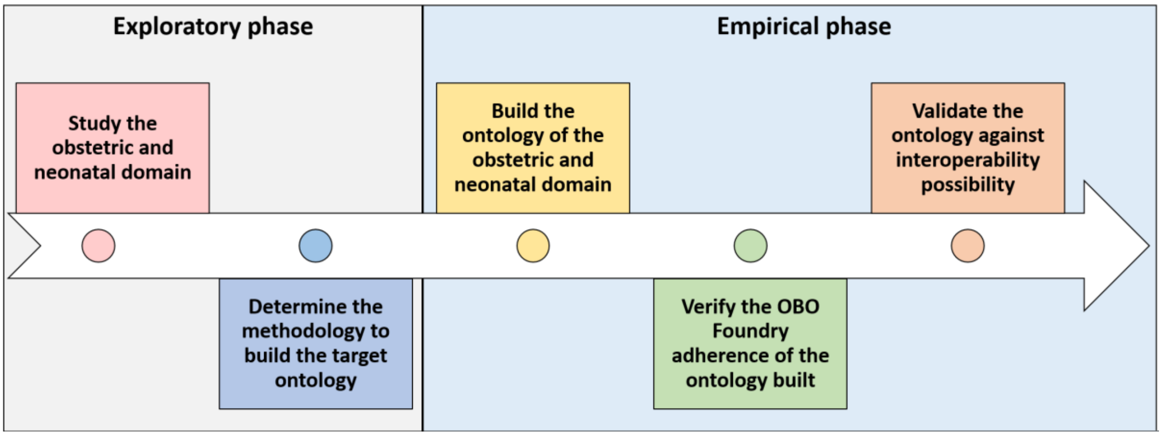
It also sought to develop a consistent bibliographical material, which serves as a reference for future work, given the shortage of similar studies in both Information Science and Biomedical Informatics fields, without the pretension of exhausting the subject. Studies found by literature review helped in forming the knowledge base for the researcher to conduct this research, promoted a better understanding of the main issues that driven to research problem identification, and validated the feasibility of the proposed solution. The results of this literature review compound part of chapter 2.

After the definition of what to investigate, it should define the set of procedures necessary to achieve the research goal. These procedures compose the research methodology described in next subsection.

3.3 Research methodology

This section presents the general description of the research methodology, explaining the methodological steps taken to achieve the specific objectives of the research. We organized these steps as presented in the scheme of Figure 32. The research was divided into two phases, namely, exploratory and empirical phases. The exploratory phase aimed to provide greater familiarity of the researcher with the knowledge involved in the research. The empirical phase consisted of the representation of the knowledge acquired in the previous phase through a domain ontology as well as performs of a proof of concept to validate the possibility of the application of the ontology in the promotion of interoperability.

Figure 32: Steps of the research methodology of this thesis.



Source: Prepared by the author.

During the exploratory phase, we performed two steps: 1st) Study the obstetric and neonatal domain; and 2nd) Determine the methodology to build the target ontology. While in the empirical phase, we performed three steps: 3rd) Build the ontology of the obstetric and neonatal domain; 4th) Verify the OBO Foundry adherence of the ontology built; and 5th) Validate the ontology against interoperability possibility.

3.3.1 Study the obstetric and neonatal domain

The first step was to ***study the domain of knowledge*** covered by the target ontology. The purpose of this step was to acquire knowledge on the obstetric and neonatal domain. Hence, to acquire knowledge, the researcher employed direct methods, for instance, literature review, document analysis, interviews, and observations.

The literature review and document analysis helped the researcher understand the knowledge of the ontology target domain. The researcher identified and selected sources of knowledge that served as a reference to all ontology development. The identified sources of knowledge include textbooks, scientific journals, reports, manuals, healthcare protocols, health record templates, case studies, user stories, among others.

We retrieved some relevant documents from the Virtual Library of the Ministry of Health of Brazil[30]. The main documents analyzed consisted of clinical protocols, primary care protocols, and therapeutic guidelines. Some of these documents are:

- Protocols of Primary Care: Women's Health[31] (Brasil, M. D. S. D., 2016).

---

[30] Virtual Library of the Ministry of Health of Brazil is available at http://bvs.saude.gov.br/

[31] Title in Portuguese: Protocolos da Atenção Básica : Saúde das Mulheres.

- Pregnant Woman's Booklet[32] (Brasil, 2014a).
- Prenatal and Puerperium: qualified and humanized care[33] (Brasil, 2006).
- Ambulatorial perinatal record (Brasil, 2014b).

We identified several health record templates or standards from different countries. Some of these health record templates are:

- ACOG Woman's Health Record (Acog, 2010b).
- ACOG Antepartum Record and Postpartum Form (Acog, 2010a).
- Children's Electronic Health Record Format (Ahrq, 2013).
- Children's EHR Format Enhancement: Final Recommendation Report (Wald Js, 2015).

The researcher resorted to textbooks of obstetrics, gynecology, embryology, anatomy, and pediatrics. The references section of this thesis lists some of these textbooks. We also, retrieval some case studies from these books:

- PROLOG Obstetrics: Critique book (Acog, 2013).
- PROLOG Gynecology and surgery: Question Book (Acog, 2009).
- 100 cases in obstetrics and gynaecology (Bottomley; Rymer, 2014).

In addition, we studied four situations (user story[34]) to understand the obstetric and neonatal context. Another investigation at the Maternity Hospital of UFMG[35], conducted by other researchers in parallel with this one, defined such stories. These stories aimed to identify the minimum set of data on prenatal care in support of communication between the primary care network and motherhood. We present below the one of these four situations studied.

History 1

A pregnant woman arrives at the maternity hospital with complaints of uterine contractions. She reports that she started prenatal care as soon as she got pregnant, had many consultations and exams, but recently lost her prenatal card. She denies any diseases, has no signs of clinical abnormalities and will be admitted for childbirth care. In your opinion, what are the most important (prenatal) data that would support you to rate gestational risk and give this woman appropriate care for her immediate needs?

[32] Title in Portuguese: Caderneta da Gestante.

[33] Title in Portuguese: Pré-natal e Puerpério: atenção qualificada e humanizada.

[34] We thank both to Dr. Zilma Reis and Thabata Sá for giving us the user story.

[35] UFMG (Federal University of Minas Gerais)

We performed interviews and observations aiming to identify straightly with health professionals and their work environment any relevant knowledge on the obstetric and neonatal domain in order to subsidize the ontology development. For this, the researcher visited the Maternity Hospital (lying-in hospital) of the UFMG's University Hospital in Minas Gerais State in Brazil in the first semester of 2015. The researcher also visited the Women & Children's Hospital of Buffalo and the UBMD[36] Obstetrics-Gynecology clinics, both in New York State in the United States, in the second semester of 2015 and first semester of 2016. During such visits at UBMD clinic, the researcher shadowing Obstetrics-Gynecology physicians during medical appointments with pregnant women and non-pregnant women.

These visits/shadowing helped to identify information on gynaecological, obstetric and neonatal infrastructure and to observe how physicians and nurses performing their healthcare during pre-pregnancy, prenatal, and postnatal. In addition, the researcher conducted direct unstructured interrogation to physicians requesting for information on the problem of interoperability lack, why the data exchange is so important, what data are needed to exchange between systems, besides collecting relevant information on the obstetric and neonatal knowledge.

3.3.2 Determine the methodology to build the target ontology

The second step of exploratory phase was ***determine the methodology to build the target ontology***. Before beginning the construction of the ontology, the ontologist should decide what the most suitable methodology to build his ontology. However, there are several methods and methodologies to build ontologies as described in subsection 2.4. In order to define a suitable methodology to build the ontology of the obstetric and neonatal domain, we performed a systematic literature review. Then, from the results of systematic literature, the researcher defined the methodology to build the target ontology.

A systematic literature review is documented method that aims to identify the data source to the literature on a particular subject of interest. The systematic literature review also evaluates and interprets all available research pertaining to a research question or subjects related to the proposed solution. Systematic review (SR) result is a mapping of the previous knowledge existing and relevant initiatives about the research topic (Jones, 2004, Kitchenham, 2009, Wright *et al.*, 2007). There is a recommendation to create a protocol with systematic review planning (Biolchini *et al.*, 2007, Mian *et al.*, 2005). Thus, the protocol used by the researcher is available in Appendix 2.

[36] UBMD (University at Buffalo Medical Doctor) - http://ubmd.com/practice-locations/practices/obgyn.html

Considering the results of this systematic literature review we noticed that both the methodology of ontological realism and NeOn methodology presenting suitable features to build an ontology. Each one covers specifics necessities in building ontologies. So that, we define our methodology combining the best features of each one of these both methodologies as described in subsection 4.2.

3.3.3 Build the ontology of the obstetric and neonatal domain

The first step of the empirical phase was ***building the ontology of the obstetric and neonatal domain***. For this, we use the knowledge acquired in the previous step and the ontology building methodology ReBORM that we described in subsection 4.2.

3.3.4 Verify the OBO Foundry adherence of the ontology built

After the ontology building, the next step was ***verify the adherence of the ontology to the principles of OBO Foundry***. For this, we just read the list of OBO Foundry principles (subsection 2.4.2) and manually checked in the built ontology an evidence of this adherence. In the cases that the ontology was not adequate according to some principle, we made the necessary adjustments in the ontology to conform to this principle. However, during the ontology construction step, the ontologist was dedicated to meeting each of the principles. Therefore, this step was just a check to ensure compliance with all OBO Foundry principles.

3.3.5 Validate the ontology against interoperability possibility

The last step was ***validate the possibility of ontology interoperability***. For this, we created a proof of concept (PoC) demonstrating the feasibility of using a formal ontology as a reference model to query data originated on relational databases. This proof of concept, PoC, is an experiment that first establishes a mapping between a hypothetical database and a draft of our ontology, second export data of each database table line to ontology instances creating an RDF triple file, and finally, it performed some queries using language SPARQL in the RDF triple file.

The environment of this PoC consists in two parts. The first part is the creation of a hypothetical relational database schema using a local MySQL[37] DBMS (database management system). Notice that every time on this thesis that it mentioned the terms table,

---

[37] Details available in https://www.mysql.com/

column, primary key, foreign key, column value, table lines, it refers to database elements. Some of these elements are structural database elements, that is, they are database metadata. Examples of database metadata include the set of tables, table columns, primary keys, foreign keys, table and column names, columns datatype and their respective size. In addition, to prepare the second part of the PoC environment, it detached a part out of the ontology build including the set of terms that corresponds to elements of the database schema. In this PoC, both the database and the ontology were created on the same local machine. Nevertheless, both could be located in different machines and may have been created at different times and by different teams and organizations.

Once the PoC environment was created, we performed two mapping. The first mapping creates a correspondence between database metadata and ontology elements. The second mapping aims to identify ontology instances from data existing in the database tables.

First mapping: relating database metadata to ontology elements

We used some recommendations of the direct mapping of relational database (RDB) to RDF guideline created by W3C consortium available at W3c (2012a). The idea of the W3C direct mapping is to specify an RDF Graph from the data existing in an RDB. However, our mapping process not intents to create an ontology from a database or create a database from an ontology. Our goal is identifying the correspondences between the database metadata and the universals of ontology with their respective ontology properties.

An RDF graph is a set triples pattern defined by subject-predicate-object. In which the nodes of an RDF graph are subjects or objects of triples, and the edges of an RDF graph connecting two nodes are predicates (property) of the triple. A subject is an IRI or a blank node, a predicate is an IRI, and an object is an IRI, a literal or a blank node (W3c, 2004c, 2014a).

From this point on, we will just call mapping the process of defining the correspondence between database metadata and an ontology element. During the first mapping, we performed a mapping to correspond each database metadata elements and ontology elements, and then record the result, the correspondence defined, in an output file as a subject-predicate-object triple. Figure 33 schematizes this first mapping performed into four stages:

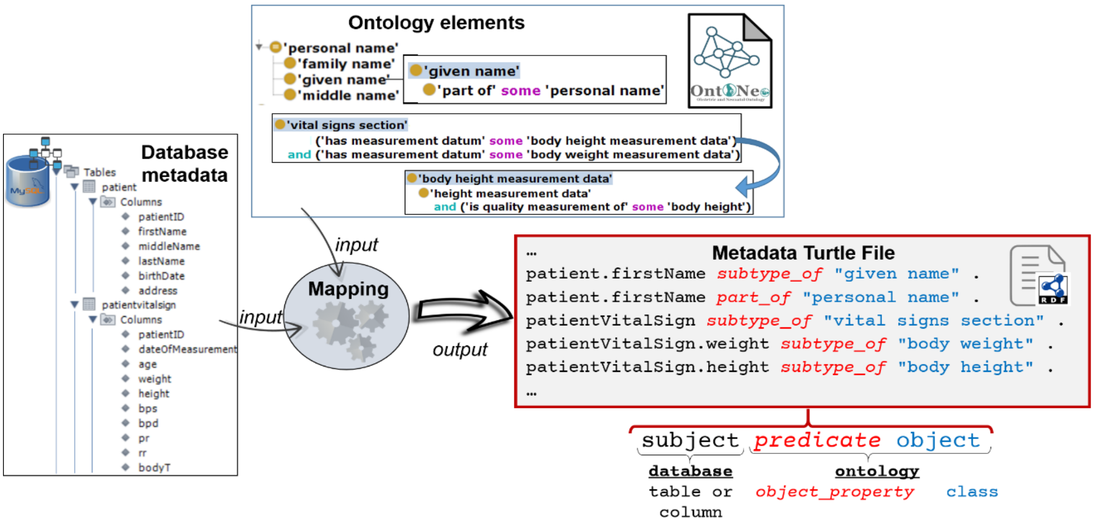
1st. Identify database metadata elements: tables, tables’ columns, primary keys, foreign keys.

2nd. Identify ontology elements: classes and properties.

3rd. Perform a mapping to correspond each database metadata elements to one ontology element, following the set of rules showed on Table 11.

4th. Generate an RDF turtle file reflecting the mapping performed in the previous step.

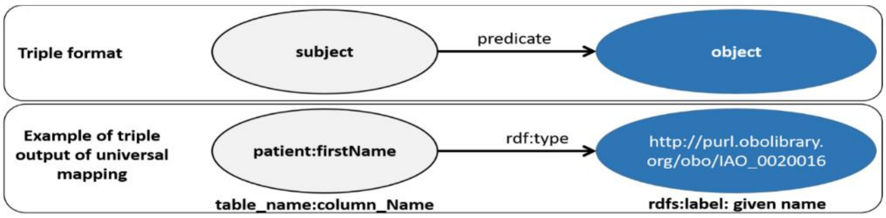
Figure 33: Schema of mapping between database metadata to ontology elements.



Source: Prepared by the author.

The generation of the output in turtle format is organized in RDF triples, namely, subject-predicate-object triple (Figure 34).

Figure 34: Example of RDF turtle file graph



Source: Prepared by the author.

Every triple component is specified via an IRI, however, in order to foster a better understanding of this thesis, the triple component here is denoted by a label instead it IRI. Each component of the resulting RFD triple corresponds to constructs of the inputs (ontology and database) as follows:

- The value "*subject*" of the triple corresponds to database metadata, which could be tables or table's columns.
- The value "*predicate*" of the triple corresponds to the relations between subjects and objects, which come from object properties of the ontology.
- The value "*object*" of the triple corresponds to classes, which come from classes of the ontology or classes that corresponds to some subject already mapped.

Table 9 presents the correspondence mapping between database metadata and ontology elements.

Table 9 – Rules of mapping between database metadata and ontology elements

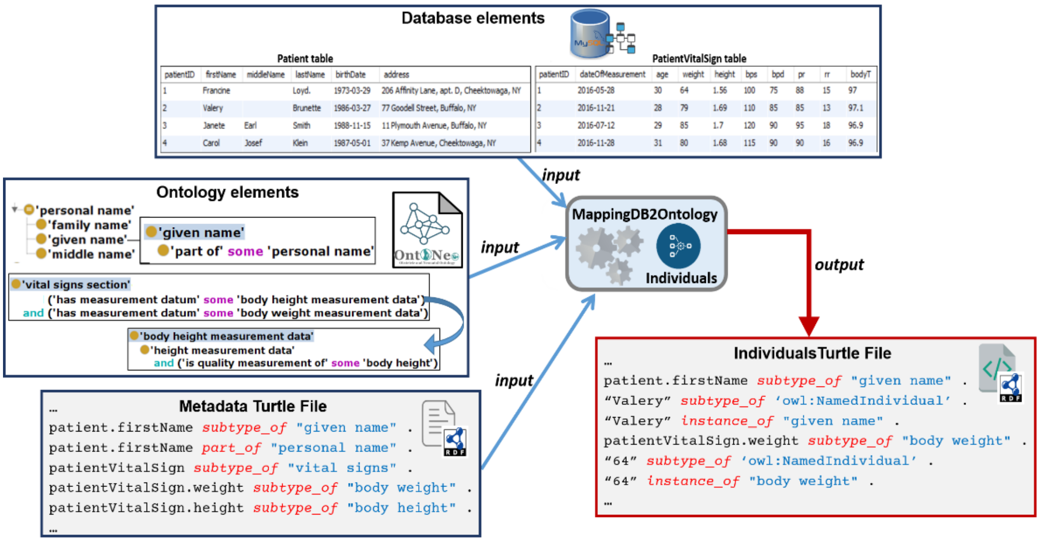
Database metadata		Ontology element
Case 1	Table	Class
Case 2	Table	Classes and relationships between these classes expressed by object properties.
Case 3	Column table	Class
Case 4	Simples PK (1 column)	Class
Case 5	Complex PK (> 1 column)	Classes and relationships between these classes expressed by object properties
Case 6	Foreign Key	Instances of relation between a parent class and a child class, expressed by object properties

Source: Prepared by the author.

Second mapping: defining ontology individuals from database data

After the first mapping, we focused on creating another triple file to be used ahead for queries. Figure 35 presents a general view of the process carried out.

Figure 35: Mapping individuals from database to triple



Source: Prepared by the author.

During this second mapping, we make the correspondence between each column value of each table row to an instance of an ontology class. The first mapping identified the correspondent class for each column, but now the value of column becomes an instance of that previously identified class. The input is the triple file (generated in the first mapping), both database and ontology elements; the output is another RDF triple file that now focuses on individuals. Table 10 presents the correspondence mapping between database elements and ontology elements.

Table 10 – Rules of mapping between database elements and ontology elements

Database metadata		Ontology element
Case 7	Line	To an instance of classes or object properties. The existence of a line indicates an instance of the target mapped.
Case 8	Column Value	*instance_of* some class

Source: Prepared by the author.

Thru this mapping, we defined an IRI to each individual found in the database. These individuals compose the "subject" of the triple. The "predicate" of the triple is composed of relations defined by classes of the ontology mapping. Finally, the "object" of the triple indicates that the subject is an element of an individual type and the class in which this individual figures as an instance type.

Performing queries

Once the data from the database was mapped to a formal ontology, the unified format can be queried using SPARQL. In this proof of concept, we use Twinkle 2.0 software to query our file of instances.

4 RESULTS

"I cannot teach anybody anything. I can only make them think".
Socrates

This chapter describes the results of the methodological steps performed, presenting evidence of the accomplishment to each methodological step. The chapter consists of four sections. Section 4.1 summarizes the obstetric and neonatal domain in which is a result of the first methodological step. Section 4.3 describes the development process of the ontology of the obstetric and neonatal domain that uses the ReBORM methodology (described in subsection 4.2), encompassing the second and third methodological steps. Section 4.4 describes how OntONeo development deals OBO Foundry principles, which refers to the fourth methodological step. Finally, section 4.5 presents a proof of concept using OntONeo to retrieval data in a relational database, which refers to the fifth methodological step.

4.1 The obstetric and neonatal domain

A motivator of this study is the Stork Network Program (*in Portuguese is* "*Rede Cegonha"*), proposed by the Brazilian Health Unified System, which aim ensure the continuity of care for women and newborns by guarantees a woman's entitlement to reproductive planning, pregnancy, childbirth, and postpartum care (Brasil, 2011a). Thus, all information related to woman's care in pre-pregnancy, prenatal, parturition, puerperal (postnatal) and newborn phases are part of the obstetric and neonatal domain.

Since EHRs are tools used to store this information (Galvão; Ricarte, 2012, Massad; Marin; Neto, 2003), all EHRs used into these phases constitute the knowledge focus inside the obstetric and neonatal domain to OntONeo. Therefore, the knowledge of the obstetric and neonatal domain covers the information generated in the both woman and newborn healthcare.

Despite the higher women life expectancy compares to men, across the lifespan, women undergo more unhealthy days than men, suffering physical and mental disturbances (Hhs; Hrsa, 2011). During their lives, women face different health care needs compared to men. There is no doubt of the biological differences encountered in the body of men and women, however, these differences lead to healthcare specificities not only in relation to the reproductive system but also in terms of anatomy, physiology, and psychological factors.

Woman healthcare involves a series of different care during her life. In Brazil, in the 1980s, the Program for Integral Assistance to Women's Health (PAISM) established guiding principles of women's health policy including educational, preventive, diagnostic, and treatment actions, including women's care in gynecological clinics, prenatal, postpartum, puerperium, family planning, sexually transmitted diseases (STD), cervical and breast cancer. Later in the late 1990s and early 2000s, women's health policy included attention to climacteric, infertility care, assisted reproduction, adolescent women's health, chronic-degenerative diseases, and infectious diseases, besides other needs identified from the population profile of women (Brasil, M. D. S. D., 2016, Brasil; Saúde; Estratégicas, 2004).

In short, women's healthcare involves reproductive health and non-reproductive health. Reproductive health involves procedures related to maternal health such as pregnancy, prenatal follow-up, complications of pregnancy, childbirth, postpartum care, congenital diseases, embryo/fetus development, woman physiology changes, and so on. Non-reproductive health involves gynaecological follow-up, surgical history, immunizations, information about allergies, diseases, treatments, and so on.

The woman's care requires mainly the medical specialties gynecology and obstetrics. Gynecology is the medical specialty dedicates to the study of the anatomy, physiology, and pathologies of the woman's body and her genital system. Obstetrics is the medical specialty that studies issues related to human reproduction with a focus on women. Obstetrics investigates the physiological and pathological aspects of gestation, childbirth and postpartum or puerperium. Almost all gynaecologists today are also obstetricians.

Gynecology or gynaecology is "the science of women," it is concerned with the well-being of women, the health of their female reproductive organs and their ability to reproduce. The gynaecologist is the specialist who takes care of the health of the female reproductive system from infancy to the postmenopausal period. Generally, after the first menstruation (menarche), most women begin their follow-up with a gynaecologist seeking behaviour that improves their intimate health. The gynaecologist is still prepared to address issues of endocrinology, female urology, pelvic malignancy, osteoporosis and other female health issues.

The American College of Obstetricians and Gynecologists (ACOG) recommends a woman to have at least one gynaecological annual examination. This annual health assessment, or annual examination, serves as preventive care and helps early health problems detection. The assessments performed in each encounter vary according to age group. ACOG also defined four age range groups to fit its recommendations (Acog, 2017):

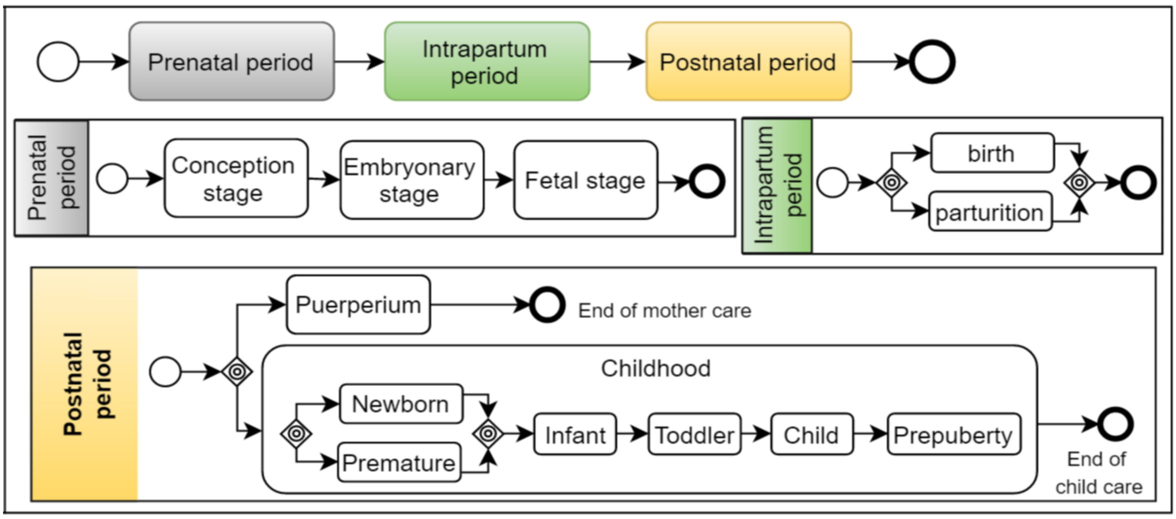
- Group 1 - women with ages between 13 and 18 years old.
- Group 2 - women with ages between 19 and 39 years old.
- Group 3 - women with ages between 40 and 64 years old.
- Group 4 - women with ages from 65 years old and older.

For each one of these groups, Acog (2012, 2017) recommends a specific set of procedures such as screening, examinations, laboratory and other tests, evaluation and counselling, and immunizations. Therefore, woman anatomy and physiology and woman development are knowledge required by those involved in gynaecological care. Once the basic facts of woman anatomy and physiology change during puberty, post-menarche, and pre-menopause periods (Beckmann *et al.*, 2010b).

Gynecological care comprises evaluation of the woman's health during her life requiring, for example, pelvic examination and breast evaluation, menstrual and contraceptive history, infertility and fertility issues, sexual behavior, and reproductive planning (Acog, 2010a, 2010b, Brasil, 2006, Brasil; Saúde; Estratégicas, 2004). For details, take a look at these references: (Acog, 2012, 2017).

Obstetrics aims to guarantee the normality of pregnancy and childbirth besides focusing on the health and quality of life of the woman during the course of pregnancy, parturition, and puerperium. The course of pregnancy, childbirth, and child development involves a series of stages as illustrated in Figure 36 and Figure 37. The pre- and postnatal periods are the periods before and after birth (*parturition*). The intrapartum period comprises the period from the beginning of parturition to the completion of baby and placenta delivery. Moreover, clinical care in each stage requires different medical specialties for example obstetrics, paediatrics. The course of pregnancy urges different lab tests and imaging procedures, as well as immunizations, screening and other healthcare-related processes (Beckmann *et al.*, 2010a, Cunningham *et al.*, 2014, Hoffman *et al.*, 2012, Mcinerny *et al.*, 2009).

Figure 36: Stages of pregnancy and child development



Source: Prepared by the author.

Similarly, monitoring the child development requires physical examinations, screening, immunizations, and so on (Aap, 2011, Beckmann *et al.*, 2010a, Cunningham *et al.*, 2014, Hay *et al.*, 2014, Hoffman *et al.*, 2012, Mcinerny *et al.*, 2009).

Figure 37: Timeline of pregnancy and child development

Fertilization and prenatal period			
		Gestational Age	
Ovulation		Week 1-2	
Conception	Fertilization	Week 3	Day 1
	Cleavage		Day 2
	Compactation		Day 3
	Differentiation		Day 4
	Cavitation		Day 5
	Implantation		Day 6
Embryonary period	Gastrulation	Week 4	
	Somitogenesis	Week 5	
	Organogenesis	Week 6	
	Histogenesis	Week 7	
	Neurulation	Week 8	
Fetal period		Week 9-40	
Normal Gestation Birth			

Puerperium or postnatal period (Mother)	
Immediate puerperium	First 24 hours after labor
Early puerperium	First week after labor
Remote puerperium	Around 6 weeks after labor

Childhood (Child)		
Premature	Baby born under 36 weeks of gestation	
Newborn	1st month after birth	
Infant	2nd to 12th month	
Toddler	1st month of 2nd year to 3 years	
Child	4th year to Prepuberty period (around 10 years)	
Adolescent	Prepuberty (girls)	Around 11 years
	Prepuberty (boys)	Around 12 years
	Puberty (girls)	11th to 15th/17th years
	Puberty (boys)	12th to 16th/17th years

Source: Prepared by the author.

Because anomalies and congenital diseases may arise during development the embryo, physicians that monitor prenatal stages are required to know embryology and embryogenesis (Cunningham *et al.*, 2013, Larsen, 2001). During obstetric care, a doctor performs a physical examination of the woman to detect any changes in her body that might indicate an underlying disease. Additionally, the examination of pregnant mother aims to detect embryo or fetal complications, and screening exams help to detect developmental anomalies. For example, a prenatal ultrasound exam provides images of the baby, amniotic sac, placenta, and ovaries. The results enable tracking of anatomical abnormalities or birth defects of the baby. Thus, knowledge of human anatomy and physiology are also foundations of obstetric care (Beckmann *et al.*, 2010b).

Summarizing, the obstetric and neonatal domain involves data gathered from several medical specialties and knowledge domains such as embryology, anatomy, gynecology, obstetric, and paediatrics.

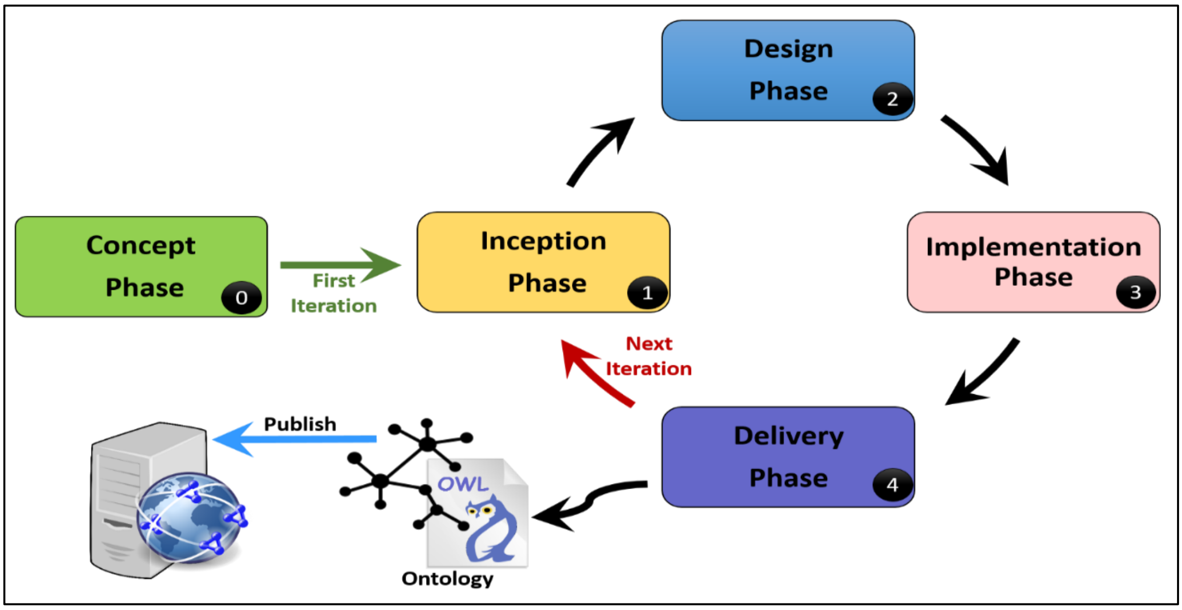
4.2 ReBORM: Combining methodologies for building ontology

In this section, we describe the methodology proposed and used in this thesis to build our target ontology. This methodology is a new methodology reborn from the combination between ontological realism methodology (subsection 2.4.5) and NeOn methodology - an ontology engineering-based methodology (subsection 2.4.6). We have named this methodology as ***Realism-Based Ontology engineeRing Methodology***, using an acronym **ReBORM** in an attempt to play with words, thus, the acronym ReBORM makes an analogy to the word *reborn*.

ReBORM adopted the principals of the methodology of ontological realism, distributed along of ReBORM phases, to foster semantic coherence, to both humans and computers. Complementing, ReBORM adopted the NeOn methodology to take advantages of the ontology engineering best practices. One advantage of using Neon Methodology is its incremental character, which is very suitable for the development of ontologies following the premise of fallibilism. Fallibilism reflects the evolutionary disposition of science and is part of the framework of ontological realism (Munn & Smith, 2008). Additionally, ReBORM relies on well-consolidated initiatives, namely those pertaining to the OBO Foundry (subsection 2.4.2), and ODPs[38] (subsection 2.4.1).

From a practical point of view, ReBORM merges the five steps of the methodology of ontological realism (Figure 21, subsection 2.4.5) with the scenarios and activities (Figure 22 and Figure 23, subsection 2.4.6) of the NeOn methodology. ReBORM follows a development life cycle iterative-incremental, as suggested by the ontological realism and foreseen by NeOn (Figure 26, subsection 2.4.6). In this context, inspiring by the scenario 1 of NeOn, ReBORM workflow contains five main phases, as shown in Figure 38.

Figure 38: Development life cycle workflow of ReBORM.



Source: Prepared by the author.

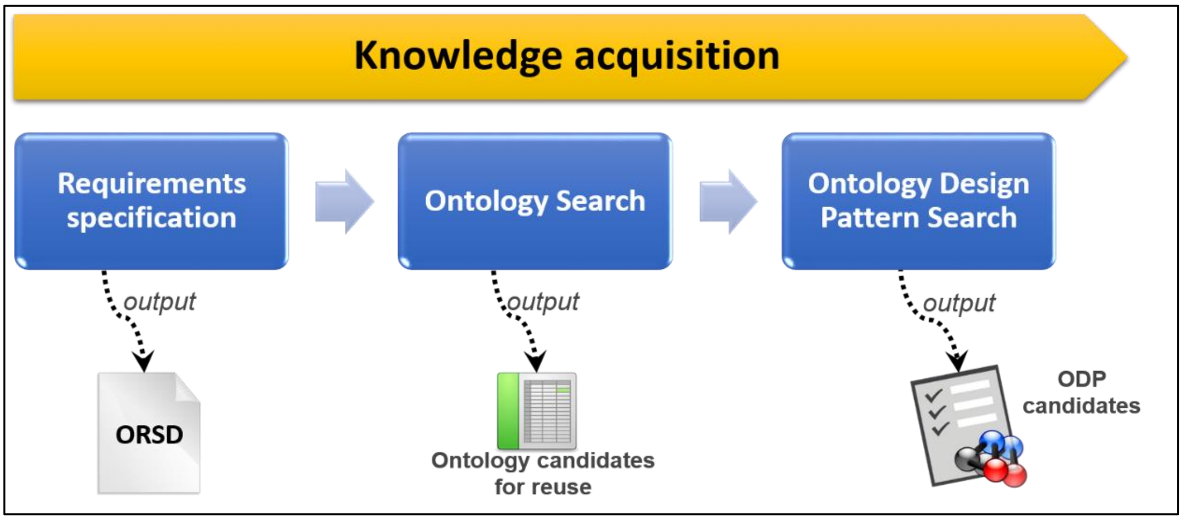
[38] ODP - Ontology Design Patterns

The phases are: 0) *concept*; 1) *inception*; 2) *design*; 3) *implementation*; and 4) *delivery*. The *concept phase*, number 0 (zero), was the beginning of the ontology development. This phase drove all others four phases and happened just one time while the phases numbered from 1 to 4 happened once in each iteration performed. The phases 1 to 4 constitute the iterative-incremental part of ReBORM development life cycle.

4.2.1 Concept phase

The first phase of ReBORM is the ***concept phase*** in which aim to identify the purpose of the ontology, the set of requirements that it should satisfy, and the competency questions used to validate the ontology. In this phase, the ontologist explored and elaborated the main idea for the whole ontology project. In addition, it collects material and acquires knowledge about the domain. Thereby, this phase embedded two principles of the methodology of the ontological realism, namely, "*demarcate the scope and domain of knowledge*" and "*gather information of the ontology domain*". In addition, according to NeOn, any ontology development starts with *ontology requirements specification* activity, which counts on the support of the activity of *knowledge acquisition*. Based on these both references of methodology, this phase consisted of the following activities: *Knowledge acquisition, Requirements specification, Ontology Search,* and *Ontology Design Pattern Search* (Figure 39).

Figure 39: Concept phase activities of ReBORM.



Source: Prepared by the author.

<u>*Knowledge acquisition*</u>

Knowledge Acquisition activity consists of gathering knowledge from several information sources. The *knowledge acquisition activ*ity was carried out during the whole ontology development (Suárez-Figueroa, 2010b). The step "*gather information of the ontology domain*" of the methodology of ontological realism suggests for ontologists acquiring knowledge from to textbooks, documents, reports of the target domain, besides consultation to domain experts (Arp; Smith; Spear, 2015).

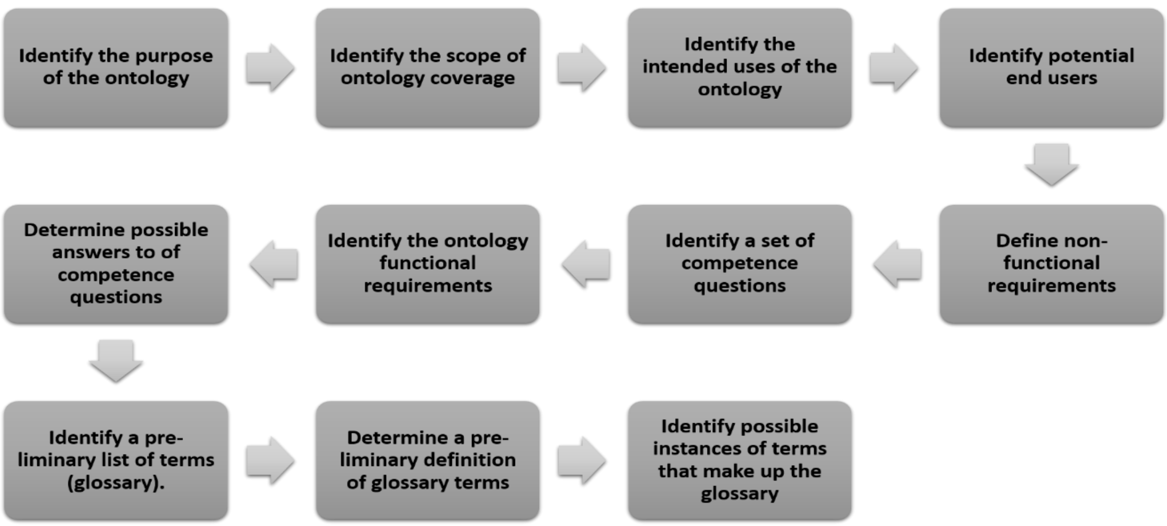
Requirements specification

Ontology *Requirements Specification* focuses on establishing the set of requirements that the ontology should meet. Such activity is equivalent to the step "Demarcate the scope and domain of knowledge" of ontological realism. The objective of *requirements specification activity* is to define why the ontology is being built, what its intended uses are, who the end-users are, and what the requirements the ontology should meet (Suárez-Figueroa, 2010b).

In order to register the set of requirements, ReBORM use the same template of ontology requirements specification document that NeON methodology, in which was presented in Figure 24 (subsection 2.4.6). We notice that the ontology requirements specification document items of NeON methodology template, are also equivalent to the questions proposed on the "*Demarcate the scope and domain of knowledge*" step of the methodology of ontological realism.

This activity should carried out according to the guidelines shown in (Suárez-Figueroa, 2010b) and (Suárez-Figueroa; Gómez-Pérez; Villazón-Terrazas, 2009). To perform this step, we defined some task organized according to the following task flow (Figure 40).

Figure 40: Task flow of the Ontology Requirements Specification activity.



Source: Prepared by the author.

After the ontology requirements specification activity, it is advisory to carry out a quick search for knowledge resources to, using as input terms included in the ontology requirements specification document (Suárez-Figueroa, 2010b).

Ontology search

Ontology Search activity is the activity in which ontologists should look for ontologies already existing that meet the requirements or part of requirements of the developing ontology (Suárez-Figueroa, 2010b). This activity meets the recommendation of the step "demarcate the scope and domain of knowledge" from the methodology of ontological realism that highlighted the importance of the identification of ontological resources and artefacts against the ontology requirements to compose a list of candidates for reuse.

There are several ontology search engines and ontology repositories to help perform this activity, such as:

- Ontobee: A web-based ontology data server and browser developed to link OBO Foundry library ontologies (Ong et al., 2017, Ontobee, 2017, Xiang et al., 2011, 2013).
- Bioportal: A web portal to a virtual library of biomedical ontologies and terminologies, which is developed and maintained by the National Center for Biomedical Ontology (NCBO). (Biopportal, 2017, Musen *et al.*, 2009, Noy *et al.*, 2008, Noy *et al.*, 2009, Rubin *et al.*, 2008, Salvadores *et al.*, 2013, Whetzel *et al.*, 2011).
- Ontology Lookup Service (OLS): A web-based repository for biomedical ontologies, which is developed and maintained by the Samples, Phenotypes and Ontologies Team (SPOT) of the European Bioinformatics Institute (EMBL-EBI) (Côté *et al.*, 2010, Côté *et al.*, 2006, Côté *et al.*, 2008, Ols, 2017).
- Swoogle Semantic Web Search Engine: An indexing and retrieval system for knowledge on the semantic web encoded in RDF and OWL languages (Ding *et al.*, 2004, Ding *et al.*, 2005, Finin *et al.*, 2005, Swoogle, 2017).
- Watson RDF search engine: A gateway for the Semantic Web, that is, a web-based search engine for searching ontologies and semantic documents. Additionally, Watson system explores the content of semantic documents (D'aquin; Baldassarre*; et al.*, 2007, D'aquin; Motta, 2011, D'aquin; Sabou*; et al.*, 2007, Watson, 2017).

- vocab.cc - RDF vocabulary search and lookup: A web portal allows users to search for linked data vocabularies and information about already existing ontologies (Stadtmüller; Harth; Grobelnik, 2013, Vocab.Cc, 2017).

Considering the existence of multiples ontology search engines and ontology repositories, a recommendation here is to check each ontology search engines analyzing ontologies features as the kind of ontology indexed, domain of coverage, formality of ontologies, top-level ontologies accepted, and so on, in order to decide in which ontology search engines and/or ontology repositories you will perform your search.

Design pattern search

Design Pattern Search consists of look for an available ontology design pattern useful to the ontology developing ontology. To perform this activity, the ontologist should check the official Catalogue of Ontology Design Pattern[39] looking for candidates to reuse on the ontology project.

To conclude the concept phase explanation, we notice that this phase helps to identify the general scope and all items that the developing ontology should covers. However, just in next phase, the ontologist will detail these items. Indeed, some ontologies project presents a big scope. Thus, we recommended dividing all the items on ontology scope in multiples iterations.

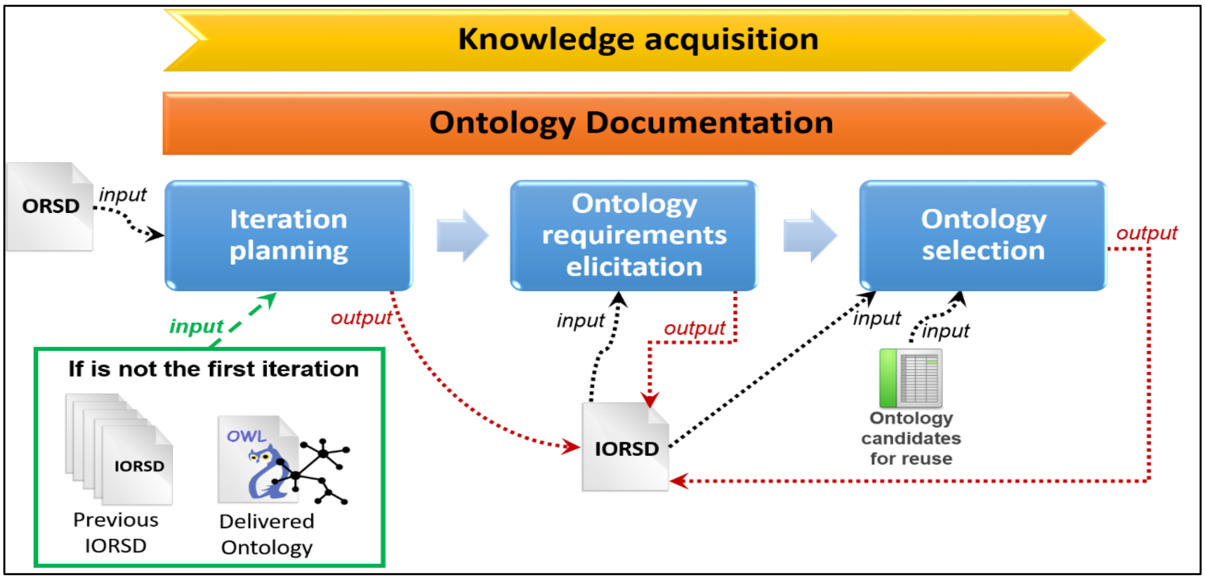
4.2.2 Inception phase

The second phase of ReBORM is the ***inception phase*** in which aim to establish the piece of the general ontology scope and requirements that the iteration will generate. Additionally, given the overall planning of the ontology project created in the concept phase, this phase defines the planning of the iteration.

The activity flow of inception phase consisted of the following activities: *Knowledge acquisition, Ontology documentation, Iteration planning, Ontology requirements elicitation,* and *Ontology selection* (Figure 41).

[39] Available at: http://ontologydesignpatterns.org/wiki/Community:ListPatterns

Figure 41: Inception phase activities of ReBORM.



Source: Prepared by the author.

Ontology documentation

Ontology documentation activity is a support activity that generates any valuable documentation to understand the ontology itself and decisions made during the ontology development. As the knowledge acquisition activity, the ontology documentation activity happens during the whole ontology development.

There is no documentation pattern to follow in this activity. The documentation created depends on each iteration. Ideally, the recommendation is to create a single repository to store the documentation avoiding both losses and replication of relevant information.

Considering the recommendations of OBO Foundry consortium, we suggest three documentation patterns: i) Annotate the ontology using annotations properties. ii) Create a project homepage. And iii) Create a development repository.

Iteration planning

At the beginning of each ontology development iteration, first, the ontologist team should define an iteration plan. This activity is the iteration planning in which the team of ontologists prioritizes the items that they will work during the iteration that begins.

Here, we follow a recommendation of ontological realism methodology of the idea of low-hanging fruit, which means, the ontology development should start with the most common entities and their relations, and move forward to the most complex entities.

Therefore, for each iteration, the ontologist defines the iteration scope using as input the requirements in the Ontology Requirements Specification Document (ORSD) (created in the conceptual phase) and, the implementations made in the previous interactions, if the iteration planning is not of the first iteration. In the case of the first iteration planning, the input is just the ORSD[40].

We register the iteration scope resulting from the planning in an Ontology Iteration Requirements Specification Document (I-OSRD), which Figure 42 present the I-OSRD template created from OSRD template presented on Figure 24 (subsection 2.4.6). Some sections of the I-OSRD[41] will be registered in post-iteration planning activities.

Figure 42: Template for the I-OSRD used by ReBORM.

Ontology Iteration Requirements Specification Document Template	
Iteration nº: *number of iteration - Title of iteration*	
1	**Purpose of iteration**
Inform the general goal of the iteration.	
2	**Scope of iteration**
Inform the coverage and the degree of detail that the piece of ontology of this iteration should have.	
3	**Ontology Iteration Requirements**
a. Non-Functional Requirements (NFR)	
Set the list of non-functional requirements that the piece of ontology to this iteration should fulfil.	
b. Functional Requirements	
Set the content specific requirements that the piece of ontology should fulfil, in the form of sentences according the sentences below. The ontology should represent knowledge of … The ontology should covers data of …	
c. Competency Questions	
A list of competency questions that the piece of ontology should aswer.	
4	**Ontology Iteration Requirements Elicitation**
The list of functional requirements with some elicitation. In others words, to each functional requirements you should presents details of to help the ontologist to define all terms and relations agaisnt this requirements. You could present a list of documents, textbooks (whole or sections), images, user stories, etc.	
5	**Pre-Glossary of Terms**
a. List of general terms	
A preliminary list of terms identified during the requirements elicitation, which the iteration should cover.	
b. Objects or Instances Examples	
Set a list of instances examples to the terms on section 5a.	
6	**Ontologies and terminologies candidates to reuse**
Inform a list of existing ontologies and terminologies candidates to be reused during the iteration.	
7	**Ontology Design Patterns candidates to reuse**
Inform a list of existing ODPs candidates to be reused during the iteration.	
8	**Literature review**
Relevant documents to be used or that serve as references in all definitions.	

Source: Author's personal collection, created from (Suárez-Figueroa, 2010b, p. 115).

[40] ORSD is an acronym to "Ontology Requirements Specification Document".

[41] I-OSRD is an acronym to "Ontology Iteration Requirements Specification Document".

Ontology elicitation

Once defined the iteration purpose, we should elicit its requirements. The elicitation activity of ontology requirements is the process of extracting knowledge from the available knowledge sources. The purpose of eliciting requirements is to gather information that is relevant to the understanding of the knowledge domain of the ontology. The researcher understands that the activity of elicitation is a knowledge acquisition activity, however with the focus on a specific subject and with the intention of detailing the knowledge about such subject.

Thus, to perform ontology elicitation activity, we suggests extending the techniques knowledge acquisition such as brainstorming, JAD (Joint Application Design), interviews with domain experts, analysis of documents, analysis of user stories, bibliographic and documentary research. The results should be documented on section 4 of the ontology iteration requirements specification document (Figure 42).

Following the recommendation of the step "g*ather information of the ontological domain*" of the methodology of ontological realism, during this activity, the ontologist also should identified a list with general terms on the scope of the iteration and register on section 5a of the ontology iteration requirements specification document (Figure 42).

Ontology selection

During the *Ontology Selection* activity, the ontologist should determine the reuse candidate ontologies most suited to the ontology iteration scope, following the recommendation of the step "*gather information of the ontological domain*" of the methodology of ontological realism. Thus, for each iteration, the ontologist selected suitable ontologies and terminologies from the list generated in *Ontology Search* activity.

During the *Ontology Search* activity, the ontologist just search for ontologies that are candidate for reuse without employment of any selection criteria. The selection criteria should be applied in this activity. However, the selection criteria for reuse of an ontology can vary according to the ontology project, considering for example elements such as the knowledge domain covered by the ontology, the language that the ontology is made available, the level of formalism of the ontology, the high level ontology in which the candidate ontology was built, etc. In this way, the definition of the selection criteria for reuse an ontology should be established in this activity.

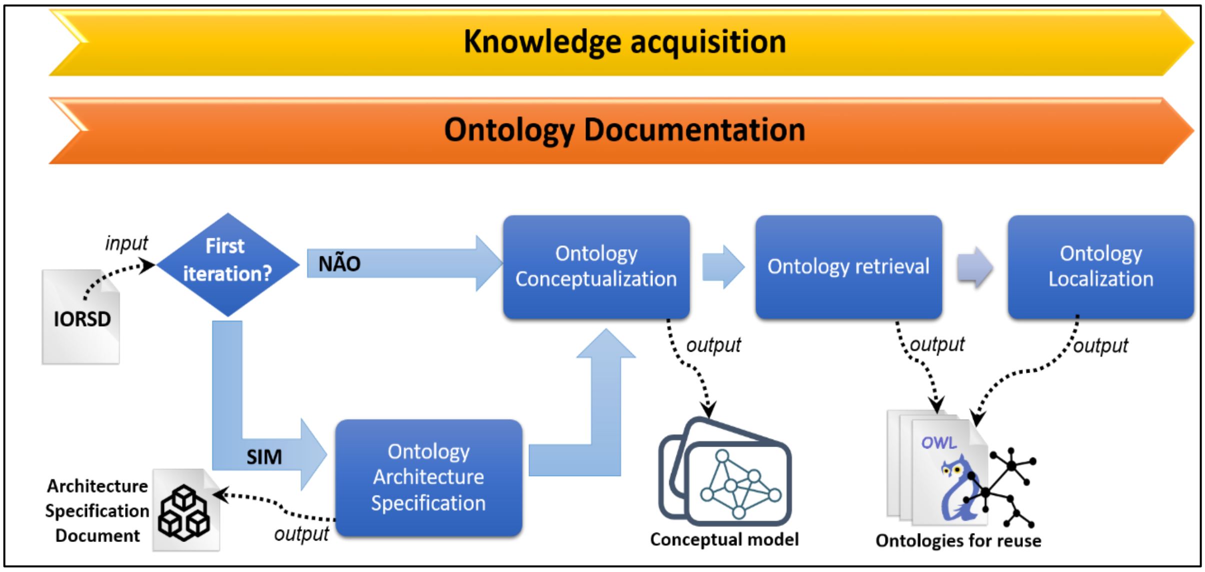
Concluding the inception phase, one could question the fact of not exist a similar activity to select the set of ontology design pattern for each iteration. Well, the ontologist understands that the list of ontology design pattern selected in the previous phase should persist during whole development. So that, the list of ODPs here is the same list defines in concept phase.

4.2.3 Design phase

The next phase, *design phase,* consists of defining detailed specifications that emphasize the implementation solution to the ontology requirements. The focus of this phase is to obtain, at the end, the definition of the architecture of the ontology, of the development environment, and of the conceptual model related to the requirements elicited in the previous phase.

In all iteration of this phase, It should be performed the follow activities *ontology conceptualization* and *ontology localization*. Moreover, just at the first iteration, it should perform the activity of *ontology architecture specification*. Figure 43 showed the activity flow of design phase.

Figure 43: Design phase activities of ReBORM.



Source: Prepared by the author.

Each activity of design phase results in an artefact that describes the development ontology. Some important artefacts generated at design phase are an ontology conceptual model possible of implementation and the files with ontological resources to reuse. Additionally, at the design phase of the first iteration should establish an ontology architecture specification that will guide the implementation phase.

<u>*Ontology architecture specification*</u>

Ontology Architecture Specification activity indicates the ontology architecture on which the ontology will be developed. We understand the ontology architecture as a set of components, their interfaces, environment, and behaviours involved in ontology development.

In general, the architectural elements constitute parts of the non-functional requirements of the ontology. Some examples of the ontology architecture elements are:

- Top-level ontology base to developing ontology
- Ontology namespace, IRI, base purl
- Project homepage
- Ontology development repository
- Development tools
- Coding language
- Default language to ontology elements and alternative languages
- Type of license of the ontology
- Necessary documentation

ReBORM defined a template to the ontology architecture specification document according to Figure 44.

Figure 44: Template for the ontology architecture specification used by ReBORM.

Ontology Architecture Specification Document		
1	**Top-level ontology to extend**	
2	**Ontology namespace**	
	PREFIX or IDSPACE	
	Ontology IRI	
	Default purl	
	Local identifier	
3	**Ontology project website**	
4	**Development repository**	
5	**Development tools**	
6	**Coding language**	
7	**Default language to ontology elements**	
8	**Alternative language to ontology elements**	
9	**License**	
10	**Minimal ontology documentation**	
11	**Additional information**	

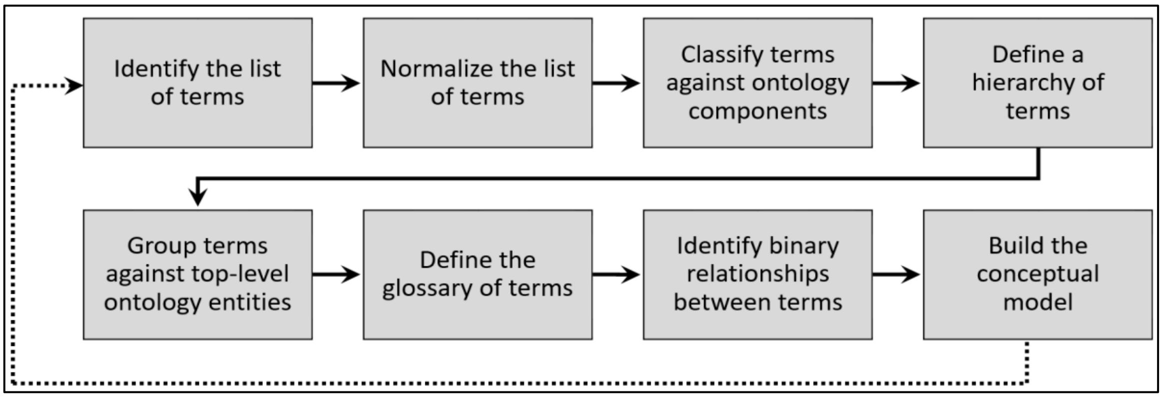
Source: Prepared by the author.

We notice that the reuse of ontology architectural patterns could help the ontologist to define his own project architecture.

Ontology conceptualization

The purpose of the conceptualization activity is organizing and structuring the information collected during the acquisition of knowledge, the specification of requirements, and ontology elicitation. During this activity, the ontologist addressed several recommendations of the ontological realism methodology from different steps. Figure 45 presents the tasks performed during each conceptualization activity of ReBORM. After figure, we explained each task.

Figure 45: Ontology conceptualization tasks of ReBORM.



Source: Prepared by the author.

We suggested a sequence to perform these tasks according to Figure 45; however, each ontologist could perform these tasks in a sequence more convenient to him. The diagram shows an idea of circularity through the dotted arrow, in which refers to low-hanging fruit principle of the ontological realism methodology. Therefore, the conceptualization should start by identifying the most commonly terms used and advance gradually from there, to identify more complex and specific terms.

During the inception phase, the ontologist defined a preliminary list of terms against the scope of iteration. However, this list is just a start point with general terms easily identifiable from the knowledge of the scope of the interaction domain. Thus, as an initial task of conceptualization, *the ontologist refined the preliminary list identifying additional terms*.

In general, the terms in the list embrace verbs, nouns, adverbs, adjectives and numerals that will be equivalent to different types of ontology components such as universals, instances, attributes, relationships. Once the terms identification happened, the ontologist should standardize these terms. Hence, *normalization of the list of terms* is the task of formatting terminology used to fit the set of rules and best practices suggested by the methodology of ontological realism.

Then, the ontologist should *classify the terms* in order to identify them as the type of ontology component presented in Table 4 (see subsection 2.3.2). Additionally, the ontologist correlated the terms classified as instances to the respective terms classified as classes or entities. Next, the ontologist should *define an is_a hierarchy* following the recommendations of the methodology of ontological realism.

Continuing the conceptualization, the ontologist should *group each term*, classified as an entity, *against the top-level ontology entities*. Next task is to *define the glossary of terms*, in other words, the ontologist created the definition of each term in the list, especially for terms classified as the entity type. The definition followed the recommendations of the methodology of ontological realism. Succeeding, the ontologist defined the binary relations between the terms, based on primitive binary relations: the class-class relation, the instance-class relation, and the instance-instance relationship.

Finally, the ontologist built the graphic conceptual model reflecting the term list. This conceptual model gave an overview of what knowledge the iteration would represent helping the validation of the ontology piece with domain experts. For this, we suggest the tool CMap Collaborative Ontology Environment (Garcia *et al.*, 2007, Hayes *et al.*, 2005, Hayes; Saavedra; Reichherzer, 2003).

As result of this task, we reach a first version of the conceptualization, with terms and relations identified and conceptually related. At each iteration, the result of the ontology conceptualization activity produced one or more conceptual models shared with the domain experts and other ontologists consulted for analysis and validation. These already validated conceptual models subsidized the activities of ontology retrieval, integration, and formalization.

<u>Ontology retrieval</u>

Ontology retrieval aims to support the ontology reuse by preparing the potentially reusable ontologies for importation during the implementation phase. To ensure ontology consistency the recommendation is to reuse, whenever possible, entities from ontologies built under the same top-level ontology.

The ontology retrieval of pre-existing ontologies could happens by two different approaches: i) retrieval of the whole ontology or ii) retrieval just some terms from an ontology. In the both approach, after retrieval the ontology, maybe, the ontologist could cut some not relevant documentation, usually registered by annotation properties, to the ontology development, for example, comments and contributor information. Notice the needs to document some relevant information presented in Table 11.

Table 11 - Information documented on ontologies reused by OntONeo.

Documentation	Annotation used to document
Retrieval date	http://www.w3.org/2000/01/rdf-schema#comment
Ontology version retrieved	http://www.w3.org/2000/01/rdf-schema#comment
Who retrieved the ontology (personal name)	http://www.w3.org/2000/01/rdf-schema#comment
Default language when such information is not documented	http://protege.stanford.edu/plugins/owl/protege#defaultLanguage

Source: Prepared by the author.

In the second approach, we follow the Minimal Information to Reference External Ontology Terms (MIREOT) guidelines. MIREOT is an initiate developed as part of the Ontology for Biomedical Investigations project that recommends the reuse just of relevant terms from others ontologies instead of the whole ontology. Therefore, according to MIREOT, an ontologist should import only necessary elements (terms, properties, and relations) from another ontology to be reused, avoiding the overhead of importing the whole ontology (Courtot *et al.*, 2011).

For this, we recommend the use of Ontofox[42], a web-based tool capable of generating an OWL output file with selected ontology elements based on users input settings. Ontofox is an acronym to "***Onto****logy tool that* ***f****etches* ***o****ntology terms and a****x****ioms*". OntoFox implements the MIREOT principle and related strategies by extracting from ontologies just the minimum information necessary according to users' inputs. (Xiang *et al.*, 2010). The OWL file containing the subset of the ontology that Ontofox generated either can be imported into the ontology under development or embedded just parts of this file on the target ontology.

Ontology localization

Ontology Localization consists of to convert an ontology from the original language and culture to another that the ontologist need. The execution of this activity depends on both the ontology requirements and the set of ontologies select as a candidate to reuse. During the requirement specification activity, the ontologist defined the language and culture of the ontology in development. According to this definition, the ontologist compared the language of the target ontology with the language of each potentially reusable ontology to identify the types of adaptations required. The necessity here is to maintain the labels and definitions of both the developing ontology and the reusable ontologies in all languages that the requirement defines.

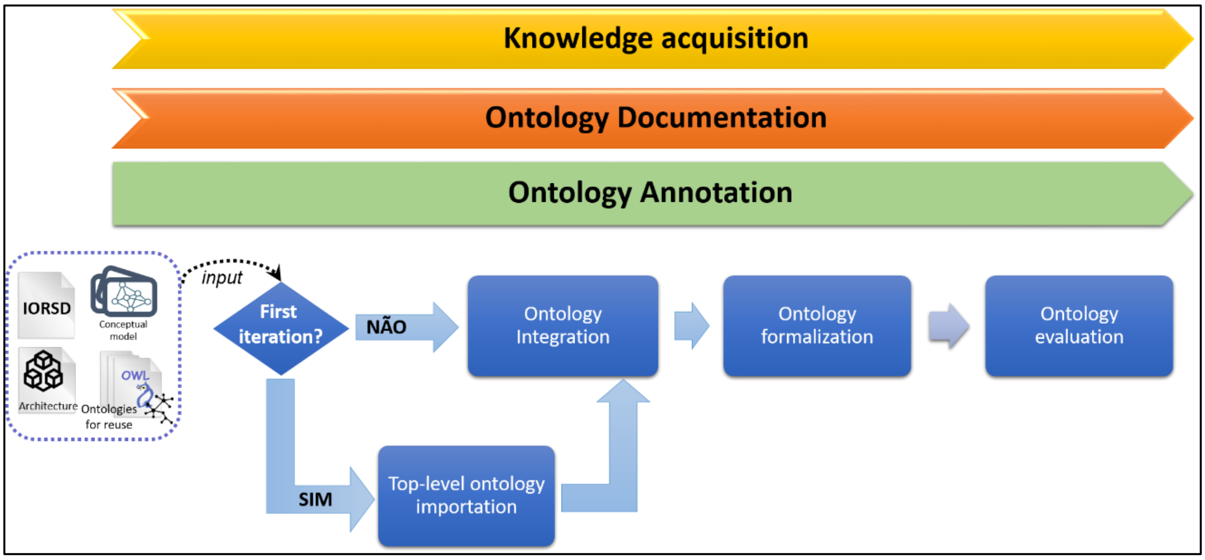
[42] OntoFox is available at: http://ontofox.hegroup.org/. Last date of access: June 13, 2017.

4.2.4 Implementation phase

The *implementation phase* aims to transform the conceptualization produced into a formal model represented by a machine-readable code. At this phase, the ontologist formalizes the conceptual model using a formal language. Notice that the conceptual model is an output of the *ontology conceptualization activity* and the definition of the formal language is an output of the requirement specification activity and which was confirmed on the activity of ontology architecture specification.

During the implementation phase, ontologits should follows the architectural issues defined at the *ontology architecture specification*. Figure 46 presented a workflow of the activities that ReBORM specify to the implementation phase.

Figure 46: Implementation phase activities of ReBORM



Source: Prepared by the author.

Top-level ontology importation

In the first execution of the implementation phase, that is, in the first iteration of development, the developer needs to import the top-level ontology that underlies the developing ontology.

Ontology integration

Ontology integration is the activity of including a pre-existing ontology (partial or total) in the ontology under construction. Then, the ontologist should import the reusable ontologies prepared during the design phase. It is important to observe the ontology interoperability approaches - alignment, mapping, combination, and integration (subsection 2.3.3) – to define a better strategy reuse such ontologies.

From the second iteration, the ontologist should be aware of the different strategies of ontology integration. In the case of the integration of a new ontology for reuse, the ontologist should import the OWL file created in the previous phase as a whole. However, in the case of the integration of an already integrated ontology on the previous iterations, in which this ontology suffered an update to the new iteration, the ontologist should consider comparison and merge tasks to reach the desired result.

Ontology annotation

The term "*annotation*" refers to a piece of data used to document or add information to some data. Annotations are metadata, i.e. data about data. In this sense, "ontology annotation" is the act of include data to other data, i.e. to add an annotation to some ontology component. Every time that an ontology component receives an annotation, it performed an instantiation of the ontology annotation activity. Therefore, the ontologist annotated the ontology during whole implementation phase according to the specific scope of annotation used. We suggested some relevant annotations to enrichment of the ontology with details about the ontology domain scope (Table 12).

Ontology formalization

Ontology formalization is the activity in which the ontologist represents the conceptual model using a formal language such OWL. This phase involves the formalization or codification of each term existing in preliminary list refined during of the inception phase. These terms could be represented through classes, relations, functions, instances and so on.

Ontology evaluation

The process of evaluation to check the technical quality of the developed ontology. We check the consistency of definitions and make use competency questions to validate whether the ontology is compatible with the planned ontology. We also analyzed the final ontology built to verify the compatibility with the requirements and competency questions. We submit the ontology developed to OOPS! evaluation in order to check the presence of pitfalls (subsection 2.4.4) in such ontology.

Table 12 – ReBORM suggestion of annotations properties.

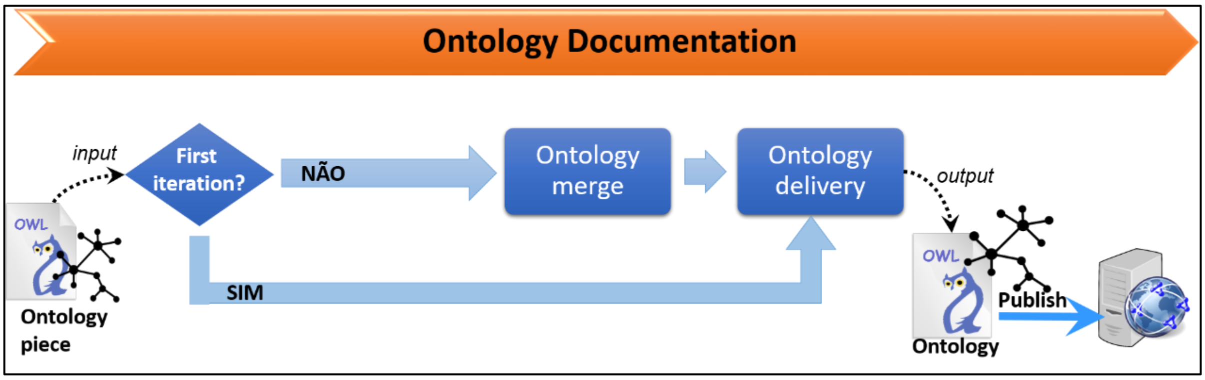
Source		Annotation used	Scope of Use
PREFIX	Namespace		
dc	http://purl.org/dc/elements/1.1/	creator	Ontology
		contributor	Ontology
		license	Ontology
		date	Any component
		source	Imported components
foaf	http://xmlns.com/foaf/0.1/	homepage	Ontology
		mbox	Ontology
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#	Description	Any component
rdfs	http://www.w3.org/2000/01/rdf-schema#	label	Ontology
		range	Properties
		domain	Properties
		comment	Ontology and entities
		isDefinedBy	Any component
owl	http://www.w3.org/2002/07/owl#	versionIRI	Ontology
		versionInfo	Ontology
		priorVersion	Ontology
obo	http://purl.obolibrary.org/obo/	IAO_0000111 editor preferred term	Entity
		IAO_0000114 has curation status	Entity and object property
		IAO_0000115 definition	Entity
		IAO_0000117 term editor	Entity
		IAO_0000118 alternative term	Entity
		IAO_0000119 definition source	Entity
		IAO_0000412 imported from	Imported components
oboInOwl	http://www.geneontology.org/formats/oboInOwl#	id	Entity
		created_by	Any component
		creation_date	Any component
		hasDbXref	Entity
protege	http://protege.stanford.edu/plugins/owl/protege#	defaultLanguage	Ontology

Source: By researcher's thesis.

4.2.5 Delivery phase

The delivery phase consists of publishing the deliverable piece of the ontology to the community for their use. During the implementation phase, ontologits should follows the architectural issues defined at the *ontology architecture specification*. Figure 47 presented a workflow of the activities that ReBORM specify to the delivery phase.

Figure 47: Delivery phase activities of ReBORM



Source: Prepared by the author.

During the *ontology merge* the already publishied ontology is conbined to the new ontology piece using one of the strategies described in subsection 2.3.3 (alignment, mapping, combination, and integration). In *ontology delivery*, the ontology could be publish in an ontology repository such as BioPortal and OBO Foundry.

4.3 OntONeo development process

The scope of OntONeo is very broad as we showed in subsection 4.1. The development of OntONeo uses the ReBORM methodology we presented in subsection 4.2. Thus, OntONeo is developed thought multiples iterations, dividing the development into stages and delivering ontology content at the end of each part. We referred to each deliverable as an ontology version or piece of ontology. Here we describe some main results of the iterations already performed. We focus on describing the first iteration because some of the activities defined in ReBORM are specifics this first iteration.

4.3.1 Concept phase of OntONeo development

The development of OntONeo begins with the ***concept phase*** and followed the activities workflow presented in Figure 39 (subsection 4.2.1). Table 13 presents a summary of main activities results.

Table 13 - Summary of concept phase results

Activity	Results
Requirements specification	Appendix 3
Knowledge acquisition	Subsection 4.1
Ontology search activity	Figure 53 and Figure 52

Source: Prepared by the author.

Ontology search activity

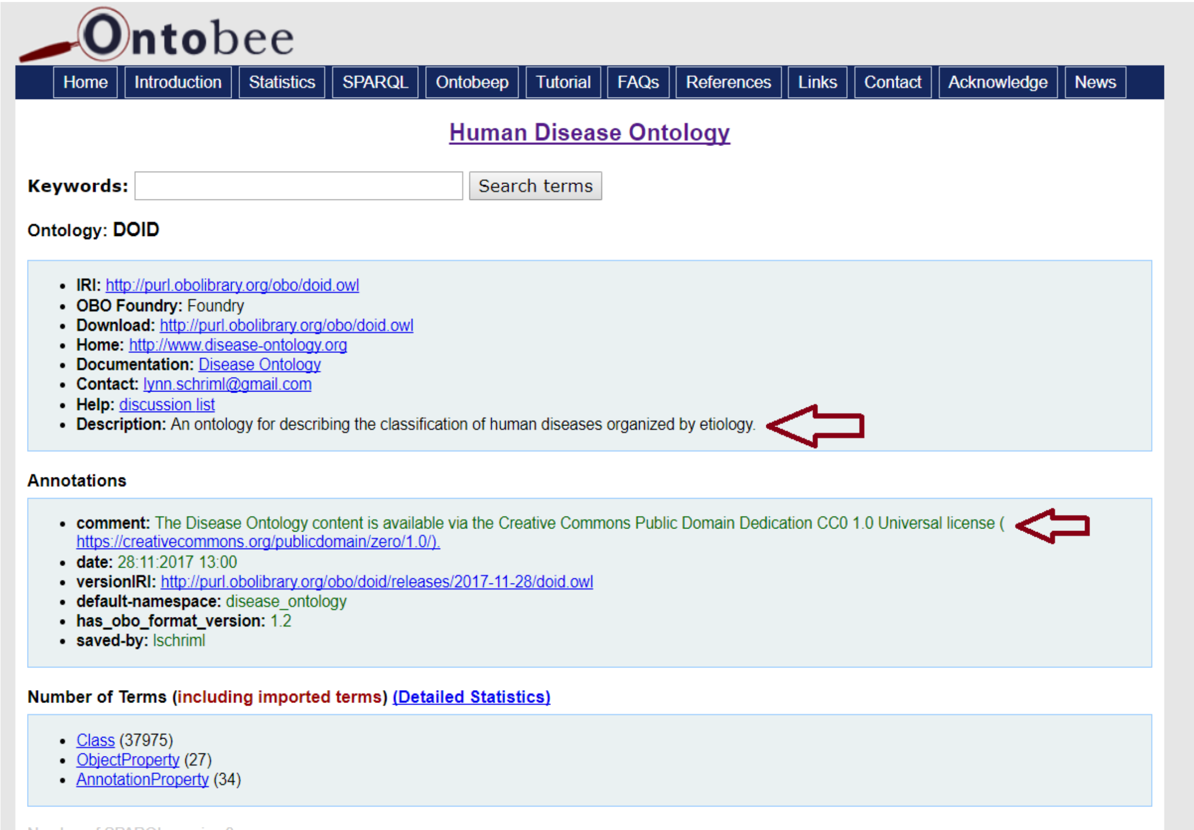
Owing to the fact that OntONeo is an ontology from the biomedical domain constructed with the goal of conforming to OBO Foundry principles, during the *ontology search activity*, our ontology search was confined to reusable ontologies on the search engines Ontobee and BioPortal, which respectively index OBO Foundry ontologies and biomedical ontologies. This decision took into consideration the fact that these two search engines cover ontologies and biomedical terminologies. However, Ontobee Portal was prioritized due the reuse of OBO Foundry ontologies.

The Ontobee search, was effectuated in two ways, search by ontology according to ontology domain of coverage (search phase 1), and search by key terms from OntONeo (search phase 2). When Ontobee search did not return any occurrence, a search by key terms was effectuated in BioPortal (search phase 3). At this moment we did not define an ontology selection criteria because search goal is just to have a list of ontologies candidates for reuse.

Search Phase 1:

For each ontology listed on Ontobee Portal, the ontology metadata (Description and comment) was analyzed to identify the coverage domain of each ontology (Figure 48).

Figure 48: Evidence of ontology search by metadata on Ontobee Portal



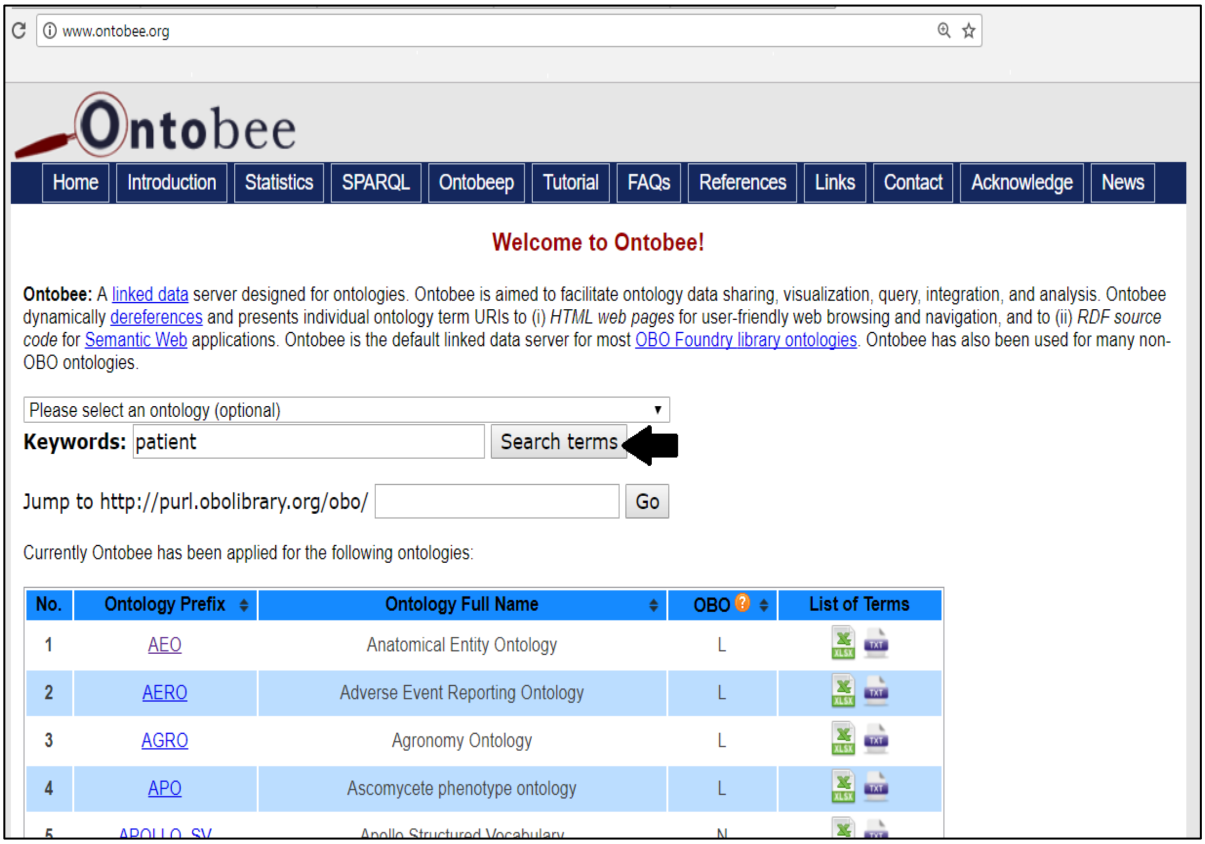
Source: Prepared by the author.

The selected ontologies addressed general subjects relevant to the OntoNeo domain and general aspects based on the EHR sections, such as: embryology, anatomy, disease, vaccine, demographic information.

Search Phase 2:

This search looked for all occurrences of all OntoNeo relevant terms in ontologies listed on the Ontobee Portal. The terms searched in Search Phase 2 are listed in section 7 of the ontology requirements specification document of OntONeo (Appendix 3). When we use the Ontobee Portal to search for ontologies that incorporate with a specific term, the search machine compare the term informed in search field with the entity property *rdfs:label*[43]. For example, according to Figure 49, we entered "*patient*" to search for ontologies that reference this term.

Figure 49: Evidence of ontology search by term on Ontobee Portal.



Source: Prepared by the author.

[43] "*rdfs:label is an instance of rdf:Property that may be used to provide a human-readable version of a resource's name (W3c, 2014f, on-line)*".

Figure 50 shows an extract of the results obtained. In some ontologies the term *patient* is referenced in isolation, as in the case of ontology IDOMAL, OAE and SIO. In others the term patient is combined with another term as is the case of IAO and ERO ontologies, which respectively include term "*patients section*" and "*patient screening*". Some cases, the term patient is referenced both in isolation and combined as is the case of OBI and ORMSE ontologies which include term "*patient*" and "*patient role*".

Figure 50: Result of search for the term patient on Ontobee Portal.

Source: Prepared by the author.

On the basis of the combined results of Search Phases 1 and 2 we selected OBI, ORMSE and SIO. These three ontologies covered the term "*patient*" and, according Ontobee description metadata…

- Ontology for Biomedical Investigations (OBI), is "an integrated ontology for the description of life-science and clinical investigations", wich means to be a reference ontology in biomedical investigations.
- Ontology of Medically Related Social Entities (OMRSE), is an ontology that "covers the domain of social entities that are related to health care, such as demographic information and the roles of various individuals and organizations", wich means to be a reference ontology in social domain of biomedical investigations.
- The Semanticscience Integrated Ontology (SIO) provides a simple, integrated ontology of types and relations for rich description of objects, processes and their attributes, wich seens to be a reference of semantic ontology integration.

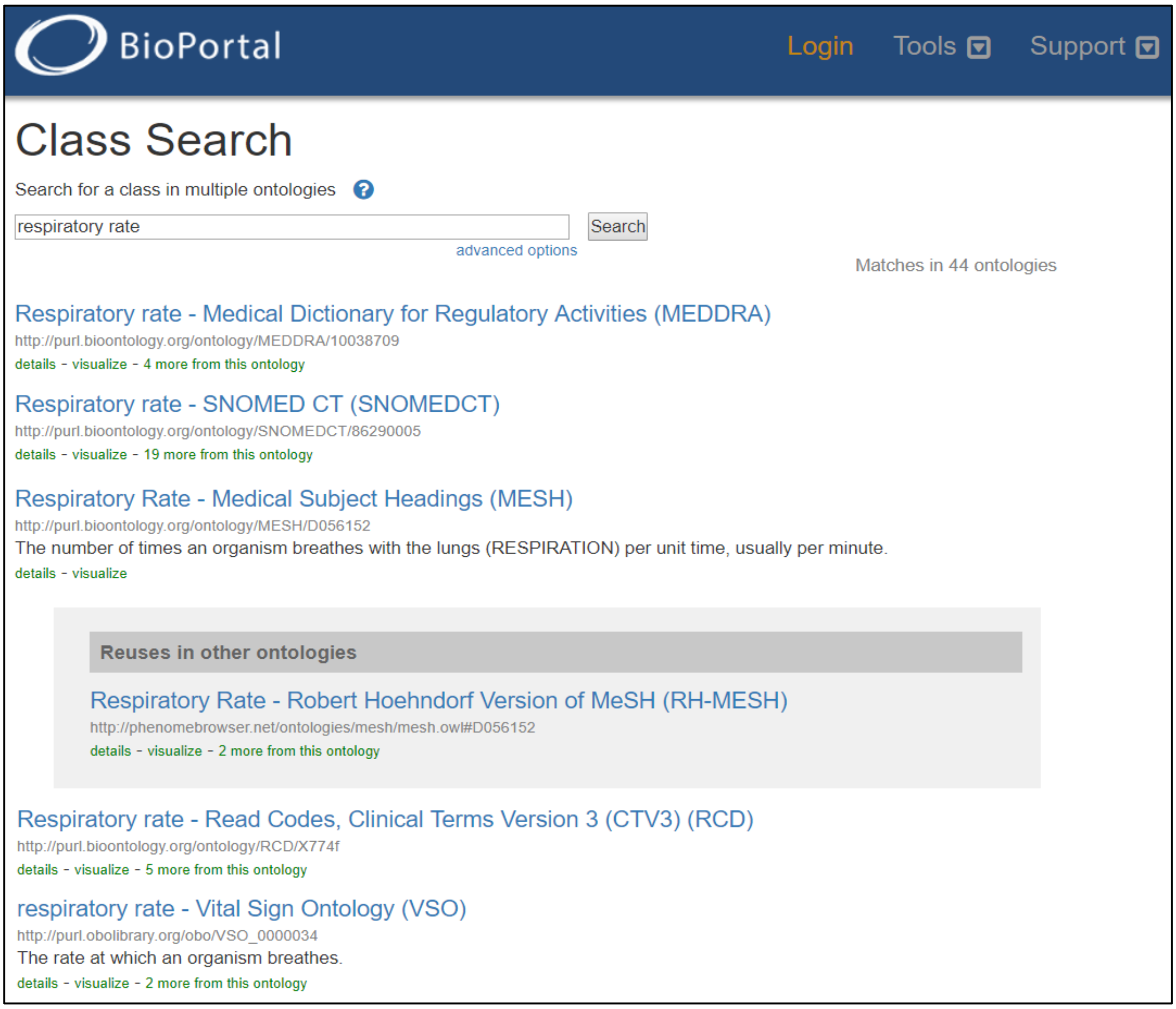
In the other hand, despite IDOMAL (Malaria Ontology) covered the term "patient" also, such ontology is an ontology that covers details of Malaria knowledge domain, on a detail level higher than OntONeo expects, thus, at this moment we did not consider IDOMAL an ontology candidate for reuse.

Search Phase 3:

This search looked for all occurrences of all OntoNeo relevant terms in ontologies listed on the BioPortal once that Ontobee Portal did not have any occurrence (Figure 51). The terms searched here are listed in section 7 of the ontology requirements specification document of OntONeo (Appendix 3).

The final result of our ontology search activity is a list of candidate ontologies for reuse (Figure 53) and a list candidate terminologies containing terms reusable in OntoNeo (Figure 52).

Figure 51: Result of ontology search by term “respiratory rate” on BioPortal.

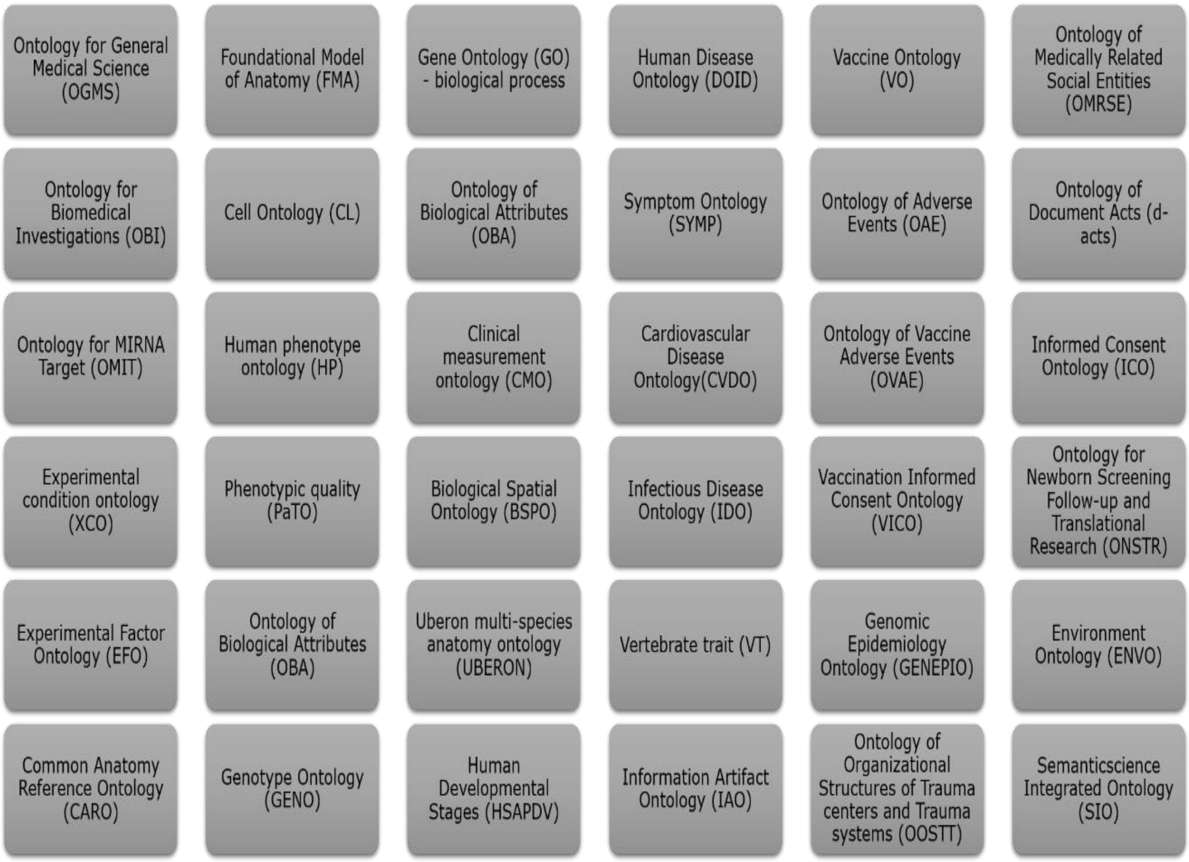


Source: Prepared by the author.

Figure 52: Candidate terminologies for reuse by OntONeo

Source: Prepared by the author.

Figure 53: Candidate ontologies for reuse



Source: Prepared by the author.

Ontology Design Pattern Search

Finalizing the ontology search activity, we performed the activity *Ontology Design Pattern Search*, which consisted in manually checking the catalogue of ontology design patterns[44] looking for candidates for reuse in the OntONeo project. In fact, after a detailed analysis of the complete list of patterns, we realized that most of such patterns are incompatible with definitions and principles of OBO Foundry ontologies, perhaps because they follow a different lineage of ontologists and are even supported by top-level ontologies other than BFO.

However, OntONeo considered some proposed Content Ontology Design Patterns such as: *Disposition*[45] (Röhl; Jansen, 2011), *PartOf*[46], Gene Ontology Top Level[47], and Participation[48].

---

[44] Available at: http://ontologydesignpatterns.org/wiki/Community:ListPatterns

[45] Details in: http://ontologydesignpatterns.org/wiki/Submissions:Disposition

[46] Details in: http://ontologydesignpatterns.org/wiki/Submissions:PartOf

[47] Details in: http://ontologydesignpatterns.org/wiki/Submissions:GO_Top

[48] Details in: http://ontologydesignpatterns.org/wiki/Submissions:Participation

Furthermore, considering a possibility to create a taxonomy from an enumeration or a list of terms, we took into account some Reengineering Ontology Design Patterns, namely, "*Classification scheme - adjacency list model - to Taxonomy*[49]" and "*Classification scheme - path enumeration model - to Taxonomy*[50]".

Given the characteristic of OntONeo development to maximize ontology reuse, the category Alignment[51] of Ontology Design Pattern should be considered. Currently, there are fourteen subcategories of Alignment Design Pattern, and during the inception and implementation phases, they were checked when necessary.

4.3.2 Inception phase of OntONeo development

In order to initiate this phase, we use the ontology requirements specification document to identify, within the broad scope of OntoNeo, a criterion to define smaller and relatively self-contained modules. To begin the first iteration, we considered dividing the scopes of iterations according to the corresponding medical disciplines of embryology, anatomy, gynecology, obstetrics, and pediatrics, which are relevant to OntONeo.

All iterations of the inception phase during OntONeo development followed the activities workflow presented in Figure 41 (subsection 4.2.2). Table 14 presents a summary of main activities results of the first iteration.

Table 14 - Summary of inception phase results of first iteration

Activity	Results
Iteration planning activity	Section 1 and 2 of Appendix 4
Ontology elicitation activity	Appendix 4
Ontology selection activity	Figure 55
Ontology documentation activity	Project homepage Development repository Ontology annotations

Source: Prepared by the author.

[49] Details in: http://ontologydesignpatterns.org/wiki/Submissions:Classification_scheme_-_adjacency_list_model_-_to_Taxonomy

[50] Details in: http://ontologydesignpatterns.org/wiki/Submissions:Classification_scheme_-_path_enumeration_model_-_to_Taxonomy

[51] Details in: http://ontologydesignpatterns.org/wiki/Category:AlignmentOP

According to ReBORM, *ontology documentation* activity starts o inception phase and persists until implementation phase. Consequently, the OntONeo documentation happens during all these phases (initialization, design and implementation) considering the type of documentation made.

OntONeo development adopted three documentation patterns: i) use of relevant ontology annotations. ii) project homepage[52]. iii) development repository[53]. We presents in subsection 4.3.3 both the project homepage and the development repository.

During the *iteration planning activity*, the scope of the first iteration of the inception phase of OntONeo development the focus was anatomic structures of human body, mainly woman body. Such decision took account of the woman physical examination performed during a medical encounter, in which physicians evaluate woman body and then they record such examination and findings on specific sections of medical records. Once the purpose and scope of the iteration have been defined, we begin the *ontology elicitation activity* involving techniques of knowledge acquisition, which generated the ontology requirements specification document of such iteration (Appendix 4).

Next, in the *ontology selection activity*, considering the iteration scope and given the list of candidate ontologies of the concept phase, we define selection criteria to establish what are the ontologies and terms that we must reuse. We observed that the domain of anatomy itself is very large in scope and is already covered by several OBO Foundry ontologies, including:

- Foundational Model of Anatomy (FMA[54]): (Musen, 2013, Rosse; Mejino, 2003, Rosse; Mejino Jr, 2008).
- Uberon multi-species anatomy ontology (UBERON[55]): (Haendel *et al.*, 2014, Haendel *et al.*, 2009, Mungall *et al.*, 2012).
- Human Developmental Anatomy (EHDA[56]): (Bard, 2012, Hunter *et al.*, 2003).

---

[52] OntONeo project homepage address: *https://ontoneo.com/*

[53] OntONeo GitHub repository address: *https://github.com/ontoneo-project/Ontoneo*

[54] Details of Foundational Model of Anatomy are available at http://www.obofoundry.org/ontology/fma.html and http://si.washington.edu/projects/fma

[55] Details of Uberon are available at http://uberon.github.io/ and http://www.obofoundry.org/ontology/uberon.html

[56] Details of Human Developmental Anatomy are available at http://www.obofoundry.org/ontology/ehdaa2.html

• Common Anatomy Reference Ontology (CARO[57]): (Haendel *et al.*, 2008).

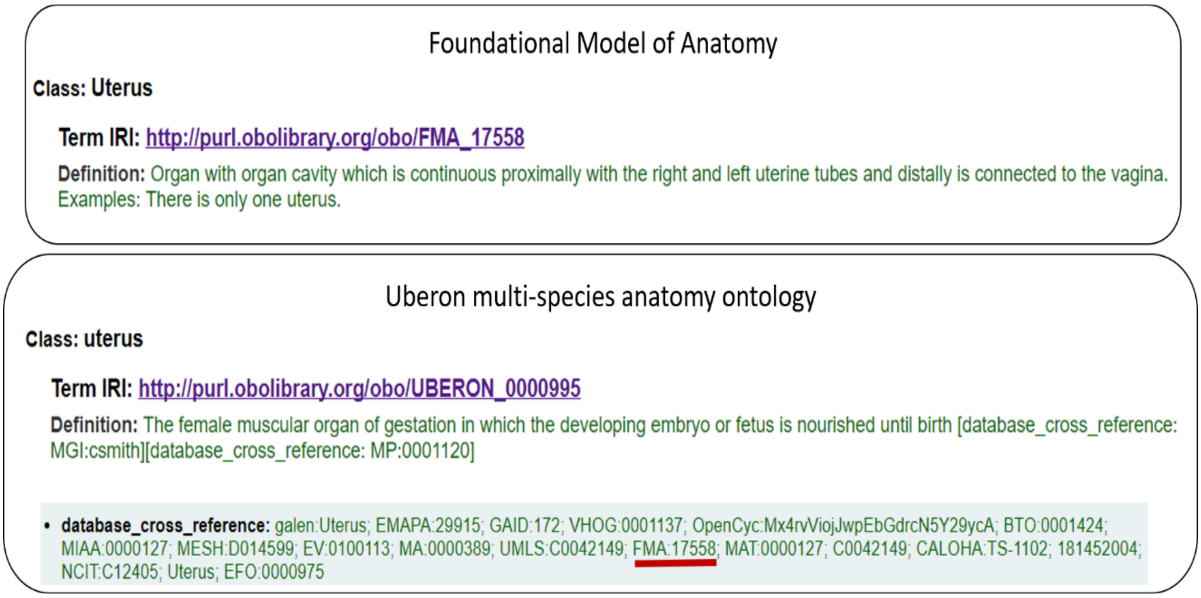
We concluded that OntONeo did not need to provide any further representation of the anatomy domain, but rather that it should reuse legacy ontology content relating to this domain.

Notwithstanding the foregoing, at this point, we highlight a challenge in ontology reuse, sometimes, during an ontology term search using some search engine as Ontobee, we can find an occurrence of such term but its definition could not express what we really need or expect.

In addition, a term search could find more the one occurrence, for instance, according to Figure 54, the term '*uterus*' exists in both FMA (as FMA_17558) and UBERON (as UBERON_0000995). Both definitions satisfy the need of OntONeo. We select the former, since this is the original source and incorporate a mapping to the latter. In such cases in which cross-reference is not defined, we vote in favor of the oldest ontology or the most referenced ontology.

So that, Figure 55 showed the suitable ontologies and terminologies selected.

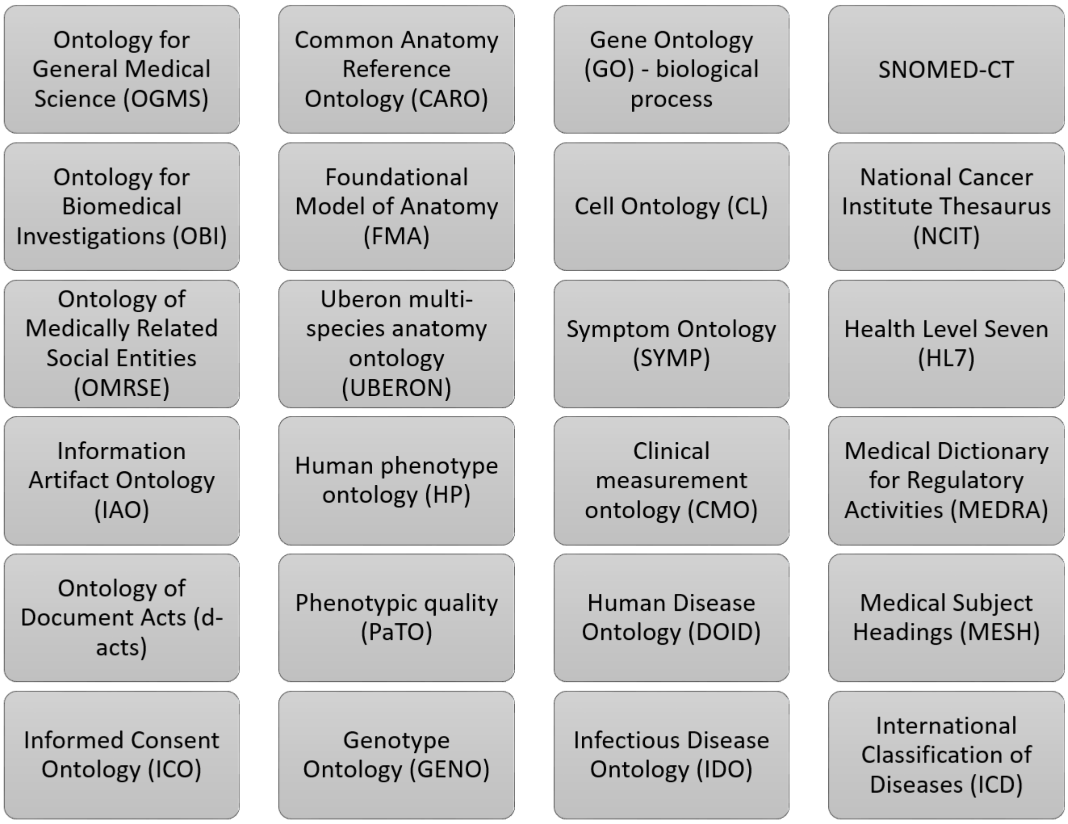
Figure 54: Multiples results search on Ontobee Portal.



Source: Prepared by the author.

[57] Details of Common Anatomy Reference Ontology are available at http://www.obofoundry.org/ontology/caro.html

Figure 55: Ontologies and terminologies candidates to reuse on iteration 1.



Source: Prepared by the author.

4.3.3 Design phase of OntONeo development

All iterations of the design phase during OntONeo development followed the activities workflow presented in Figure 43 (subsection 4.2.3). Table 15 presents a summary of main activities results of the first iteration.

Table 15 - Summary of design phase results of first iteration

Activity	Results
Ontology architecture specification	Appendix 5, Figure 56 and Figure 57
Ontology conceptualization	Example → Figure 58
Ontology retrieval	Example → Figure 59
Ontology localization	Examples → Figure 60, Figure 66 and Figure 67

Source: Prepared by the author.

Ontology architecture *activity:*

Starting the design phase of first iteration, we established the ontology architecture used during whole OntONeo development. Since the beginning of the project, there was a desire of the ontologist to publish OntONeo ontology as an OBO Foundry library. Thus, the ontologist sought, on the OBO Foundry website, information against OBO Foundry principles (subsection 2.4.2) relevant to define the ontology architecture.

So that, we specified the OntONeo architecture and register our architecture decisions on the document of architecture specification (Appendix 5), following the template of ReBORM (Figure 44). In short, some decisions are:

- The ontology elements should support at least English and Portuguese languages.
- The ontology should follow the principles of the OBO Foundry.
- The implementation language must be OWL 2 Web Ontology Language.
- The ontology should use as top-level ontology Basic Formal Ontology version 2.0.
- The ontology should reuse other ontologies already accepted by the OBO Foundry whenever possible.

Next, we describe some steps performed to define others OntONeo architectural elements.

Define the ontology PREFIX or ID-Space.

The idea here is to create a short word such an acronym to be associated to your ontology in different ways. When it is necessary to simplify the namespace (see next subsection), it is common to use a PREFIX. Additionally, each ontology element (e.g. class, relation, and instance) has a unique URI identifier usually formed by the concatenation of the default namespace and the local identifier.

In order to determine the PREFIX of the ontology of the obstetric and neonatal domain, a research in the OBO Foundry ontologies list was performed to ensure that none ontology already uses the target prefix, which is ONTONEO. Then, the ontologist used the service *prefix.cc*, to guarantee that none other resource uses the ONTONEO prefix. Prefix.cc[58]

[58] Prefix.cc service is available at http://prefix.cc/. Last date of access: June 13, 2017.

is a service created for RDF developers in which help them on the search for URI prefixes and namespaces.

The OBO Foundry recommends that the local identifier should be a concatenation of the prefix (also labelled as ID-Space), a character "_" (*underscore*), and a sequence of numbers, as the formula below.

OBO Foundry local identifier: <IDSPACE>_<NUMBER>
Example: ONTONEO_00000001

Define the ontology namespace.

In short, a namespace is a mechanism to provide generic context for encapsulating items. A namespace is a sequence of characters or symbols that serve to uniquely identify a collection of names or identifiers. A default namespace is a valid URI or IRI followed by a character '/' or '#'. For this, it defined a persistent URL (purl) using the PURL service[59].

Following OBO Foundry recommendations, the ontology of the obstetric and neonatal domain (OntONeo PURL[60]) is located under the OBO PURL domain[61], as subdomain. Before creation of OntONeo PURL, we also used the service *prefix.cc* to verify the existence or not of a similar namespace.

Ontology development repository

A primary intention of creating a development project is to enable a collaborative worldwide development. As a software development project, it is desirable that an ontology development project maintains a centralized repository but with distributed access to potential project developers. Ideally, this repository should allow the versions control of the project, parallel development of project parts, integration of different development branches, request reviews, and tracking of issues and requests.

Thus, OntONeo Project uses GitHub[62] repository platform due to the familiarity of the researcher with it. Figure 56 showed an overview of OntONeo repository created under the URL: *https://github.com/ontoneo-project/Ontoneo*.

---

[59] PURL service is available at *https://archive.org/services/purl/*. Last date of access: June 13, 2017.

[60] The ONTONEO PURL domain of OntONeo created is *http://purl.obolibrary.org/obo/ontoneo* (simplifying "*/obo/ontoneo*").

[61] The OBO PURL domain is *http://purl.obolibrary.org/obo/* (simplifying "*/obo*").

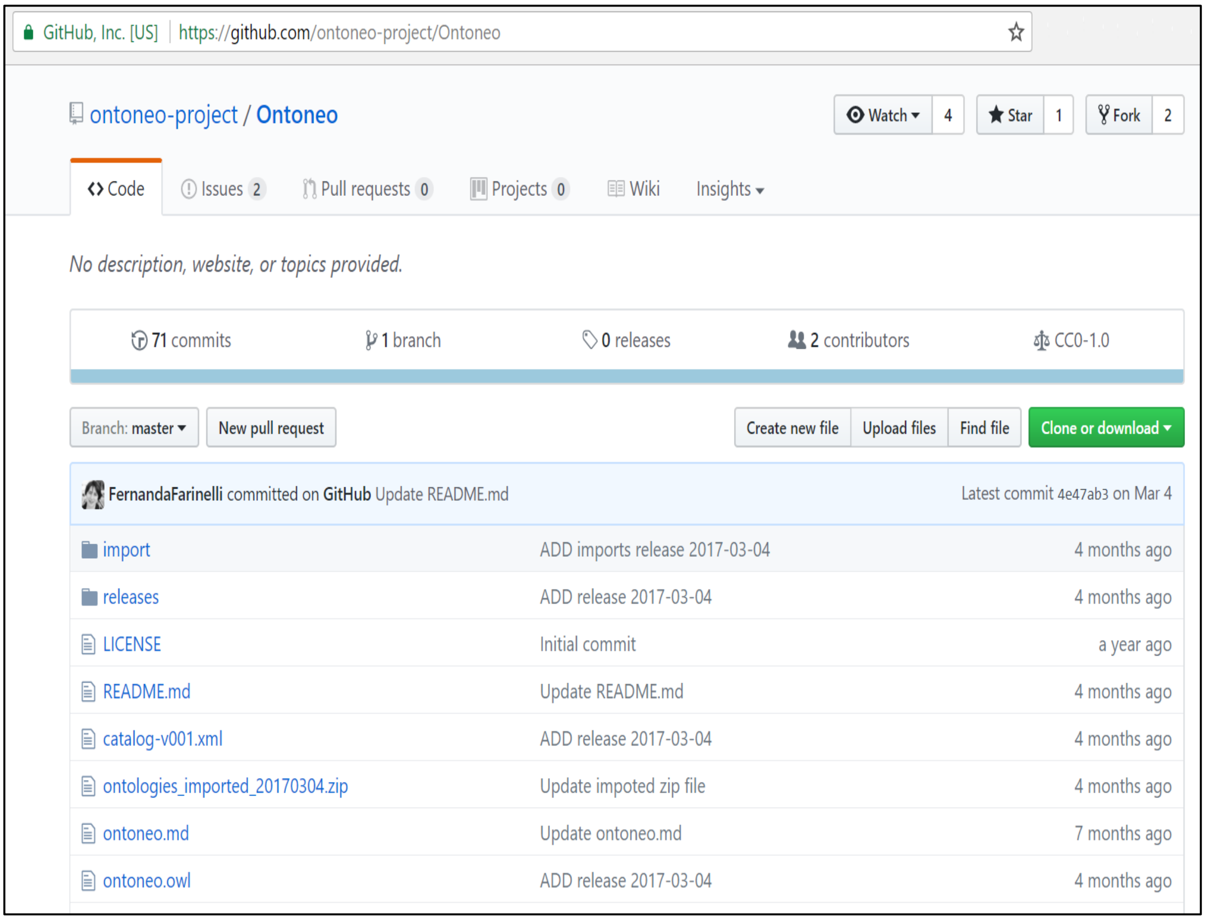
[62] GitHub platform is available at https://github.com/. Last date of access: June 13, 2017.

In order to better configure the repository, we chose to create a general user with an administrator profile. This administrator user was the creator of the ontoneo project repository. One advantage of using such a user is to maintain the independence of the project repository with the researcher. In the future, if convenient, others can assume the OntONeo project.

Project homepage

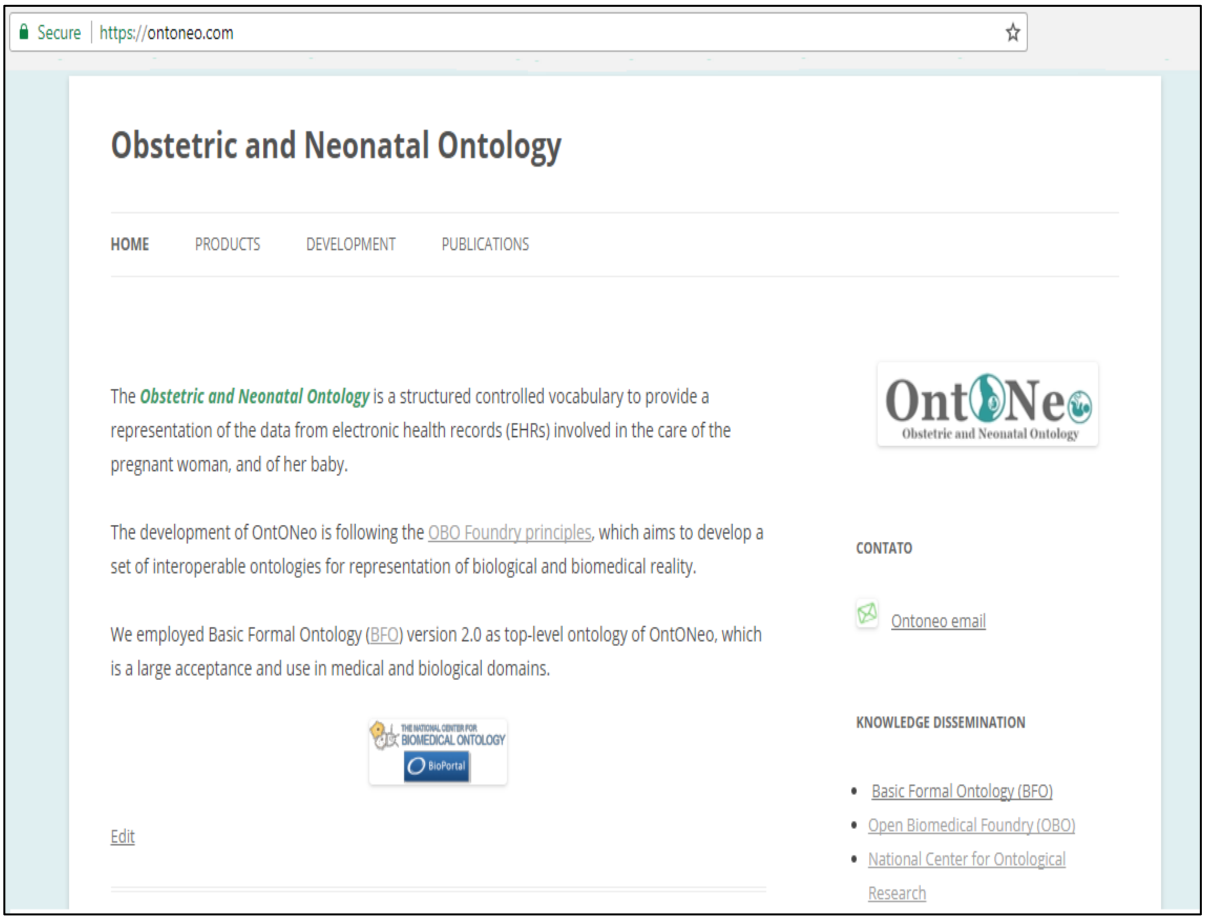
This thesis's researcher created the OntONeo project homepage in order to turn this project available to the community of users and ontologist both the project itself and any information related to the project. The researcher used the WordPress web platform for creating this home page and acquired the domain of the homepage. The OntONeo project homepage address is https://ontoneo.com/. Figure 57 showed the main page of OntONeo project.

Figure 56: Overview of GitHub repository to OntONeo project.



Source: Prepared by the author.

Figure 57: Overview of the OntONeo project website.



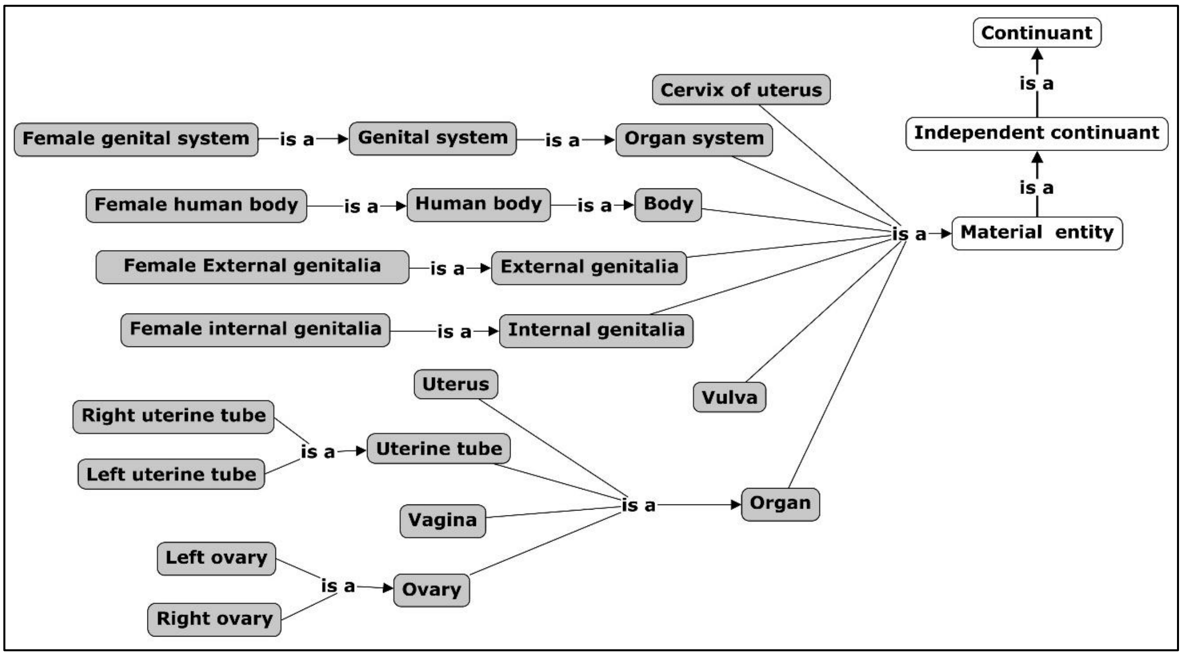
Source: Prepared by the author.

Ontology conceptualization:

After the architecture specification, we start the *ontology conceptualization*, in which we define a first list of terms retrieving such terms from sources of knowledge about the obstetric and neonatal domain. These sources include textbooks, scientific journals, reports, manuals, clinical protocols, medical guidelines, medical records forms and templates, case studies, user stories, among others. In addition, the term identification also retrieved terms quoted by domain experts during interviews and observations.

Then, we *group each term*, classifying as an entity, *against the top-level ontology* BFO (version 2) (see subsection 2.3.5). Finally, we built the graphic conceptual model reflecting the term list using CMaps tool. Figure 58 showed an example of the conceptual model created during OntONeo development.

Figure 58: OntONeo's conceptual model from the is_a hierarchy of the first iteration.



Source: Prepared by the author.

Ontology retrieval:

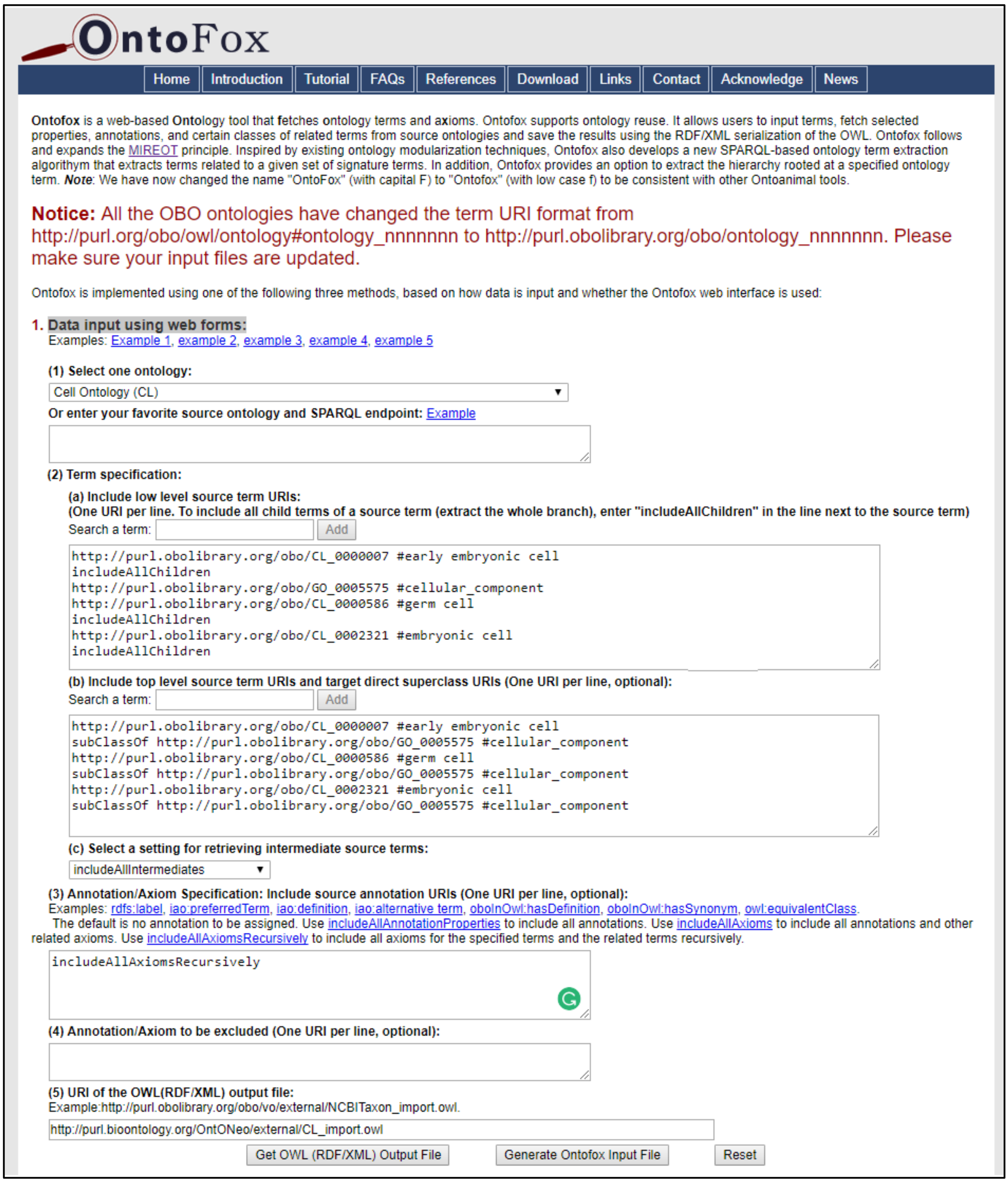
Next activity performed was the *ontology retrieval,* in which we prepared the ontologies for reuse in OntONeo. As we mentioned above, OntONeo development privileges the reuse of pre-existing ontologies that use BFO top-level ontology.

Some ontologies were completely retrieved, as was the case with BFO and RO. For this, the ontologist recovered these two ontologies from the OBO Foundry portal downloading the OWL files available.

Next, we prepared the ontologies using the approach of MIREOT, in which we reuse just parts from others ontologies instead of the whole ontology. We used Ontofox tool to support our retrieval for ontology parts. So, we selected the option "*Data input using web forms*" on OntoFox and filled all fields. For example, Figure 59 presents an example of retrieving elements of Cell Ontology (CL).

All retrieval ontology was saved in an OWL language because was the formal language defined to OntONeo development. After retrieval of the OWL files for all ontologies, we removed some documentation from these files that are not relevant to OntONeo development, namely: non-relevant comments and contributor information. In addition, the ontologist also included documentation according to Table 11. Note that such documentation was done through annotation properties, thus consisting of an annotation of the *ontology annotation* activity.

Figure 59: Example of Ontofox use to retrieval Cell ontology.



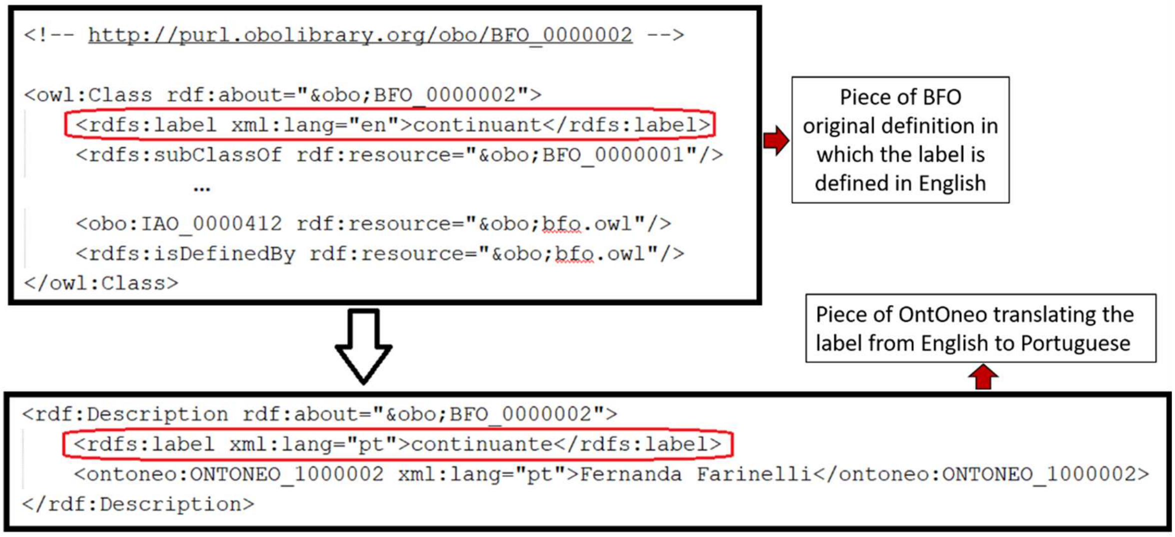
Source: Prepared by the author.

Therefore, in order to meet the non-functional requirement of the developing ontology, namely:

> The ontology must support a multilingual scenario at least the following languages: English and Portuguese.

Finally, we followed to *ontology localization* activity to prepare such ontology files in both English and Portuguese (Figure 60). Thus, during the location of the ontology, we focused on translating and adapting the both label and definition of the ontologies or piece of ontologies prepared on ontology retrieval activity.

Figure 60: Evidence of performed the ontology localization activity.



Source: Prepared by the author.

4.3.4 Implementation phase of OntONeo development

All iterations of the implementation phase of OntONeo development followed the activities workflow presented in Figure 46 (subsection 4.2.4). Table 16 presents a summary of main activities results of the first iteration.

Table 16 - Summary of implementation phase results of first iteration

Activity	Results
Top-level ontology importation	Appendix 5, Figure 56 and Figure 57
Ontology integration	Example → Figure 61
Ontology annotation	Examples → Figure 63, Figure 66, Figure 67
Ontology formalization	Examples → Figure 62, Figure 64, Figure 65, Figure 68, Figure 69, Figure 72, Figure 73, Figure 74, Figure 78
Ontology evaluation	Figure 83

Source: Prepared by the author.

As defined in the OntONeo architecture specification document (Appendix 5), the ontology editor used during implementation phase is the Protégé desktop (versions 4.3 and 5.2). Protégé is a widely used open-source ontology development tool supporting users to define ontologies with a graphic friendly user interface (Noy *et al.*, 2003). To start the ontology development we first configured the ontology development environment in Protégé according to the guide presented in Appendix 6.

Top-level ontology importation:

Starting OntONeo development, in the first iteration, we imported the OWL file of Basic Formal Ontology (BFO[63]) version 2.0 (Arp; Smith; Spear, 2015, Smith *et al.*, 2015, Smith; Kumar; Bittner, 2005), the top-level ontology that grounds OntONeo. For this, we followed the steps presented in Appendix 7. In addition to BFO, we imported also the OWL file of Relations Ontology (RO[64]) (Smith *et al.*, 2005).

We imported OWL files of both BFO and RO that we prepared during the design phase. We chose a design approach to import the whole ontology from an external file instead to import straight from the IRI due to the need for adjustments in these ontologies (BFO and RO), for instance, to translate of label and definitions to Portuguese. According to Figure 61, BFO and RO are ontologies imported to OntONeo.

Ontology integration:

During the ontology integration activity the OWL files of the reusable ontologies prepared during the design phase in our project were imported (Appendix 7). We decided to import the OWL files instead of embedded just parts of these files to ensure the modularization feature and the easy maintenance of the reused resources. Since OntONeo development is iterative-incremental, at each iteration, the ontologist can include new elements in the ontologies in reuse or include new ontologies for reuse. Figure 61 presents a piece of OntONeo with all imports until the iteration 5.

We realized the possibility of contributing to all ontology projects reused by OntONeo, adding the translations of the labels and definitions for Portuguese language in its specific project, thus expanding the possibility of reusing these ontologies in other projects that cover the Portuguese language. However, considering the limited time to prepare this thesis,

---

[63] Details of Basic Formal Ontology are available at http://basic-formal-ontology.org/ and http://www.obofoundry.org/ontology/bfo.html.

[64] Details of Relations Ontology are available at http://www.obofoundry.org/ontology/ro.html and https://github.com/oborel/obo-relations/.

we chose to keep the translated versions only in our project. In the future, we will demand the inclusion of these translations in the specific ontologies projects.

Figure 61: OntONeo Piece showing ontologies in reuse

```
<owl:Ontology rdf:about="http://purl.obolibrary.org/obo/ontoneo/ontoneo.owl">
    ...
   <owl:imports
rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/omiabis_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/oae_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/bcgo_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/pato_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/symp_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/omrse_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/hp_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/iao_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/d-acts_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/sio_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/fma_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/ogms_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/uberon_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/bfo_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/obi_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/mpath_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/doid_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/ero_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/mp_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/cl_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/go_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/ro_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/vo_import.owl"/>
   <owl:imports rdf:resource="http://purl.obolibrary.org/obo/ontoneo/import/efo_import.owl"/>
   ...
</owl:Ontology>
```

Source: Prepared by the author.

Ontology formalization:

At this point we formalized the conceptual model defined in the design phase. Ontologies in reuse already embrace several terms covered by OntONeo. Thus, we represent only the elements of ontology that were not found in preexisting ontologies.

During the formalization activity, OntONeo project created new annotation properties, object properties, classes, and relations between classes. Following, we present some of the main contributions formalized in the OntONeo project.

Formalization of annotation properties:

An important detail about OntONeo is the fact that ontology maintains its own components and such in reuse both in English and Portuguese, according to a non-functional requirement defined. In the development of OntONeo, the default language of an ontology is determined by the *defaultLanguage*[65]. Thus, both the OntONeo and the ontologies in reused have the *defaultLanguage* property set to English (en). Following the idea of *defaultLanguage* annotation property, which defines the standard language used for ontology development, OntONeo defined an annotation property labelled "*alternative language*". Figure 62 showed the definition code of “*alternative language*" annotation property.

Figure 62: Definition of the “alternative language” annotation property of OntONeo.

```
<owl:AnnotationProperty rdf:about="&obo;ontoneo/ONTONEO_1000000">
  <rdfs:label xml:lang="en">alternative language</rdfs:label>
  <rdfs:isDefinedBy rdf:datatype="&xsd;anyURI">
     ontoneo/ontoneo.owl
  </rdfs:isDefinedBy>
  <obo:IAO_0000115 xml:lang="en">
    An alternative language in which the ontology is represented.
  </obo:IAO_0000115>
  <obo:IAO_0000117 xml:lang="en">
    PERSON:Fernanda Farinelli
  </obo:IAO_0000117>
  <obo:IAO_0000111 xml:lang="en">alternative language</obo:IAO_0000111>
  <obo:IAO_0000114 rdf:resource="&obo;IAO_0000125"/>
  <rdfs:range rdf:resource="&xsd;language"/>
</owl:AnnotationProperty>
```

Source: Prepared by the author.

The purpose of the "*alternative language*" annotation is to show for ontology users in which alternative languages the ontology components are defined. Thus, Figure 63 presents a piece of the formalization of OntONeo in OWL language evidencing the use of both *defaultLanguage* and "*alternative language*" annotations properties.

Figure 63: Piece of OntONeo code with language definitions properties.

```
<owl:Ontology rdf:about="http://purl.obolibrary.org/obo/ontoneo/ontoneo.owl">
  ...
  <protege:defaultLanguage xml:lang="en">en</protege:defaultLanguage>
  <ontoneo:ONTONEO_1000000 xml:lang="pt">pt</ontoneo:ONTONEO_1000000>
  ...
</owl:Ontology>
```

Source: Prepared by the author.

Just as it is important to annotate who is the creator and/or editor of a component of the ontology, it is necessary to indicate the responsible for the translation of such

[65] The defaultLanguage annotation property was defined at http://protege.stanford.edu/plugins/owl/protege.

components. For this, OntONeo defined an annotation property labelled "*translator*" to indicate the personal name of the responsible for translate an ontology from default language for the alternative language as showed on Figure 64.

Figure 64: Definition of the “translator” annotation property of OntONeo.

```
<owl:AnnotationProperty rdf:about="&obo;ontoneo/ONTONEO_1000001">
  <rdfs:label xml:lang="en">translator</rdfs:label>
  <rdfs:isDefinedBy rdf:datatype="&xsd;anyURI">
      “ontoneo/ontoneo.owl”
  </rdfs:isDefinedBy>
  <obo:IAO_0000115 xml:lang="en">A person how is responsible for
     translating the entity annotations from English to another
     language.
  </obo:IAO_0000115>
  <obo:IAO_0000117 xml:lang="en">
     PERSON:Fernanda Farinelli
  </obo:IAO_0000117>
 <obo:IAO_0000111 xml:lang="en">translator</obo:IAO_0000111>
  <rdfs:range rdf:resource="&xsd;Name"/>
</owl:AnnotationProperty>
```

Source: Prepared by the author.

Additionally, we realized that some ontology could be translated into several languages, then, OntONeo defined the annotation property labelled “*translator to Portuguese*” as a sub property of “*translator*” defined as showed in Figure 65.

Figure 65: Definition of the “translator to Portuguese” annotation property of OntONeo.

```
 <owl:AnnotationProperty rdf:about="&obo;ontoneo/ONTONEO_1000002">
  <rdfs:label xml:lang="en">translator to Portuguese</rdfs:label>
  <rdfs:isDefinedBy rdf:datatype="&xsd;anyURI">
    “ontoneo/ontoneo.owl”
  </rdfs:isDefinedBy>
  <obo:IAO_0000115 xml:lang="en">A person how is responsible for
    translating the entity annotations from English to Portuguese.
  </obo:IAO_0000115>
  <obo:IAO_0000117 xml:lang="en">
    PERSON:Fernanda Farinelli
  </obo:IAO_0000117>
  <rdfs:subPropertyOf rdf:resource="&obo;ontoneo/ONTONEO_1000001"/>
  <obo:IAO_0000111 xml:lang="en">
    translator to Portuguese
  </obo:IAO_0000111>
  <rdfs:range rdf:resource="&xsd;Name"/>
</owl:AnnotationProperty>
```

Source: Prepared by the author.

Therefore, Figure 66 and Figure 67 present pieces of formalization of OntONeo in OWL language evidencing the use of “*translator to Portuguese*” annotations properties in two different forms.

Figure 66: Example 1 of OntONeo using translator annotation property.

```
<!-- http://purl.obolibrary.org/obo/BFO_0000015 -->
<rdf:Description rdf:about="http://purl.obolibrary.org/obo/BFO_0000015">
   <rdfs:label xml:lang="pt">processo</rdfs:label>
   <ontoneo:ONTONEO_1000002 xml:lang="en">
      PERSON: Fernanda Farinelli
   </ontoneo:ONTONEO_1000002>
</rdf:Description>
```

Source: Prepared by the author.

Figure 66 showed an entity of an ontology in reuse, in this case, is the entity "*process*" of BFO. Note on line 3 of this example that the label property (*rdfs:label*) was used jointly with an attribute that specifies the language of the element content (*xml:lang*), indicating that such label is expressed in Portuguese (*<rdfs:label xml:lang="pt"*>), followed by the content of such label expressed in Portuguese, i.e. "*processo*".

Figure 67 showed the second form example, in which the ontologist annotated an entity defined by OntONeo, the entity "*ONTONEO_00000067*". In this case, according to lines group 1, the label property was defined both in English and Portuguese, respectively as "*prenatal encounter*" and "*acompanhamento pré-natal*".

So, such entity was defined using the annotation property IAO_0000115, from Information Artifact Ontology (IAO[66]) labeled "*definition*", according to lines group 2. In addition, according to lines group 3, we used the annotation property IAO_0000118 labeled "*alternative term*", establishing synonyms to such entity. Such annotations was created both in English and Portuguese.

In both cases, Figure 66 and Figure 67, the ontologist used the annotation property "*translator to Portuguese*" (ontoneo:ONTONEO_1000002) to inform who translated elements to Portuguese.

[66] Details of Information Artifact Ontology are available at

Figure 67: Example 2 of OntONeo using translator annotation property.

```
<!-- http://purl.bioontology.org/OntONeo/ONTONEO_00000067 -->

<owl:Class rdf:about="http://purl.bioontology.org/OntONeo/ONTONEO_00000067">
   <rdfs:subClassOf rdf:resource="&obo;OGMS_0000097"/>

   <rdfs:label xml:lang="en">prenatal encounter</rdfs:label>                  (1)
   <rdfs:label xml:lang="pt">acompanhamento pré-natal</rdfs:label>

   <obo:IAO_0000115 xml:lang="en">A health care encounter throughout pregnancy, in which tests
are performed to monitor the health of the mother and baby.</obo:IAO_0000115>      (2)
   <obo:IAO_0000115 xml:lang="pt">Um encontro de cuidados de saúde que ocorre durante a
gravidez, nos quais são realizados exames e testes para monitorar a saúde tanto da mãe quanto
do bebê.</obo:IAO_0000115>

   <obo:IAO_0000118 xml:lang="en">prenatal care, prenatal monitoring</obo:IAO_0000118>
   <obo:IAO_0000118 xml:lang="pt">cuidado pré-natal, consulta pré-natal, monitoramento pré-   (3)
natal</obo:IAO_0000118>

   <obo:IAO_0000117 xml:lang="en">PERSON: Fernanda Farinelli</obo:IAO_0000117>
   <ontoneo:ONTONEO_1000002 xml:lang="en">PERSON: Fernanda Farinelli</ontoneo:ONTONEO_1000002>
   <rdfs:isDefinedBy rdf:datatype="&xsd;anyURI">
                      http://purl.obolibrary.org/obo/ontoneo/ontoneo.owl</rdfs:isDefinedBy>

</owl:Class>
```

Source: Prepared by the author.

Formalization of object properties:

Although the development of OntoNeo adopts as a basis the OBO Relations Ontology (Smith *et al.*, 2005, Smith; Grenon, 2004, Smith; Rosse, 2004) to determine the relationships between entities, the ontologist has observed the need to create new relations or properties to determine specifics meanings in the OntONeo definitions. For exemplify the first couple of object properties created by OntONeo, consider the gestational age.

Gestational age indicates the pregnancy time, that is, is a measure of the pregnancy age since the beginning of the last menstruation. To estimate a gestational age, one of the methods used by physicians is considering the menstrual history, that is, a gestational age is defined, in weeks or in complete days, from the date of the last menstruation (DUM) of the pregnant woman.

Therefore, once a gestational age is defined calculating the days since the beginning of the last menstrual period if the date of the last menstruation changes the gestational age consequently changes. In order to cover these situations described above, OntONeo ontologist created the object properties labelled “*defines*”, and an inverse relation labelled “*is defined by*” according to definition presented in Figure 68 and Figure 69. Thus, we have the example of both definitions on Figure 67.

Figure 68: Description of “*defines”* object property of OntONeo.

Y *defines* **X** = def. **Y** *defines* **X**, if and only if

1) **Y** is a *quality* or a *temporal region*, and
2) **X** is an *entity* (*continuant* or *occurent*), and
3) the value of **X** *is defined* from the value of **Y**, and
4) a change in the value of **Y** causes a change in the value of **X**.

Source: Prepared by the author.

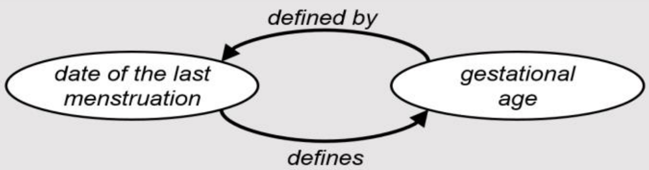
Figure 69: Description of “*defined by”* object property of OntONeo.

X *is defined by* **Y** = def. **X** *is defined by* **Y**, if and only if

1) **X** is an *entity* (*continuant* or *occurent*), and
2) **Y** is a *quality* or a *temporal region*, and
3) the value of **X** *is defined* from the value of **Y**, and
4) a change in the value of **Y** causes a change in the value of **X**.

Source: Prepared by the author.

Figure 70: Example 1 of object property created by OntONeo.



Where:

- “*date of the last menstruation*” is a “*temporal region*” of BFO.
- “*gestational age*” is a "*temporal measurement*" defined by the Experimental Factor Ontology.
- "*temporal measurement*" is "*information content entity*" defined by the Information Artifact Ontology, which is a subtype of "*generically dependent continuant*" of BFO.
- the value of “*gestational age*” is defined from the value of “*date of the last menstruation*”.
- a change in the value of “*date of the last menstruation*” causes a change in the value of “*gestational age*”.

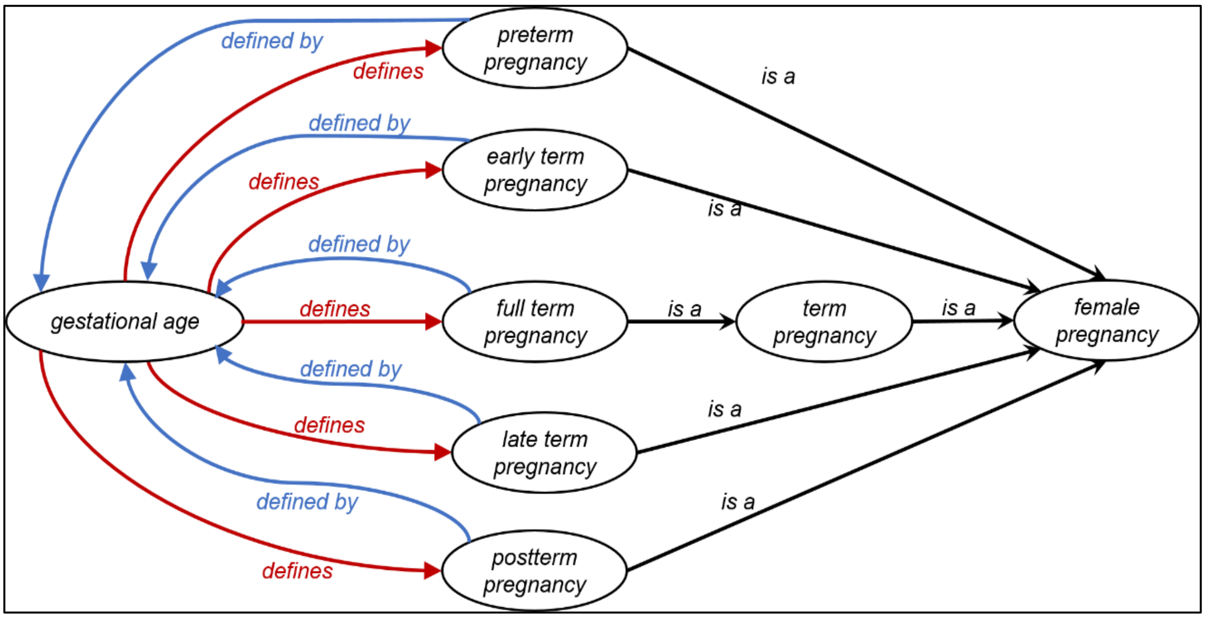
Source: Prepared by the author.

Additionally, concerning the date when the baby was born during the pregnancy, physicians use gestational age to determine preterm, term (early, full or late) and postterm pregnancy, namely:

- A preterm pregnancy is a pregnancy that ends with the birth of a baby before 37 weeks of gestational age.
- An early term pregnancy is a pregnancy that ends with the birth of a baby between 37 weeks through 38 weeks and 6/7 days.
- A full-term pregnancy is a pregnancy that ends with the birth of a baby between 39 weeks through 40 weeks and 6/7 days.
- A late term pregnancy is a pregnancy that ends with the birth of a baby between 41 weeks through 41 weeks and 6/7 days.
- A postterm pregnancy is a pregnancy that extends to 42 weeks of gestational age and beyond.

Thus, we have the definition showed on Figure 71.

Figure 71: Example 2 of object property created by OntONeo



Source: Prepared by the author.

Another situation that demanded the creation of new object property refers to the relationship between an injury and an injured organism. One can argue that an injury is inhered in some organism, or an organism is a bearer of some injury. However, such a relationship refers to the idea of belonging or possession. That is, if an injury exists, such injury belongs to an organism because the injury occurred in the referred organism. That is, if an organism suffers an adverse situation that generates an injury in the organism because of this situation, such injury is an injury of this organism.

OntONeo ontologist analyzed some object properties found in OBO Foundry ontologies as candidates to define the relationship between injury and organism. Such object properties "*inheres in*" and "*bearer of*", displayed in Figure 72 and Figure 73 respectively.

Figure 72: Description of “*inheres in”* object property of RO.

Object Property: *inheres in*

Term IRI: *http://purl.obolibrary.org/obo/RO_0000052*

Definition: A relation between a *specifically dependent continuant* (the dependent) and an *independent continuant* (the bearer), in which the dependent specifically depends on the bearer for its existence.

Source: Prepared by the author based on Ontobee (2017).

Figure 73: Description of “*bearer of*” object property of RO.

Object Property: *bearer of*

Term IRI: *http://purl.obolibrary.org/obo/RO_0000053*

Definition: A relation between an *independent continuant* (the bearer) and a *specifically dependent continuant* (the dependent), in which the dependent specifically depends on the bearer for its existence.

Source: Prepared by the author based on Ontobee (2017).

The ontologist considered none of these properties adequate for OntONeo's needs. An injury is an independent continuant defined by the Ontology for General Medical Science. Just as the term organism, defined by the Ontology for Biomedical Investigations, is an independent continuant. Thus, considering the definitions of these two properties (Figure 72 and Figure 73), they are not used to establish the relationship between two independent continuants (the injury and the injured organism).

In order to establish a relationship between the injury and the injured organism, the OntONeo ontologist created a new object property labelled "injury of" described in Figure 74, and its inverse property "*has injury*".

Figure 74: Description of “*injury of”* object property of OntONeo.

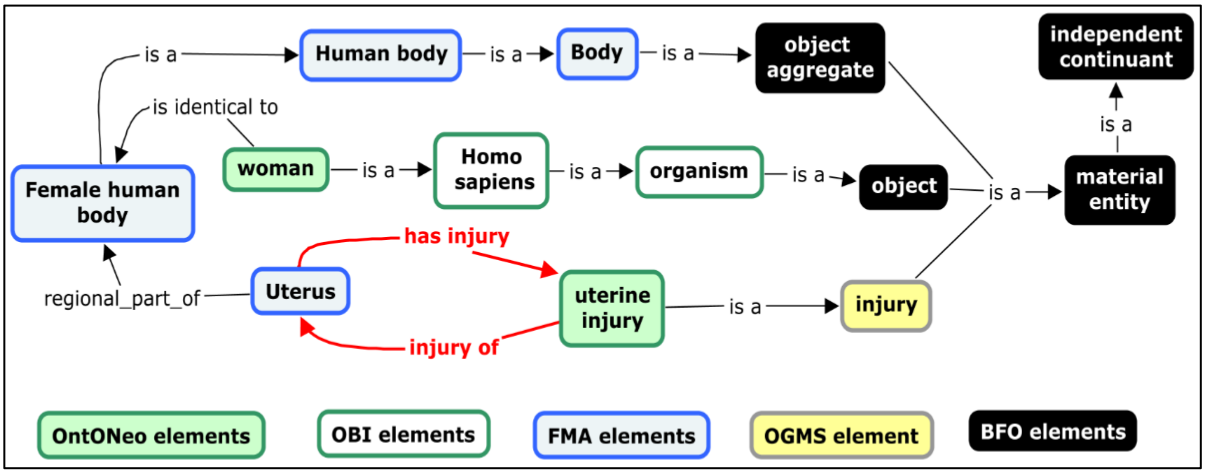
X *injury of* **Y** = def. **X** is an *injury of* **Y**, if and only if

1) **Y** is an *Organism, Body, Organ* or *“Organ system”*, and
2) **X** is an *injury, and*
3) **X** is a structural damage undergone by **Y** in such way that **X** exists only as *part of* **Y.**

Source: Prepared by the author.

Thus, Figure 75 and Figure 76 present examples of the use of "*injury of*" and "*has injury*" object properties by OntOneo.

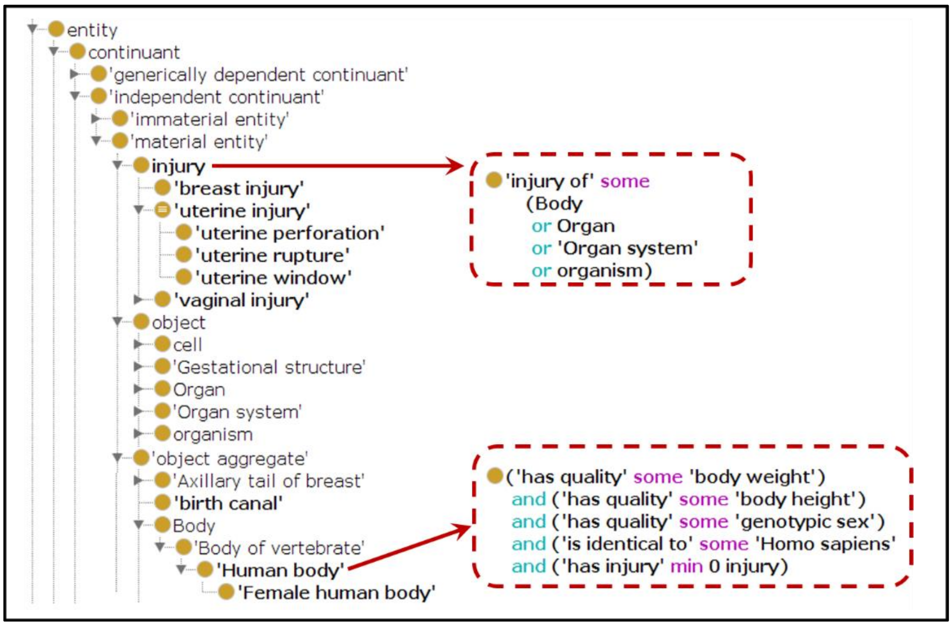
Figure 75: Example 3 of object property created by OntONeo.



Source: Prepared by the author.

According to Figure 75, a *uterine injury* is an *injury of* some *Uterus* in which is part of some *Female human body* or some *woman*. As shown in Figure 76, if an *injury* exists, it exists in some *Body*, *Organ*, *Organ system* or *organism*, on the other hand, a *Human body* could not have an *injury* (min 0).

Figure 76: Example 4 of object property created by OntONeo.

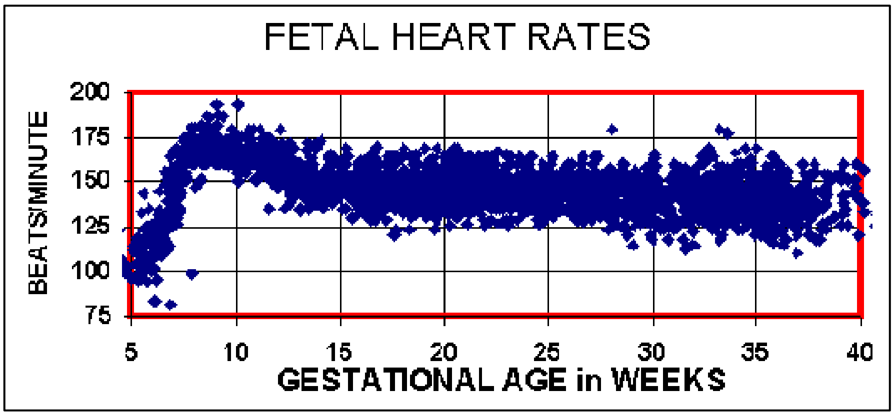


Source: Prepared by the author.

Finally, the last object property defined in OntONeo until now is the property labelled "*measure varies according to*". This property intends to establish the relationship between a measurement and another entity that causes a change in such measurement.

For instance, during prenatal follow-up, obstetricians measure the fetal heart rate in order to verify some possible related pathology. A normal fetal heart rate (FHR) usually ranges from 120 to 160 beats per minute (bpm). However, FHR varies along gestation, increasing until the 10th week of gestation and decreasing from the 10th week of gestation until term. Thus, as exemplified in Figure 77, a *fetal heart rate* measure varies according to an increase of the *gestational age*.

Figure 77: Example of fetal heart rates measurement against gestational age



Source: Retrieved from Dubose (2012).

Therefore, OntONeo ontologist defined "*measure varies according to*" object property as presented in Figure 78.

Figure 78: Description of "*measure varies according to"* object property of OntONeo.

X *measure varies according to* **Y** = def. **X** *i measure varies according to* **Y**, if and only if

1) **X** is a *quality* that denotes a measured value such a *rate* or *duration*, and
2) **Y** is a *process* or a *time region*, and
3) the value of **X** could change, increasing or decreasing, by some variation of **Y**.

Source: Prepared by the author.

OntONeo classes and relations

OntONeo is a domain ontology to cover knowledge of the obstetric and neonatal domain concerned the woman's care in pre-pregnancy, prenatal, parturition, puerperal (postnatal) and newborn phases are part of this domain of knowledge. In this way, any terms that the ontology covers relate in some ways to the core entities of this domain, which are a woman, pregnancy, and her baby (Embryo or Fetus stage).

In the following, we describe how the ontology deals with such core elements (Figure 79). OntONeo ontologist defined both the terms *woman* and *pregnant woman*, and the relationship between classes designated by "*has quality*", "*participates in*" and "*is identical to*" labels.

According showed in Figure 79, to define the class *woman*, OntONeo reuses a hierarchy of Ontology for Biomedical Investigations (OBI)[67] defined to deal with the term *Homo sapiens*. Note that *Homo sapiens* is a class of the NCBI Taxonomy, and *organism* is a class of OBI. *Gestational structure*, *Fetal structure*, *Embryonic structure*, *Fetus*, *Embryo, Body*, *Human body*, *Female human body* are classes of the ontology of Foundational Model of Anatomy (FMA)[68]. We also imported from Gene Ontology (GO)[69] the classes *biological_process*, *reproductive process*, and *female pregnancy*. *Physical object quality*, *organismal quality, biological sex, genotypic sex,* and *female genotypic sex* are classes of the Phenotypic quality ontology (PATO)[70].

Additionally, everything around OntONeo, originated in a healthcare encounter around pre-pregnancy, pregnancy, and post- pregnancy. Thus, OntONeo deals with these terms according to Figure 80, Figure 81 and Figure 82 below.

Some terms OntONeo imported from Ontology for General Medical Science (OGMS)[71] such as *diagnostic process, health care process, health care encounter, physical examination, inpatient encounter, outpatient encounter*, and *hospitalization*. The terms *physician role* and *human patient role*, OntONeo imported from Ontology of Medically Related Social Entities (OMRSE)[72]. Finally, OntONeo imported from Ontologized MIABIS (OMIABIS)[73] the term *medical record*.

---

[67] Available at http://obi-ontology.org/

[68] Available at http://si.washington.edu/projects/fma

[69] Available at http://www.geneontology.org/

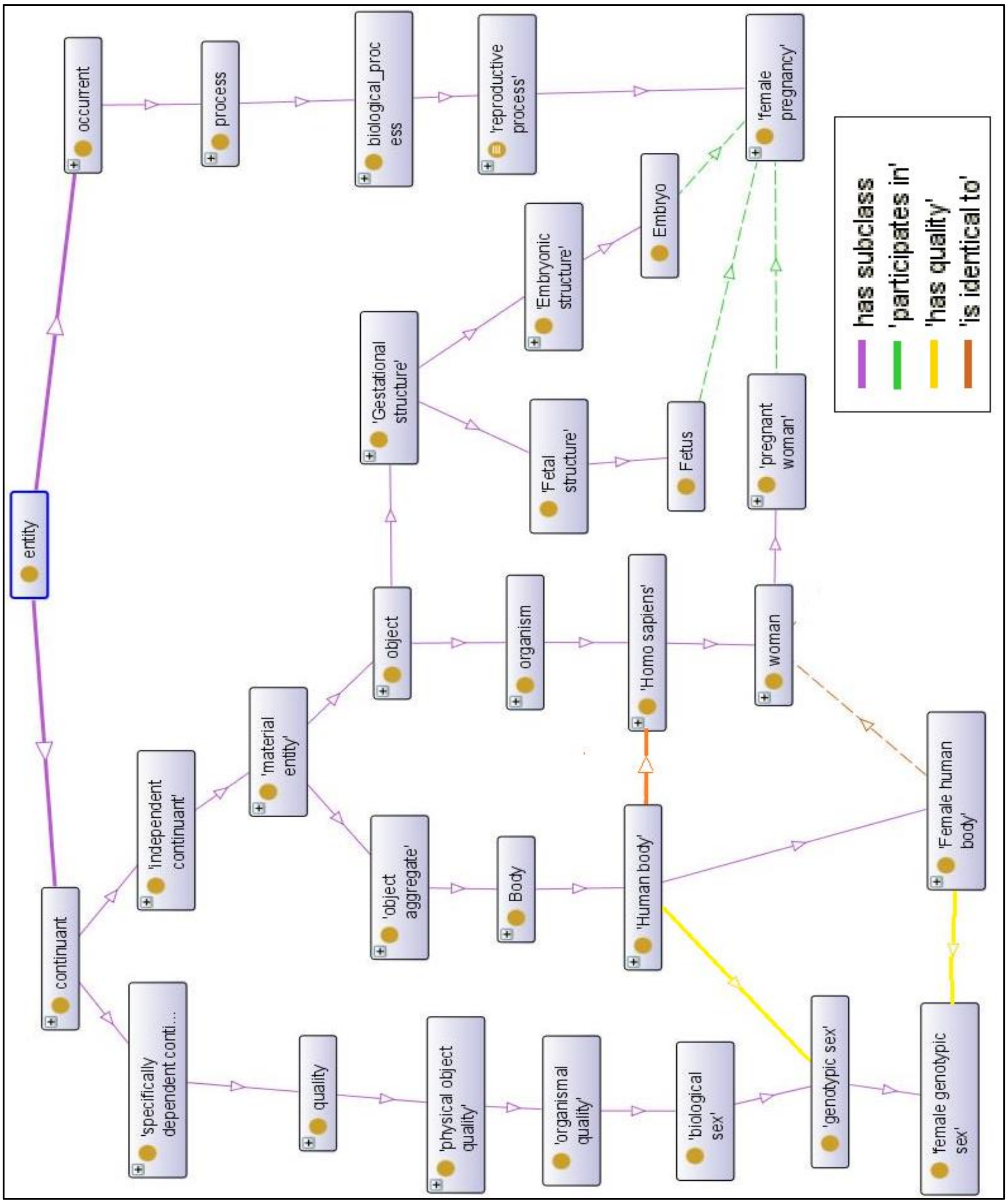
[70] Available at https://github.com/pato-ontology/pato/

[71] Available at https://github.com/OGMS

[72] Available at https://github.com/ufbmi/omrse

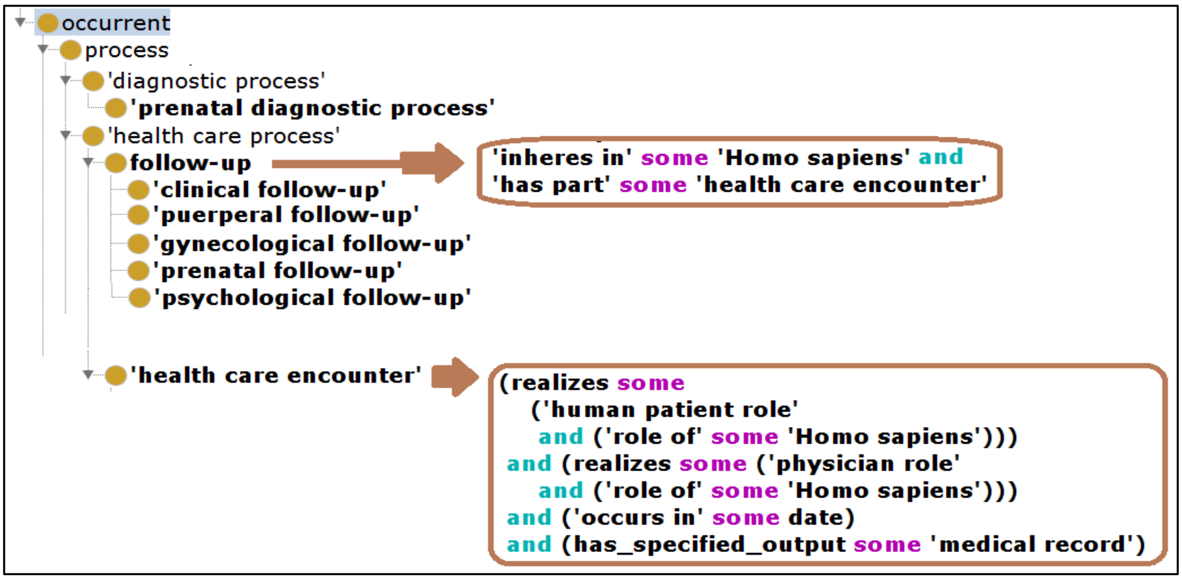
[73] Available at https://github.com/OMIABIS/omiabis-dev

Figure 79: Main representational units OntONeo ontology.



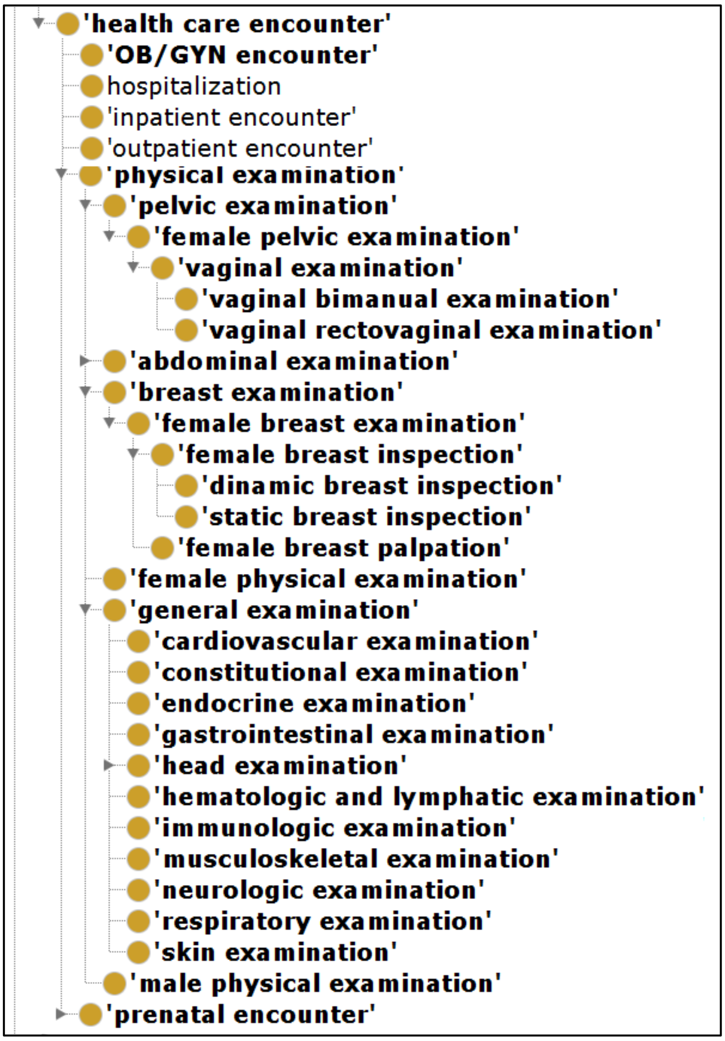
Source: Prepared by the author.

Figure 80: Example 1 of healthcare process in OntONeo ontology.



Source: Prepared by the author.

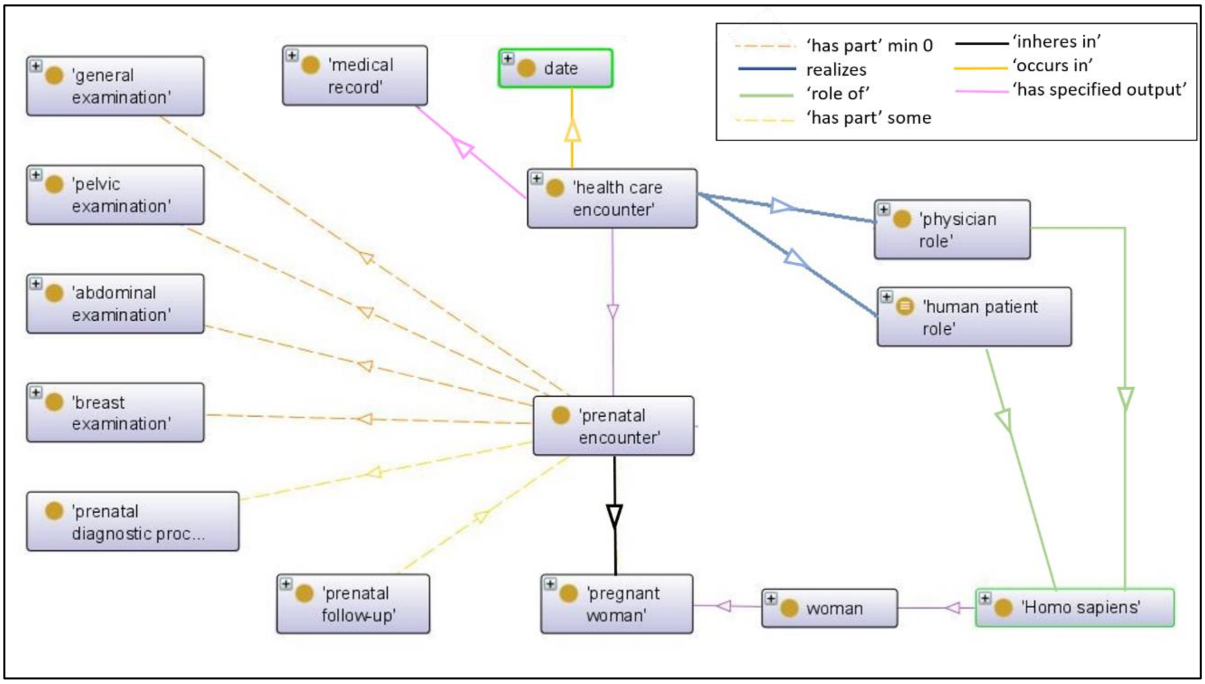
Figure 81: Subtypes of health care encounter used by OntONeo ontology



Source: Prepared by the author.

The ontologist defined the remaining terms in the OntONeo ontology. Such terms are *OB/GYN encounter*, *prenatal encounter*, *date*, *follow-up* and all its subtypes, and all subtypes of *physical examination* displayed in Figure 80 and Figure 81. Finally, the ontologist also defined new relationships between OntONeo terms and imported terms, and between imported terms, necessary to express the knowledge domain of OntONeo. For example, such relationships defined to the term *health care encounter* displayed in Figure 82.

Figure 82: Example 2 of healthcare process in OntONeo ontology.



Source: Prepared by the author.

Ontology annotation:

ReBORM defined a specific activity to deal with annotations, presenting that each time that we annotate a resource an instance of *annotation activity* is established.

For example, during the Protégé environment configuration (beginner of the first iteration of implementation phase), we use several annotations in order to maintain the ontology documented. We annotated, in ontology level, some relevant elements in order to maintain the ontology documentation, namely: *creator*, *contributors*, *license*, *homepage*, *mailbox*, *current version information*, *default language*, *alternative language*, and *sources* of knowledge used.

In fact, OntONeo resorted to annotations defined by several others ontological resources, and such OntONeo annotations properties presented in subsection 4.3.4.

Ontology evaluation:

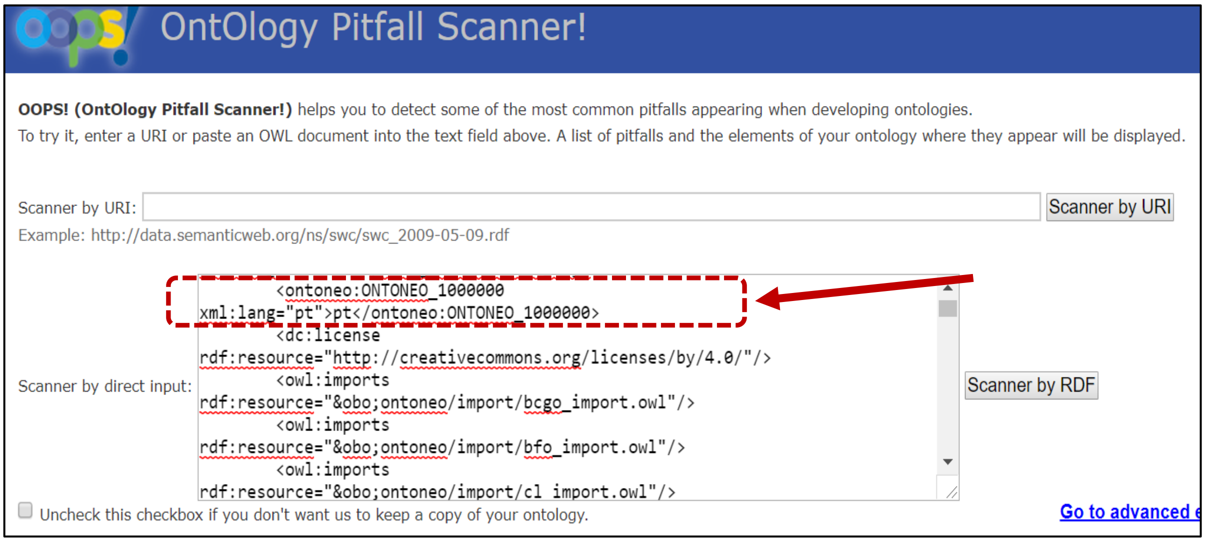
At the end of each implementation phase iteration, we submitted OntONeo to OOPS! evaluation. Figure 83 presents an example of the OOPS! evaluation results. We verifies each result reported by OOPS!. According to these results, we verify that OOPS! uses different references than OntONeo developing, for example, the results for pitfall 22 about naming conventions. OntONeo use as naming convention reference the same reference of OBO Foundry ontologies, which is Schober *et al.* (2009) instead of Noy and Mcguinness (2001). Notice that some OOPS! alerts, such the naming conventions, appears in each evaluation submission, and since we detected that it not apply to OntONeo we ignored it on subsequent evaluations.

Figure 83: OOPS! evaluation of OntONeo ontology.

Source: Prepared by the author.

We also verify some wrong analyses of OOPS!, which is the case of result for pitfall 41 reporting an omission of the license information of OntONeo ontology. OntONeo used some metadata terms recommended by DCMI (Dublin Core Metadata Initiative), for instance, the license metadata (http://purl.org/dc/elements/1.1/license), according to highlighted in Figure 84. Maybe, the license metadata verified by OOPS! is different from the license metadata used in OntONeo. We reported this observation to OOPS! maintain time.

Figure 84: Evidence of license metadata of OntONeo ontology.



Source: Prepared by the author.

In this way, OntONeo ontologist evaluated each pitfall reported prioritizing the main ontology building references used, the methodology of ontological realism and OBO Foundry principals, instead of *OOPS!* references. Another point considered is the fact of we just acted to solve pitfalls found in OntONeo ontology, we did not consider solving or acting against pitfalls found in reusing ontologies.

4.3.5 Delivery phase of OntONeo development

OntONeo was delivered by publishing such ontology in both OBO Foundry Portal[74] and BioPortal[75] in addition to OntONeo Project webpage and OntONeo's GitHub[76] repository. Appendices 8 and 9 showed details of how we published OntONeo in such both ontologies

[74] OntONeo OBO Foundry URL address: http://www.obofoundry.org/ontology/ontoneo.html

[75] OntONeo BioPortal URL address: https://bioportal.bioontology.org/ontologies/ONTONEO.

[76] OntONeo GitHub repository address: https://github.com/ontoneo-project/Ontoneo

portals. All iterations of the delivery phase during OntONeo development followed the activities workflow presented in Figure 47 (subsection 4.2.5).

4.4 Implementation of OBO Foundry principles

In following paragraphs, we highlight each OBO Foundry principle using bold and italic text format. In order to understand how to register your ontology on OBO Foundry, read the guide on Appendix 8.

First, it is relevant to highlight that principles 10 and 13, respectively ***commitment to collaboration*** and ***maintenance***, are challenging to present concrete evidence.

We can underline the commitment of the OntOneo Project to ensure that the ontology should reflect the scientific knowledge of its domain scope and the knowledge evolution. As OntONeo is an ontology still in development, wherein each iteration, will represent new elements enriching its possibilities of use. In addition, the OntOneo Project provides in its repository a feature to submit maintenance requests[77], both for new terms and for correction of problems. The project also provides an open mailing list[78] that a community of both ontology developers and the general public can submit suggestions, new demands and start any discussions related to OntONeo scope.

Therefore, until the moment of this thesis presentation, we can say that OntOneo development was carried out in a collaborative way just by exchange of opinions and knowledge with OBO Foundry ontologists. Some of this exchanges happened via mailing list of OntONeo, however, the OntONeo developer acquired the majority of contributions in person.

Following the ***OBO Foundry principle*** of the ontology must be ***open*** (principle 1); the license of OntONeo is copyrighted under a Creative Commons CC-BY license version 4.0. For evidence of this recommendation, we present Figure 85 with a piece of OntONeo RDF-XML code showing the license annotations.

Additionally, giving appropriated credits to others ontologies in reuse, OntONeo ontologist specified for each imported isolated component the annotation property "*imported from*[79]" from the Information Artifact Ontology. The value attributed to this annotation refers the URI/IRI of the ontology from which the component was imported. In the case of using several

[77] OntONeo issue tracker available at: https://github.com/ontoneo-project/Ontoneo/issues

[78] OntONeo project discussing group available at http://groups.google.com/group/ontoneo-discuss

[79] *Imported from* annotation property IRI is http://purl.obolibrary.org/obo/IAO_0000412

classes of an ontology, OntONeo used the approach of MIREOT and imported the extracted file of the ontology in reuse, whole or part.

Figure 85: Implementation evidence of OBO Foundry Principle 1 to OntONeo.

```
<owl:Ontology rdf:about="http://purl.obolibrary.org/obo/ontoneo/ontoneo.owl">
  ...
 <dc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/"/>
 <rdfs:comment xml:lang="en">OntoNeo, is developed by Fernanda Farinelli and
   her contributor group, is licensed under CC BY 4.0.</rdfs:comment>
  ...
</owl:Ontology>
```

Source: Prepared by the author.

OntONeo is adherent to the OBO principle of ***common format*** (principle 2) once that is available in a common formal language, OWL 2, in an accepted concrete syntax, which is RDF/XML (see http://purl.obolibrary.org/obo/ontoneo.owl).

During the Ontology Architecture Specification activity (see subsection 4.3.2), the OntONeo developer created the default namespace for OntONeo located under the OBO PURL domain. Moreover, to reach the OBO principle of the ***URI/Identifier Space*** (principle 3), the developer established a unique URI identifier to each element created during OntONeo formalization following the OntONeo PREFIX and its ID Space patterns described in subsection 4.3.3. See an example in Figure 87.

In order to reach ***versioning*** principle (principle 4), we provided an IRI versioning and some annotation properties to highlight each specific version on OntONeo. Note that at the end of each iteration a new version of OntONeo was delivered. There are some annotation properties to deal with versioning in OWL, they are *priorVersion*, *versionInfo*, and *versionIRI* (W3c, 2004b). Figure 86 shows an evidence of OntONeo versioning mechanisms.

With regard to the ontology ***scope*** (principle 5), as presented in the ontology requirements specification document (Appendix 3), OntONeo had its scope clearly delimited considering the clinical records of obstetric and neonatal care.

Figure 86: Implementation evidence of OBO Foundry Principle 4 to OntONeo.

```
<owl:Ontology rdf:about="&obo;ontoneo/ontoneo.owl">
  …
 <owl:versionIRI rdf:resource="&obo;ontoneo/releases/2017-03-04/ontoneo.owl"/>
 <owl:versionInfo rdf:datatype="&xsd;dateTime">2017-03-04</owl:versionInfo>
 <owl:versionInfo rdf:datatype="&xsd;string">v1.4</owl:versionInfo>
 <owl:priorVersion rdf:datatype="&xsd;string">v1.3</owl:priorVersion>
  …
</owl:Ontology>
```

Source: Prepared by the author.

The principle of ***textual definition*** (principle 6) of OntONeo is partially reaching. OntONeo used the annotation properties labelled of "*definition*"[80] from the Information Artifact Ontology to store these definitions (Figure 87). Currently, OntONeo has 1343 classes of which 486 have no textual definition (about 30%). However, some these 1343 classes are ontological resources in reuse. Several of these reusable resources do not have a textual definition, which influences directly the metric "classes with no definition" of OntONeo.

Figure 87: Implementation evidence of OBO Foundry Principles 3, 6 and 12 on OntONeo.

```
<!-- http://purl.obolibrary.org/obo/ontoneo/ONTONEO_00000157 -->

<owl:Class rdf:about="http://purl.obolibrary.org/obo/ontoneo/ONTONEO_00000157">
   <rdfs:label xml:lang="en">woman</rdfs:label>
   <rdfs:label xml:lang="pt">mulher</rdfs:label>
   <rdfs:subClassOf rdf:resource="http://purl.obolibrary.org/obo/NCBITaxon_9606"/>
   <obo:IAO_0000115 xml:lang="en">A human being that is biologically consider
      female because has the genotype female.</obo:IAO_0000115>
   <obo:IAO_0000115 xml:lang="pt">Um ser humano que é considerado biologicamente
      feminino pois possui o genótipo feminino.</obo:IAO_0000115>
   <obo:IAO_0000117 xml:lang="en">PERSON: Fernanda Farinelli</obo:IAO_0000117>
   <ontoneo:ONTONEO_1000002 xml:lang="en">
      PERSON: Fernanda Farinelli</ontoneo:ONTONEO_1000002>
   <rdfs:isDefinedBy rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI">
      http://purl.obolibrary.org/obo/ontoneo/ontoneo.owl</rdfs:isDefinedBy>
</owl:Class>
```

Source: Prepared by the author.

OntONeo privileged the use of relations defined in Relation Ontology to fit the OBO principle of ***relations*** (principle 7). We consider OntONeo a well-documented ontology as expected by OBO principles ***documentation*** (principle 8) and ***documented plurality of users*** (principle 9). All OntONeo development decisions are available on OntONeo Homepage or OntONeo repository. The ontologist annotated the ontology with several relevant information,

---

[80] *Definition* annotation property IRI is http://purl.obolibrary.org/obo/IAO_0000115

for instance, the state of development ontology, scope and purpose of this ontology, project homepage, contact email, ontology responsible, creator and contributors, main knowledge sources and resources. In addition, these cited annotation shows an evidence of the ***locus of authority*** principle (principle 11) by indicating the responsible or communications between the community and the ontology developers using the contact email.

OntONeo follows the principle of ***naming conventions*** (principle 12) described in Schober *et al.* (2009), which are also according to terminology recommendations of the ontological realism (see subsection 2.4.5). Each OntONeo element created include one single label express in English, additionally, in order to attend the non-functional requirements, OntONeo has this single label express also in Portuguese, which is defined using the element *label*[81] from RDF schema (Figure 87). In both cases, English and Portuguese, we wrote everything using plain language. Abbreviations and synonyms were included using the annotation property labelled *alternative term*[82] defined by Information Artifact Ontology.

4.5 Proof of concept: retrieving data with OntoNeo

This section describes the proof of concept (PoC) in order to validate the ontology against interoperability possibility. To this PoC we consider the follow scenario:

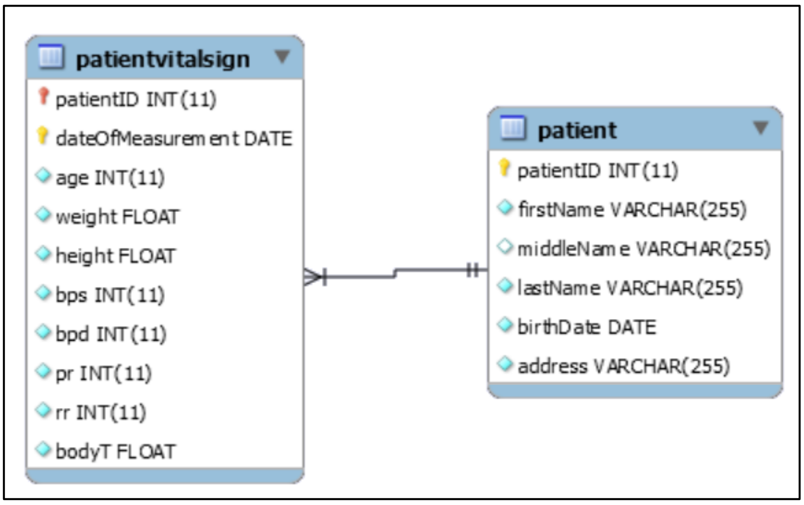
> "A patient can participate in several medical encounters. In each encounter, the physician checks the weight, height, vital signs, and identifies the full name, date of birth, age at the encounter date, and the patient's address. The vital signs of the patient are body temperature, pulse rate, respiratory rate, and blood pressure (systolic and diastolic). All data gathered are stored in a database."

The proof of concept environment consisted of a relational database schema implemented in a local MySQL DBMS using the database model of the Figure 88, which represents the situation described above. According to this model, a patient has multiples vital signs measurement. The database structure consists of two tables called *patient* and *patientVitalSign*. Figure 89 and Figure 90 presents, respectively, the data stored in table *Patient* and *PatientVitalSign*.

[81] *Label* annotation property IRI is https://www.w3.org/2000/01/rdf-schema#label

[82] *Alternative term* annotation property IRI is http://purl.obolibrary.org/obo/IAO_0000118

Figure 88: Database schema of the proof of concept.



Source: Prepared by the author.

The *patient* table, displayed in Figure 89, contains the identification data of a patient and includes the columns *firstName*, *middleName*, *lastName*, *birthDate* and *address*. The primary key (PK) of *patient* table is the column *patientID* that is a sequential number.

Figure 89: Proof of concept - Data of patient table.

patientID	firstName	middleName	lastName	birthDate	address
1	Francine		Loyd.	1973-03-29	206 Affinity Lane, apt. D, Cheektowaga, NY
2	Valery		Brunette	1986-03-27	77 Goodell Street, Buffalo, NY
3	Janete	Earl	Smith	1988-11-15	11 Plymouth Avenue, Buffalo, NY
4	Carol	Josef	Klein	1987-05-01	37 Kemp Avenue, Cheektowaga, NY
5	Carol	Josef	Klein	1987-05-01	37 Kemp Avenue, Cheektowaga, NY

Source: Prepared by the author.

The second table (Figure 90), *patientVitalSign*, contains the vital signs that are measurement during a medical encounter such as systolic blood pressure and diastolic blood pressure (columns table *bps* and *bpd* respectively), pulse rate (column table *pr*), respiratory rate (column table *rr*) and body temperature (column table *bodyT*).

Figure 90: Proof of concept – Data of Patientvitalsign table.

patientID	dateOfMeasurement	age	weight	height	bps	bpd	pr	rr	bodyT
1	2016-05-28	30	64	1.56	100	75	88	15	97
2	2016-11-21	28	79	1.69	110	85	85	13	97.1
3	2016-07-12	29	85	1.7	120	90	95	18	96.9
3	2016-11-28	31	80	1.68	115	90	90	16	96.9

Source: Prepared by the author.

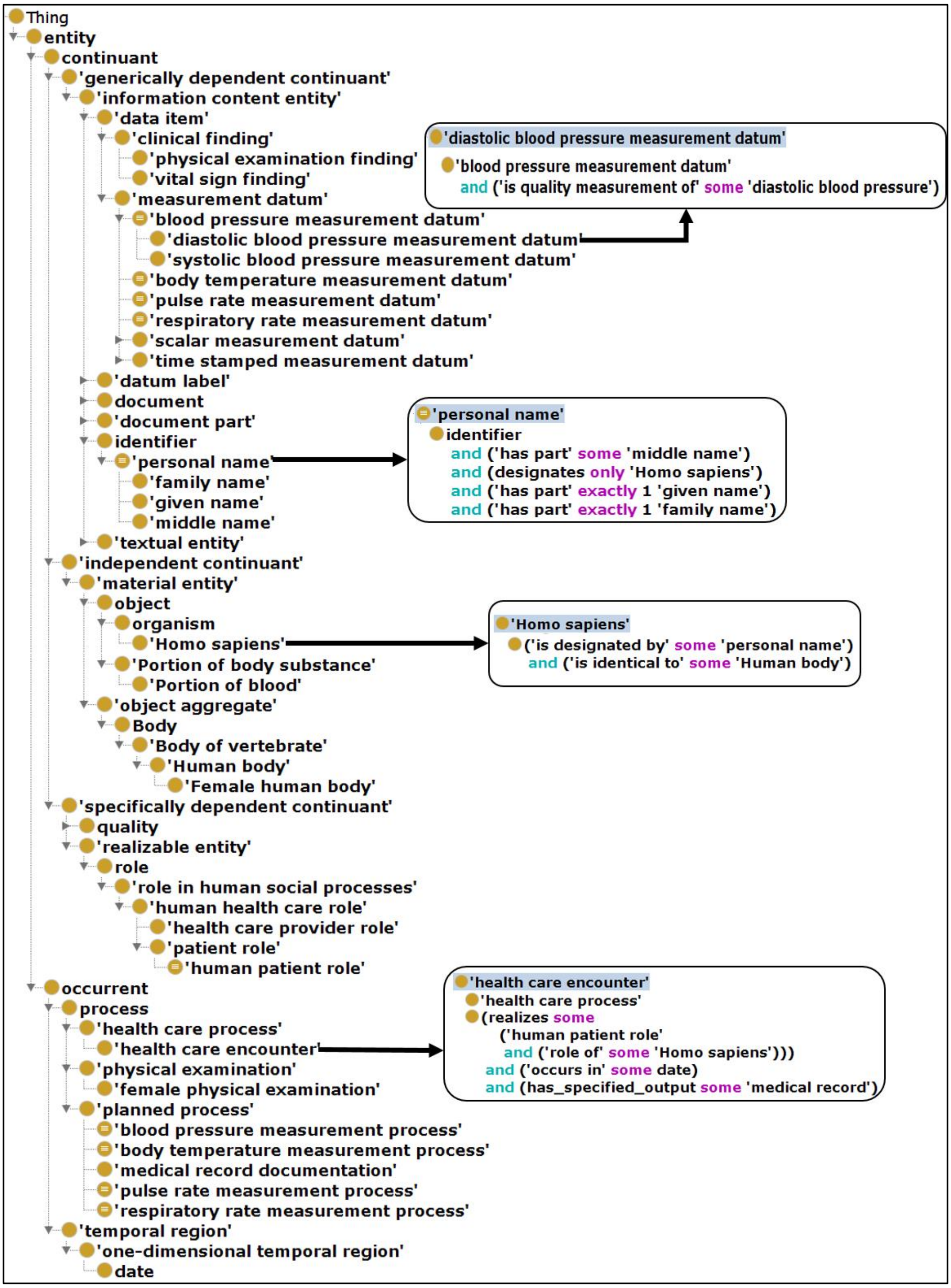
In addition, this table includes the date in which the medical encounter happens and the vitals were measured (column table *dateOfMeasurement*) and also some human body constitutional data such as body weight and height (columns table *weight* and *height* respectively) and the patient age at the medical encounter day (column table *age*). The column *patientID* is a foreign key of *patientVitalSign* table referring to the relationship between the two tables (*patient* and *patientVitalSign*), which means that a patient can have many vital sign measurements. The columns *patientID* and *dateOfMeasurement* compose the primary key of the *patientVitalSign* table.

We detached a part out of OntONeo including the set of terms that corresponds to elements of the database schema according to Figure 91. The left side of figure presents a hierarchic part of OntoNeo taxonomy, while the right-side displays some balloon details that exemplify the ontology logical definitions for some terms. These logical definitions were highlighted because are important to understanding the mapping between the database and the ontology described in next subsection.

After the creation of the PoC environment, we performed the both mapping. Table 17 presents some examples of the first mapping (relating database metadata to ontology elements) created according to the rules of Table 9 (subsection 3.3.5), and Source: Prepared by the author.

Table 18 presents some examples of the second mapping (defining ontology individuals from database data) created according to the rules of Table 10 (subsection 3.3.5).

Figure 91 : Piece of ontology used in the proof of concept



Source: Prepared by the author.

Table 17 - Examples of cases of first mappings

	Examples of first mapping
Case 1	The *patientVitalSign* table was mapped to the *vital sign* class
Case 2	The *patient* table was mapped according to the following structure: '*inheres in*' some ('Homo sapiens' and ('*participates in'* some 'health care encounter') and ('*has role'* some 'human patient role'))
Case 3	The *firstName* column of table *patient* was mapped to the *given name* class
Case 4	The *patient* table PK is the column *patientID* and mapped to the *patient identifier* class
Case 5	The *patientVitalSign* table PK corresponds to the *dateOfMeasurement* and *patientID* columns, denoting that some process of vital sign measurement happened. The mapping made is represented by the following structure: ''vital sign measurement process'' *'has temporal occupant*' some date and *'has participant*' some ('Homo sapiens' and ('*has role'* some 'human patient role'))
Case 6	The *patientVitalSign* table FK denotes a patient *vital sign* and the mapping corresponds to the following structure: '*is quality of*' some ('Homo sapiens' and ('*participates in'* some 'health care encounter') and ('*has role'* some 'human patient role'))

Source: Prepared by the author.

Table 18 - Examples of cases of second mappings

	Examples of second mappings
Case 7	Considering the patient and patientVitalSign tables and their metadata, each table line denotes that there are instances of a class mapped by the table.
Case 8	The value "Valery" of *firstName* column at line 2 of *patient* table is an *instance_of* some *given name* class

Source: Prepared by the author.

Figure 92 presents an example of the turtle file generated in the first mapping, while Source: Prepared by the author.

Figure 93 presents part of the file containing the triples for each individual, displaying three columns corresponding to the elements of the new triple, respectively: subject, predicate and object.

Figure 92: Example of metadata RDF turtle file.

```
@base <http://purl.obolibrary.org/obo/poc/PoC.owl> .
@prefix : <http://purl.obolibrary.org/obo/poc/PoC.owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
...
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix obo: <http://purl.obolibrary.org/obo/> .
@prefix poc: <http://purl.obolibrary.org/obo/poc/> .

<http://purl.obolibrary.org/obo/poc/PoC.owl> rdf:type owl:Ontology .
<jdbc:mysql://localhost:3306/poc> rdf:type <http://purl.org/dc/dcmitype/Dataset> .

patientVitalSign            subtype_of    Class .
patientVitalSign            subtype_of    "vital signs section" .
patient.firstName           subtype_of    Class .
patient.firstName           subtype_of    "given name" .
patientVitalSign.weight     subtype_of    Class .
patientVitalSign.height     subtype_of    Class .
patientVitalSign.weight     subtype_of    "body weight measurement data" .
patientVitalSign.height     subtype_of    "body height measurement data" .
```

Source: Prepared by the author.

Figure 93: Fragment of second mapping turtle file output with individuals

```
patientVitalSign.line1  subtype_of    owl:NamedIndividual .
patientVitalSign.line1  instance_of   "vital signs section" .
"Francine"              subtype_of    owl:NamedIndividual .
"Francine"              instance_of   "given name" .
"64"                    subtype_of    owl:NamedIndividual .
"64"                    instance_of   "body weight measurement data"
"patient.line1"         subtype_of    owl:NamedIndividual .
"patient.line1"         instance_of   "Homo sapiens" .
```

Source: Prepared by the author.

For the sake of simplicity, we use labels instead of IRIs for Figure 92 and Source: Prepared by the author.

Figure 93, however, during this mapping, we defined IRI to each universal or instance found in the database. These individuals compose the “subject” of the triple.

Source: Prepared by the author.

Figure 93, one can observe that the mapping of the column *firstName* from table *patient* corresponds to the ontology class “*given name*” (*patient.firstName subtype_of* “*given name*”). Thus, the mapping makes the correspondence of the value of the column *firstName* to the first line of *patient* table (value “Francine”), in which “Francine” is a type of individual (“*Francine*” *subtype_of owl:NamedIndividual);* and “Francine” is an instance of the class “*given name*” (“*Francine*” *instance_of* “*given name*”).

Finally, after both mappings, we were ready to perform queries on our dataset. For this, we used the SPARQL Query Tool Twinkle[83]. Figure 94 presents an example in which we retrieved information from the turtle file with class instances. This example recovers the values of ID, first, middle and last name of a patient and joins the three different class instances of name to form the full name of a patient using a function CONCAT, which returns the concatenation of a string sequence.

Figure 94: Example of querying the individual turtle file

Part 1

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX fn: <http://www.w3.org/2005/xpath-functions#>
PREFIX obo: <http://purl.obolibrary.org/obo#>
PREFIX poc: <http://purl.obolibrary.org/obo/poc/>
PREFIX patient: <http://purl.obolibrary.org/obo/NCBITaxon_9606>
PREFIX patientID: <http://purl.bioontology.org/PoC/PoC_00000194>
PREFIX personalName: <http://purl.obolibrary.org/obo/IAO_0020015>
PREFIX designates: <http://purl.obolibrary.org/obo/IAO_0020012>
PREFIX givenName: <http://purl.obolibrary.org/obo/IAO_0020016>
PREFIX secondName: <http://semanticscience.org/resource/SIO_001317>
PREFIX familyName: <http://purl.obolibrary.org/obo/IAO_0020017>
PREFIX has_part: <http://purl.obolibrary.org/obo/BFO_0000051>
```

Part 2

```
SELECT ?ID ?firstName ?middleName ?lastName
      (fn:concat(?firstName,?middleName,?lastName) as ?FullName)
WHERE {
    ?patient rdf:type patient: .
    ?patientID rdf:type patientID: .
    ?patientID designates: ?patient .
    ?patientID rdfs:label ?ID .
    ?personalName rdf:type personalName: .
    ?personalName designates: ?patient .
    ?personalName rdfs:label ?Name .
    ?givenName rdf:type givenName: .
    ?secondName rdf:type secondName: .
    ?familyName rdf:type familyName: .
    ?personalName has_part: ?givenName, ?familyName, ?secondName .
    ?givenName rdfs:label ?firstName .
    ?secondName rdfs:label ?middleName .
    ?familyName rdfs:label ?lastName .
}
ORDER BY ?ID
```

Part 3

ID	firstName	middleName	lastName	fullName
1	Francine		Loyd	Francine Loyd
2	Valery		Brunette	Valery Brunette
3	Janete	Earl	Smith	Janete Earl Smith
4	Carol	Josef	Klein	Carol Josef Klein

Source: Prepared by the author.

[83] Details of this tool at: *http://www.ldodds.com/projects/twinkle/*

The bottom part of Figure 94 (part 3) shows a table with the query results. The part 2. The part 2 of this Figure 94 is the query wrote in SPARQL Language. A basic SPARQL query comprises the clauses SELECT and WHERE. The SELECT clause is the result clause, that is, refers to what data or variables expect to show as a query result. Note that the elements preceded by a question mark (?) are the variables of the query. The WHERE clause is the query pattern, that is, refers to the query restrictions that define the pattern of triples to match against the set of data queried.

Finally, the part 1 of Figure 94 is the PREFIX section. SPARQL language allows the creation of prefixes to replace a specific URI. This practice becomes a query more readable and understandable to human language. For example, according to Figure 94, we define the prefixes *patient:* , *personalName:* and *designates:* , the triples used at clause WHERE were:

```
?patient rdf:type patient: .
?personalName rdf:type personalName: .
?personalName designates: ?patient .
```

Instead of:

```
?patient rdf:type <http://purl.obolibrary.org/obo/NCBITaxon_9606> .
?personalName rdf:type <http://purl.obolibrary.org/obo/IAO_0020015> .
?personalName <http://purl.obolibrary.org/obo/IAO_0020012> ?patient .
```

To conclude, considering the results shown in Figure 94, one can see that it is possible to retrieve data from multiple databases using an ontology as a common vocabulary. Once the structure of one or more databases can be mapped to an existing ontology and the data of one or more databases can be mapped to individuals of ontology, the differences found in database structures can be overcome.

Based on these simple steps, our proof of concept demonstrated the possibility of mapping different databases to the same resource of knowledge representation, which is a formal ontology, and conducting a unified query to retrieve information. In the scope of the already mentioned context of Brazilian network healthcare assistance, this means the possibility to foster communication between distinct information systems and providing progressive improvements in medical assistance to the population.

5 DISCUSSION

“There are no facts, only interpretations”.
Friedrich Nietzsche

This thesis introduces a simple framework within which to be possible the use of ontology to solve semantic interoperability issues across information systems. This chapter summarizes the thesis, discusses its findings and contributions, and points out the challenges found during the investigation. The chapter comprises three sections. Section 5.1 presents a summary of the thesis. Section 5.2 presents a discussion around ontology building methodologies, processes, and activities. Section 5.3 discusses of the obstetric and neonatal ontology presenting some details and metrics of it. Finally, section 5.4 presents a discussion on ontology providing semantic interoperability across information systems considering some architecture possibilities.

5.1 Summary of the Thesis

Chapter 1 presented the motivation and the question of research, which guided the development of this work. We observed a fragmentation of the care information of the obstetric and neonatal patient into various information systems with minimal semantics standardization among them. Such fragmentation with gaps in semantic standardization makes it difficult to exchange data between EHR systems dealing with obstetric and neonatal care data. This exchange of data is necessary to ensure continuity of care for women and newborns, mainly in a distributed environment of care.

Promotion of data exchange among obstetric and neonatal EHR systems depends on a semantic solution able to represent this knowledge domain without ambiguity. At this point, according to this thesis, we claim that formal domain ontologies could help to solve semantic interoperability problems among the terminological heterogeneity. Ontology has its origin in the field of Philosophy as the study of the nature of being, processes of transformation, fact, existence, and reality, as well as the basic categories of being and their relationships. Rapidly, ontology reaches its place as a research theme in Information Science, pervading the use of ontology for knowledge representation and organization. However, we notice a lack of formal and unambiguous representation under the obstetrical and neonatal knowledge, which, for this study, is a necessary condition to promote semantically interoperable solutions for the exchange of EHR data of the patient undergoing on prenatal, childbirth and postnatal care.

Accordingly, this study set out to demonstrate an alternative for semantic interoperability among information systems by using a representation of the obstetrical and neonatal domain of knowledge using ontological realism approach. In view of that, to reach the specific objectives, we defined a set of steps according related in Table 19 below.

Table 19 – The relation between specific objectives and research methodology steps.

Specific objective	Research methodology step
Identify the scope of knowledge of the obstetric and neonatal domain that needs representation.	Study the obstetric and neonatal domain
Identify methodologies to build formal ontology compatible with semantic interoperability issues. Determine the methodology to build the obstetrical and neonatal ontology.	Determine the methodology to build the target ontology
Build a domain ontology covering the knowledge scope of the obstetrical and neonatal domain.	Build the ontology of the obstetric and neonatal domain Verify the OBO Foundry adherence of the ontology built
Demonstrate the use of the ontology built to retrieval data of obstetrical and neonatal records.	Validate the ontology against interoperability possibility

Source: Prepared by the author.

Moving on, chapter 2 presents the main references that supported this investigation, creating a solid theoretical base for the research. In subsection 2.1, it was evident in the literature review the inclusion of knowledge organization as an area of study and specific practice in Information Science. Within this research topic, we identify the organization of knowledge as mechanisms and/or tools of representation and organization of knowledge. We highlighted, among KOS, ontologies as types of schemes of knowledge organization higher levels of complexity.

Then, subsection 2.2 presents basics concepts of interoperability and subsection 2.3 presents some fundamentals concepts of ontology. There is a consensus among several researchers that ontologies are able to provide a shared and common understanding of some domain that can be communicated between people and information systems or computational systems. Different research fields recognize formal domain ontologies as a suitable solution to knowledge representation without ambiguity promoting information retrieval effectively. Hence, ontologies help us to build more efficacious and interoperable information systems in any field.

However, it was not clear technically how ontologies work as an alternative solution for semantic interoperability and how to build ontologies in such way to ensure the absence of ambiguity among ontology terms. Under these circumstances, it became evident that the references on ontology building were spread in diverse sources, from books, journal papers, conference papers, to guides of best practices. Thus, as a potentially contribution of this thesis, section 2.4 compiles set of relevant approaches on ontology building.

Additionally, subsection 2.5 presents some relevant semantic web technologies to understand some specific parts of ontology building, namely, the components of semantic web architecture. As a final point, subsection 2.6 brings together references on informational context in the field of healthcare. Moreover, this subsection examined two relevant documents on recording data of patient's health, namely, electronic health record and electronic discharge summary. It also explored some concepts on biomedical ontologies and terminologies, and the subsection finishes with an analysis of the interoperability relevance between healthcare information systems.

Chapters 3 and 4 presents respectively the research methodology demarcating the methodological steps used to reach the specific goals and the results obtained from these methodological steps. Next,

5.2 Issues on ontology Building

During the research, a systematic literature review showed the existence of several methodologies to build ontologies, without consensus on the best methodology, it being up to the ontologist to choose the most appropriate methodology for his project. From this perspective, the methodology of the ontological realism and NeOn methodology were the most suitable methodologies to OntONeo development. However, each of these separate ontologies did not configure our expectation for ontology building.

As the name of the methodology itself says, it is based on realistic philosophical thinking, implying that a realistic ontology represents entities of reality by itself, and not what people think about reality. This principle corroborates to guarantee the semantic coherence of the ontology, for both humans and computers, avoiding ambiguous representations as well concepts or mental representations of the people. Moreover, the methodology of ontological realism guides the construction of ontology from the point of view of "*how to do*" describing how the ontologist should do (i) to identify the universals and particulars; (ii) define the entities and their respective relations; (iii) organize the entities in hierarchies is_a; (iv) write the definitions for each entity; and so on. Despite a certain orientation to the sequence that the steps of ontological realism methodology, the sequencing of the "*how to do*" does not become a process-oriented structure.

At this point, NeOn methodology assumes its role on ontology building. NeOn methodology take advantages of the ontology engineering best practices providing an extensive guide of "*what to do*" and "*when to do*" during the ontology development.

Consequently, in order to establish a complete methodology to build OntONeo, that defines "*what to do*", "*when to do*" and "*how to do*" we create the methodology labelled **ReBORM** (**Re**alism-**B**ased **O**ntology enginee**R**ing **M**ethodology), a methodology that has been reborn from the harmonization of the methodology of the ontological realism and NeOn methodology. We hope that ReBORM serve as guide to increasing the knowledge and ability of both researches beginners and potential ontologists.

Still on the ontology building, during the development of OntONeo, we faced challenging situations for decision-making in the concept, inception and design phases. Challenges mainly related to the reuse of pre-existing ontological resources in the OBO Foundry Portal.

Given the prerogatives of the OBO Foundry consortium, our expectation was not to find redundancy among the ontologies with respect to the defined entities. Indeed, contrary to our expectations, during the ontology search activity, and consequently in ontology selection activity on Ontobee engine, several times the search results returned more than one ontology that represents a term. That is, a term search returned two or more IRI available, on different ontologies. For instance, searching for the term "*patient*", we found it in at least three different, namely IDOMAL[84], OAE[85] and SIO[86] ontologies, each one defined it specific IRI. We have found several other cases similar to this that we consider as inconsistencies.

Another case was when the term searched existed not as the main label in ontology, but as a synonym or alternative term. As is the case of the term *patient* found in OBI[87] and OMRSE[88] ontologies as an alternative term for the main label "*patient role*". To make matters worse, IDOMAL, OBI, and OMRSE ontologies considering BFO types, defined the patient entity or patient role as continuant entity, specifically dependent, realizable entity, role. OAE and SIO ontologies defined the patient entity as a continuant entity, independent continuant, material entity. Moreover, in the case of the OAE ontology, the definition of the patient class is "*An organism that has the role of patient role*", furthermore, as defined in this ontology, *patient* is equivalent to "*organism and (has role some patient role)*", and "*patient role*" was imported to OAE from OBI. This definition of equivalence is conflicting and inconsistency, since according to OBI the term *patient* is an alternative term for the term *patient role*.

---

[84] patient IRI on IDOMAL: http://purl.obolibrary.org/obo/IDOMAL_0000603

[85] patient IRI on OAE: http://purl.obolibrary.org/obo/OAE_0001817

[86] patient IRI on SIO: http://semanticscience.org/resource/SIO_000393

[87] patient role IRI on OBI: http://purl.obolibrary.org/obo/OBI_0000093

[88] patient role IRI on OMRSE: http://purl.obolibrary.org/obo/OMRSE_00000011

Situations like this have required extra effort to maximize the reuse of existing ontologies. The idea of reuse ontological resources is often understood as an advantage for the development of ontologies; however, the effort expended on analysis is often greater than the effort to define a new term. It may be for this reason that the reuse of ontologies is not yet a common practice among ontologists.

During the ontology search activity, the only selection criteria established to choose an ontology was the general scope of the ontology. This is because the idea in this activity was to identify ontologies that are candidates for reuse and not to select the ontology for reuse itself. Nevertheless, in the case of the ontology selection activity and consequently in the ontology retrieval activity, the establishment of selection criteria was fundamental. This, because at this moment the selection of ontology elements should fit the semantics expected by the ontology in development. Thus, based on the personal experience of the researcher in dealing with data management and data governance, the selection criteria established by OntONeo development were: i) the label of the term fits the ontology requirements; ii) the definition of the term fits the ontology requirements; iii) definition of cross-reference between different ontologies; iv) the most used term among others ontologies; v) the oldest ontology, in which probably defined the term first.

5.3 The obstetric and neonatal ontology

The Ontology of Obstetric and Neonatal domain has been developing with the aim of representing the knowledge concerned the woman’s care in pre-pregnancy, prenatal, parturition, puerperal (postnatal) and newborn phases are part of this domain of knowledge. This ontology comprises terms from several medical disciplines, including at least embryology, anatomy, physiology, gynecology, obstetrics, pediatrics, and neonatology. Its scope includes terms already existing in others ontologies, for instance, image exams, laboratory tests, vaccines, immunizations, disorders, diseases, surgeries, hospitalizations, diagnoses, treatments, demographic data. Moreover, OntONeo not only defines terms related to the obstetric and neonatal domain but also groups and organizes terms of other ontologies representing the knowledge necessary to ensure continuity of care for the woman and her baby.

The structure of OntONeo adheres to the Basic Formal Ontology version 2 and support the reuse of ontologies that are within the scope of OBO Foundry ontologies. The definitions of OntONeo follow both the OBO Foundry principles and the recommendations of the methodology of the realism ontological as far as possible. Next subsections present some details of the main definitions done by OntONeo.

OntONeo is a biomedical domain ontology currently in version number 1.4 with 2367 representational units among classes, object properties, data properties, and annotation properties. Figure 95 presents some metrics of OntONeo generated from Ontobee engine. This ontology development design avoids regeneration of new ontology terms that already existing among OBO Foundry ontologies, thus OntONeo supports efficient ontology reuse, following such OBO Foundry principles. Currently, given the reuse principle, this ontology imports 36 ontologies. OntONeo comprises 1767 are classes in which defines 280 and reuse 1487 classes. In addition, OntONeo defines five object properties and three annotation properties.

OntONeo ontology is an ongoing initiative that seeks to facilitate the exchange of data providing a universal semantic of this knowledge, and improves both the use and analysis of data on its domain to both humans and computers, serving several purposes, for instance:

- as a formal knowledge representation on the obstetric and neonatal domain.
- as the core vocabulary for the development of semantically interoperable systems.
- as a base for computational inferences.
- as a knowledge base for educational purposes.
- as a tool to support decision-making subsidizing clinical practice with information.
- as a conceptual model to information systems against its domain.

Figure 95: Statistics of OntONeo according to Ontobee (retrieval on September/2017).

Index	Ontology Prefix	Class	ObjectProperty	DatatypeProperty	AnnotationProperty	Instance	Total
1	BCGO	0	1	0	0	0	1
2	BFO	35	12	0	0	0	47
3	CL	8	0	0	0	0	8
4	CMO	5	0	0	0	0	5
5	CVDO	4	0	0	0	0	4
6	DOID	215	0	0	0	0	215
7	EFO	7	0	0	0	0	7
8	ERO	5	0	0	0	0	5
9	FMA	404	0	0	0	0	404
10	GO	87	0	0	0	0	87
11	HP	82	0	0	1	0	83
12	IAO	37	16	4	32	0	89
13	IDOMAL	2	0	0	0	0	2
14	MP	40	0	0	0	0	40
15	MPATH	8	0	0	0	0	8
16	NCBITaxon	8	0	0	0	0	8
17	OAE	66	0	0	0	0	66
18	OBI	15	18	2	0	0	35
19	OBO_REL	0	1	0	0	0	1
20	OGMS	31	0	0	0	0	31
21	OMIABIS	3	1	0	0	0	4
22	OMRSE	31	2	1	1	0	35
23	ONTONEO	280	5	0	3	3	291
24	PATO	104	0	0	0	0	104
25	RO	0	353	0	68	0	421
26	SIO	2	9	0	0	0	11
27	SYMP	213	0	0	0	0	213
28	UBERON	38	0	0	0	0	38
29	UBPROP	0	0	0	5	0	5
30	UPHENO	0	1	0	0	0	1
31	VO	17	0	0	0	0	17
32	VSO	18	0	0	0	0	18
33	bao	2	0	0	0	0	2
34	cl	0	6	0	0	0	6
35	doap	0	0	1	0	0	1
36	doid	0	0	0	3	0	3
37	fma	0	7	0	0	0	7
38	hp-logical-definitions-subq	0	1	0	0	0	1
39	oboInOwl	0	0	0	15	0	15
40	owl	0	0	0	1	0	1
41	pato	0	4	0	6	0	10
42	rdf-schema	0	0	0	1	0	1
43	NoPrefix	0	3	0	13	0	16
Total	-	1,767	440	8	149	3	2,367

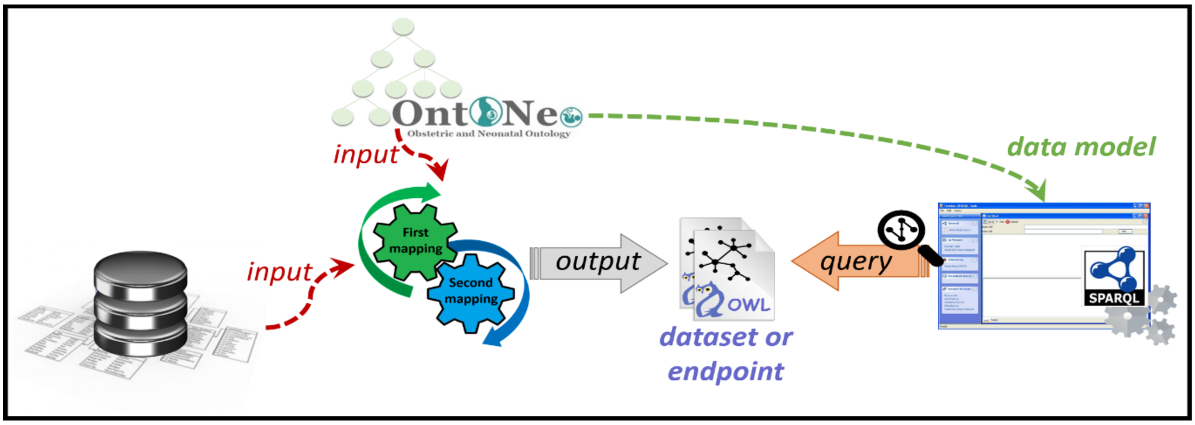
Source: Prepared by the author.

OntONeo was published on both BioPortal and OBO Foundry Portal, in this last with “Library” status. Such ontology implemented the thirteen OBO Foundry principles established by OBO Foundry Consortium as described in subsection 4.3. During the research, we did not find much documentation explaining how to adhere to OBO Foundry principles. Therefore, this thesis' subsection (4.3) contributes with future ontologists understanding how to implement these principles in an ontology development.

5.4 Use of OntONeo in semantic interoperability

There are several applications for ontologies, mainly in biological and medical domains. OntoNeo use is suitable in many of these applications since it has been designed to provide both organization and consensual representation of information within the obstetric and neonatal domain. In subsection 4.4, we describe a basic way to use the OntONeo ontology to query data originated on relational databases.

Figure 96: Proof of concept schema of this thesis.



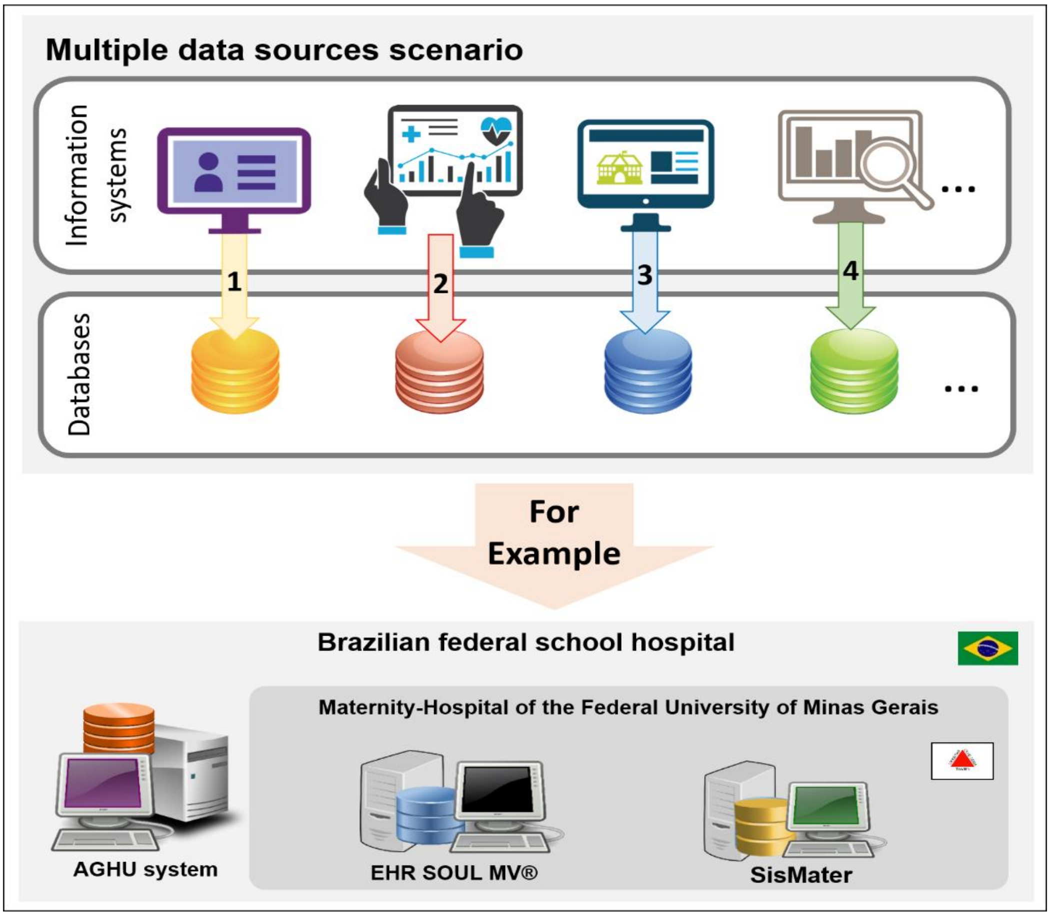
Source: Prepared by the author.

Basically, as shown in Figure 96, given a database (database schema, metadata and data of each table) against a domain of business, and an ontology that represents the domain of knowledge toward such business. After some steps of mapping between database and ontology, a resulting dataset available in OWL language could be queried using a Sparql query language. Although basic, the results obtained in this proof of concept are consistent with our expectations.

This proof of concept configures a foundational knowledge because it focuses on how to map existing databases to an existing formal ontology, without it being necessary changes in such database or such ontology. This is one of the advantages brought by OntONeo, the promotion of data exchange available in different information systems without changes in such systems and their databases. However, if the ontology does not cover some existing knowledge in the database, some maintenance on the ontology is necessary to contemplate such knowledge.

Starting from this basic scenario of this proof of concept, we can consider a scenario with multiple data sources. For instance, a situation that exists four different information systems, each one with its specific database. Such database could be implemented using the same database management system or not. In addition, they could follow different data models. For example, consider the schema of Figure 97 below, such multiple data sources could be, for instance, the sources of data observed in the operation of the Maternity Hospital from the Federal University of Minas Gerais showed in Figure 1 of chapter 1. In this schema (Figure 97), we worked with different colors to promote a better understanding of what this figure shows. We follow this same pattern of colors in subsequent figures.

Figure 97: Example of scenario with multiple data sources.

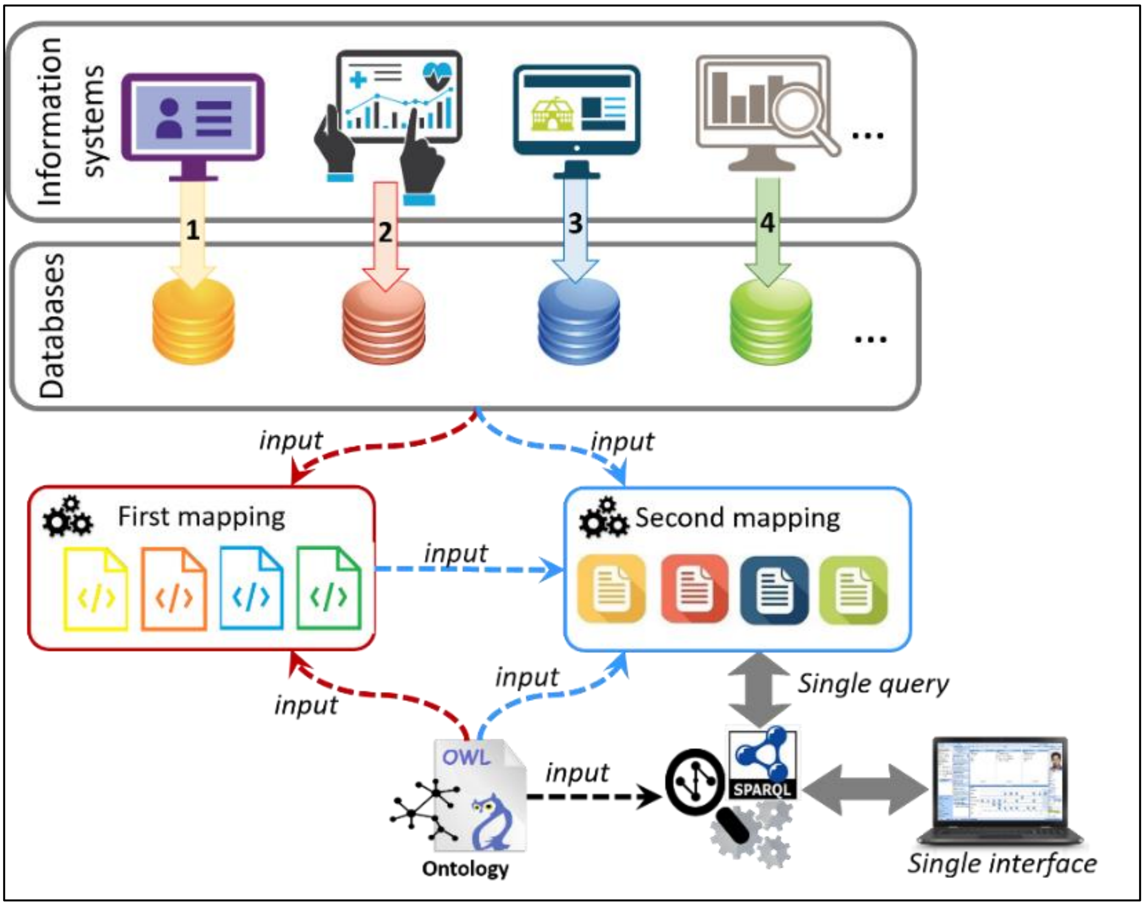


Source: Prepared by the author.

So, based on the proof of concept schema (Figure 96), an architecture to promote a database integration involving the use of ontologies was demonstrated in Figure 98.

Following the color pattern, everything related to the information system number one is in yellow, number two is red, number three is blue, and number four is green.

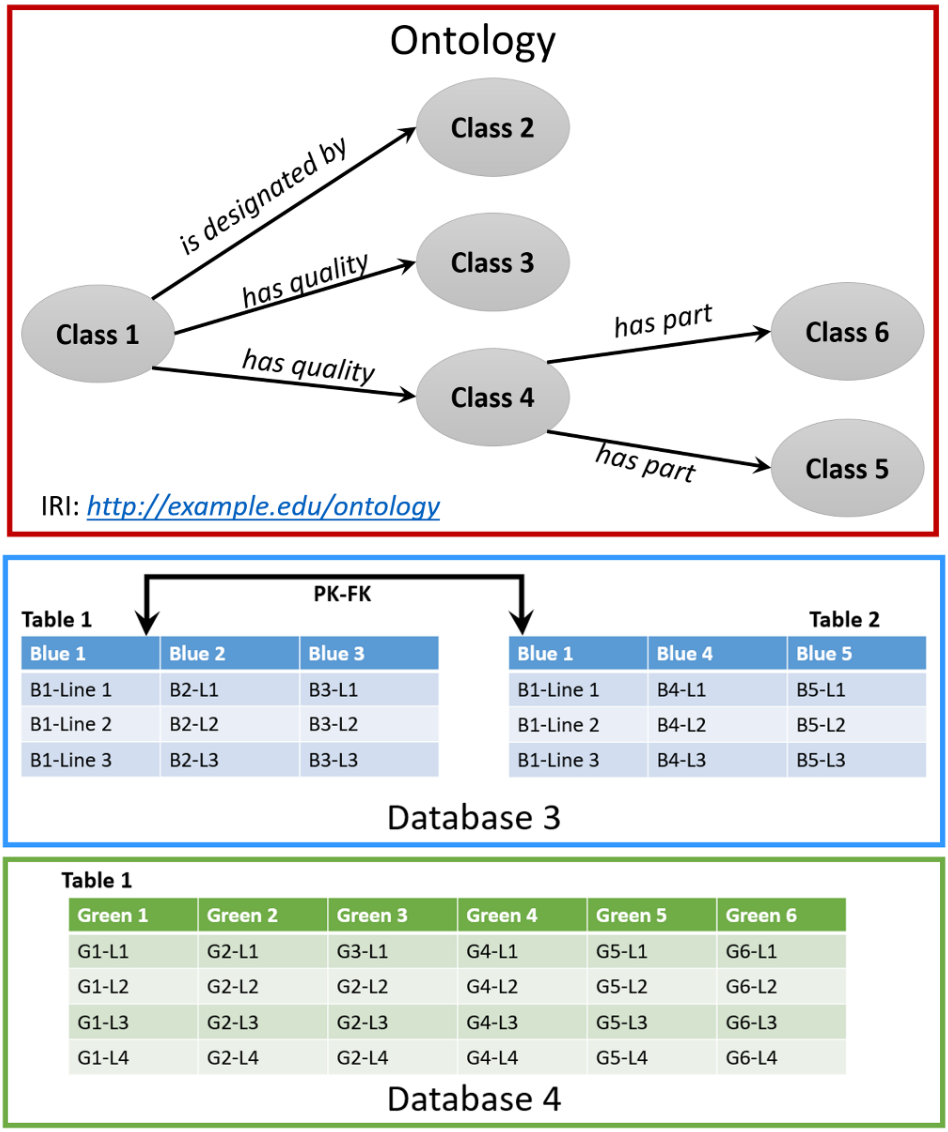
Figure 98: Conceptual schema to use ontologies as database integration solution.



Source: Prepared by the author.

As described in subsection 4.5, the first mapping generates an RDF turtle file creating the correspondences between the database metadata and the ontology elements to each database. Both the database metadata and ontology elements are inputs to this first mapping process. The second mapping generates a second RDF turtle file tanking as input to this process the data from the database, the ontology, and the RDF turtle file generated in first mapping. The output is one RDF triple file to each database with the set of instances corresponding to database elements. We will label this second RDF triple file as a dataset and the first RDF triple file as database-ontology scheme. Once the mapping happened and both the *database-ontology schema* and the *datasets* were generated you are able to query the data using a single SPARQL query and to visualize the query results in a single interface. In order to exemplify this approach, we prepare the follow ontology and databases presented in Figure 99.

Figure 99: Example of structures to use ontologies as database integration solution.



Source: Prepared by the author.

Take in mind that this is just conceptual examples and as future work, we will test in practice this. The IRI mentioned to such ontology is a false or fake IRI defined just to exemplify the solution idea to use ontology as a database integration. Firstly, you should map each database to the ontology. Figure 100 shows the example of mapping of against the database structures and the ontology showed in Figure 99. The colors maintain the reference to the origins, thus blue correspond to Database 3 and green to Database 4.

Figure 100: Example of mapping multiples databases to the same ontology.

database-ontology schemas

Blue 1	rdf:type	Class 1	Green 1	rdf:type	Class 1
Blue 2	rdf:type	Class 2	Green 2	rdf:type	Class 2
Blue 3	rdf:type	Class 3	Green 3	rdf:type	Class 3
Blue 4	rdf:type	Class 4	Green 4	rdf:type	Class 4
Blue 5	rdf:type	Class 5	Green 5	rdf:type	Class 5
			Green 6	rdf:type	Class 6

Datasets

Blue dataset IRI: http://example.edu/blueDB			Green dataset IRI: http://example.edu/greenDB		
B1-Line 1	instance of	Class 1	G1-L1	instance of	Class 1
B1-Line 2	instance of	Class 1	G1-L2	instance of	Class 1
B1-Line 3	instance of	Class 1	G1-L3	instance of	Class 1
B2-L1	instance of	Class 2	G2-L1	instance of	Class 2
B2-L2	instance of	Class 2	G2-L2	instance of	Class 2
B2-L3	instance of	Class 2	G2-L3	instance of	Class 2
B3-L1	instance of	Class 3	G3-L1	instance of	Class 3
…			…		
B4-L1	instance of	Class 4	G4-L1	instance of	Class 4
…			…		
B5-L1	instance of	Class 5	G5-L1	instance of	Class 5
…			…		

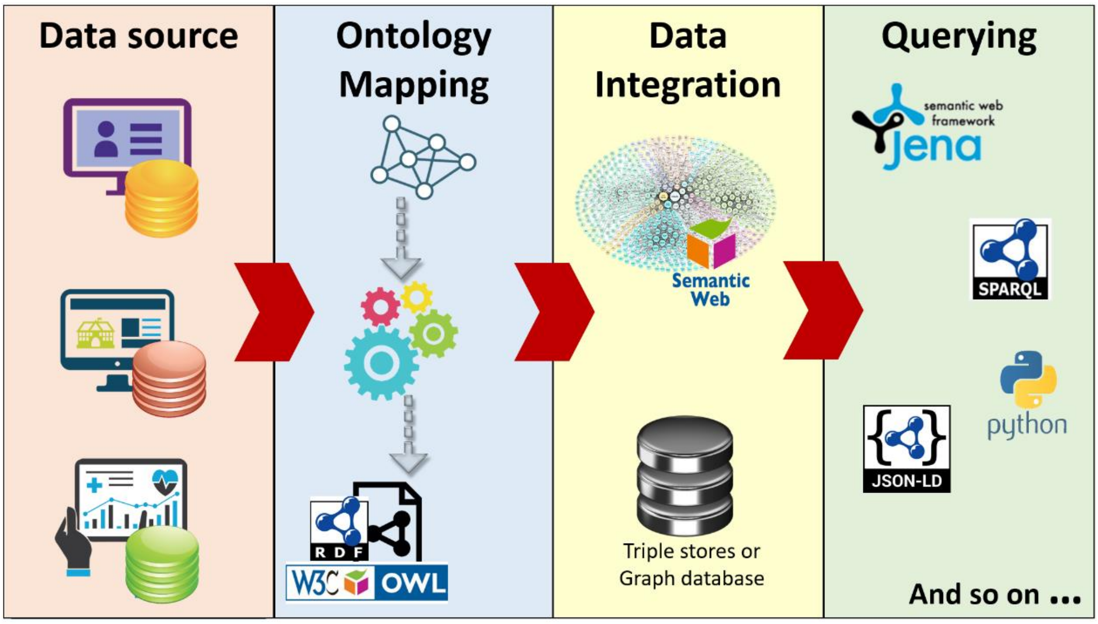
Source: Prepared by the author.

Therefore, after mapping, we have a database-ontology schema and datasets to each database mapped. Just as the false IRI for the ontology, the IRI for each set of data is also false, i.e. these IRIs were created just for exemplification. Note that these mapping aims to create the correspondence both between database metadata to ontology elements in the level of universals (first mapping) and between database data or values of lines and columns to ontology elements in the level of particulars (second mapping). Consequently, the output files of each database will be in the same structure once that the ontology definition is the common point between these databases. For example, according to Figure 100, the database-ontology scheme originated from Table 1 and Table 2 of Database 3 (Blue), the table columns Blue 1, Blue 2, Blue 3, Blue 4, and Blue 5, correspond to Class 1, Class 2, Class 3, Class 4, and Class 5, respectively. In the same way, the database-ontology scheme originated from Table 1 of Database 4 (Green) has the table columns Green 1, Green 2, Green 3, Green 4, Green 5, and Green 6, correspond to Class 1, Class 2, Class 3, Class 4, Class 5, and Class 6, respectively. Consequently, the table column Blue 1 from Table 1 of Database 3 is equivalent to the table column Green 1 from Table 1 of Database 4 since they correspond to the same ontology class, which is Class 1.

Thus, each dataset could be queried using a SPARQL query. Since the datasets originated from each database based on these database-ontology schemes, the triple structure of data follows the same ontology definition, consequently, the data inside these datasets could be retrieved using the same SPARQL query. By the way, these datasets are in RDF format and in order to be querying they also could be unified in a database based on RDF triple stores such as Sesame[89], Virtuoso[90], and Jena TDB[91], or in a graph databases such as Neo4J[92] and Ontotext GraphDB[93]. In this both cases, an ontology, like OntONeo, could be the semantic schema used to create the databases structures.

Thus, a general schema to provide data integration fostering semantic interoperability is illustrated in Figure 101. Independent of which data source we have, by mapping such data source against an ontology we can integrate the data or in databases based on ontologies or RDF technologies and consequently query the data using different type of query languages.

Figure 101: General schema of data interoperability using ontologies.



Source: Prepared by the author.

The real contribution of the proof of concept in subsection 4.5 was to demonstrate the feasibility of using OntoNeo as a common vocabulary that could enable the data exchange between different information systems. Once defined such vocabulary, the efforts to provide interoperability among legacy systems could be minimized by involving less maintenance on

[89] Details in: https://www.w3.org/2001/sw/wiki/Sesame

[90] Details in: https://virtuoso.openlinksw.com/

[91] Details in: https://jena.apache.org/documentation/tdb/

[92] Details in: https://neo4j.com/

[93] Details in: http://graphdb.ontotext.com/free/index.html

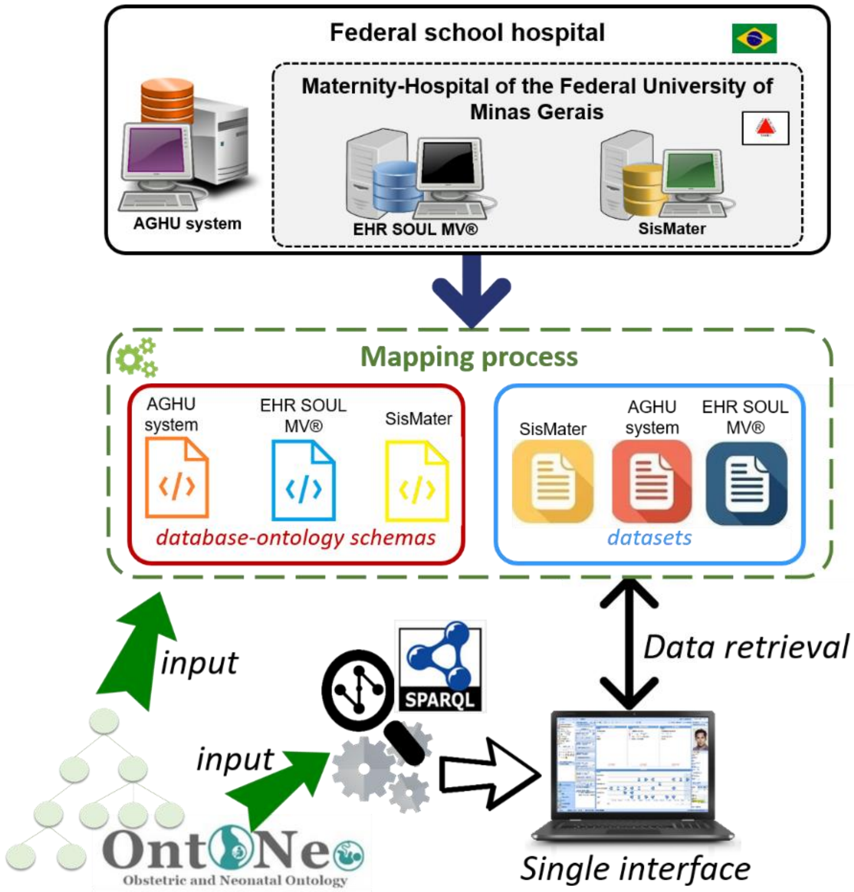
such systems. It involved building a middleware layer between the legacy system and the ontology providing the mappings between the legacy system database and the ontology. Then, from a common query interface, we can submit a query to multiples dataset unifying the query results.

Remembering, the problem of lack of semantic interoperability evidenced in the UFMG hospital involves multiples legacy systems, that is, existing and consolidated information systems. Hence, an ideal solution to promote the exchange of data between such systems is a solution that causes the least impact and less maintenance in these legacy systems. That is the role of formal ontology in a data exchange between information systems, to promote a semantic integration point. You do not need to make changes to legacy systems, just create a middleware layer that consists of the mapping between such database system and the formal ontology. Then, it could be created a single interface to visualize the queries results performed in all relevant databases independent of the physical location of the data (Figure 102). Additionally, you can create a dataset integrating data from different sources, either using the web of data, a linked data pattern, an RDF Triple store, or a database, in which it can be queried through a unified interface or multiple decentralized interfaces.

Considering the problem of the Maternity Hospital from the Federal University of Minas Gerais described in chapter 1 (Figure 1), OntONeo can be used as the semantic integration point between the different systems adopted by the federal school hospital of UFMG (such as AGHU system, EHR SOUL MV, and SisMater), as shown on Figure 102. For each of these systems, one must determine the database-ontology schemas and datasets using as input to the mapping process both OntONeo and such databases. Then, through a single interface, you can trigger a query on the datasets involved to retrieve the required data, and consequently, help to provide physicians with information to foster the continuity of care of women and newborns.

At this stage of this research, we provided here only a conceptual simulation of the way of employing the ontology as a semantic interoperable solution on the federal school hospital of UFMG. However, there are plans for making OntoNeo effective in the daily information processing routine in the scope of medical institutions. These plans concern the application of the OntoNeo as a terminological reference for systems working on the continuity of care of women and newborns in the Brazilian Health Unified System (SUS).

Figure 102: Proposal scheme of OntONeo use on Federal school hospital of UFMG.



Source: Prepared by the author.

Another point to consider about OntONeo ontology is the fact of this ontology follows mainly the recommendations of the methodology of the ontological realism. As presented in subsection 2.4.1, the driving principle of ontological realism takes an ontology as representational artefact committed to representing the reality of a given scientific domain. Ontological realism deals with the study of what kinds of things exist, which entities exist in the universe. That is to say, an adequate ontology must represent reality itself and not what one thinks or interprets about reality. Because of this, OntONeo is ad-hoc of computer technology in the sense of the technology used to develop both the information system and database does not influence in the semantic meaning of the ontology components.

Furthermore, when the domain ontology is ontological realism-based the ontology represents a single version of the true meaning of things existing in reality. We can reflect on the idea of data meaning federation that we base on the database federation in which is understood by "*an architecture in which middleware, consisting of a relational database management system, provides uniform access to a number of heterogeneous data sources*" (Haas; Lin; Roth, 2002). In this sense, a realism-based domain ontology enabling a data meaning federation by filling the gap of semantic interoperability among heterogeneous data models. A realism-based domain ontology would be a kind of middleware in a system architecture establishing a single version of the truth against the data manipulated by different information systems.

Finally, ontologies could help us to represent, organize, transmit, reuse, and share data of any knowledge domain. From the overview of formal ontologies and ontological realism recommendations, it follows that formal domain ontologies are a suitable for providing semantic interoperability between information systems both by providing a representation of knowledge without ambiguities and by representing only the things that exist in reality.

We believe that OntoNeo can be employed in many applications since it has been designed to provide both organization and consensual representation of information within the obstetric and neonatal domain. Even though we envisage its applicability, the obstetric and neonatal ontology is an ongoing initiative that needs improvements and validation to be properly adopted in real information systems.

6 FINAL REMARKS

“Ignorance more frequently begets confidence than does knowledge: it is those who know little, not those who know much, who so positively assert that this or that problem will never be solved by science”.
(Charles Darwin, 1871 - The Descent of Man)

This final chapter describes our final remarks drawn from our exploratory investigations that were designed to verify whether formal ontologies are suitable to foster semantic interoperability between information systems. This thesis was an applied qualitative research and exploratory involving a case study procedure that investigates the semantic gap of data exchange in the medical field, more specifically in obstetric and neonatal healthcare.

This thesis has revolved around two aspects: The first part aimed at providing a theoretical background for developing the researcher's understanding of the subject of research, analyzing the fundamentals of interoperability, ontologies, and healthcare information context. It also compiled the spread of studies on ontology building issues. The second part provided the empirical development of the research in which we built the ontology of obstetric and neonatal domain and evaluated its application in promoting semantic interoperability between patients' records gathered during prenatal, intrapartum and postnatal healthcare.

Our research focuses on the field of Information Science that is a multidisciplinary field realizing a relevant role as a provider of solutions regarding information issues for other scientific fields. Amidst the Information Science competencies, one can highlight the methods for knowledge representation and for information retrieval. The research on ontologies has been advancing in the Information Science field as a tool suitable to represent knowledge.

In light of Information Science field, an assumption preceding this study was the use of formal ontologies as instruments for knowledge representation able to promote semantic interoperability among information systems. Such assumption as a subject of study is still little understood and explored in the literature of Information Science, mainly regarding the process of ontology construction and the use of ontologies to foster semantic interoperability. So that, we expect with this study to enrich the discussion of the Information Science social role in helping other scientific fields in reference to solutions related to representation and organization knowledge seeking to improve semantics on data exchange

Our literature review suggests that the medical terminologies heterogeneity employed in the Brazilian healthcare systems constitutes a major challenge for the integration of healthcare information. Brazilian government initiatives for cooperation and interoperability among healthcare information systems become ineffective when information exchange is not efficient. Despite the importance of data from both the electronic medical record and electronic discharge summary to clinical practice and research, mainly to ensure the continuity of care, these data remain scattered throughout a range of disparate sources. Additionally, the lack of standardization and systematization of these data concerning the use and definition of terms poses problems for physicians and biomedical professionals.

Investments in studies to provide solutions for semantic interoperability among healthcare information systems are justified since they promote improvements in the quality of healthcare for citizens. The interoperability among healthcare systems ensures the exchange of information between electronic medical record and electronic discharge summary, which can allow the monitoring of the patient's history, enable improvements in their healthcare, reduce errors in diagnosis and prescriptions, and minimize duplicity of information. Thus, there is a need to establish interoperability standards that, in addition, to share and organize information, allow the unambiguous retrieval of the clinical information assigned to them. To accomplish this, our findings indicated that solutions directed to formal ontologies have much to contribute by allowing the formal unambiguous representation of concepts and their relationships.

The general goal of this research was to demonstrate a solution for semantic interoperability among information systems through the knowledge representation based on a formal ontology approach. Then, our case study investigated the data exchange gap between systems that deal with obstetrical and neonatal electronic healthcare records. First, we sought to review the main definitions and interpretations of the concepts related to our study focus with the aim of establishing the theoretical and conceptual framework to the research. In doing this, we hope to be contributing to a greater terminological clarification of the subject ontologies in the field of Information Science.

Furthermore, this study introduced the OntONeo Ontology (the ontology of obstetric and neonatal domain), an ongoing initiative designed to serve as a controlled vocabulary for use in organizing information about obstetric and neonatal care, adopting principally the recommendations methodology of the ontological realism. To build OntONeo, this study defined ReBORM, a Realism-Based Ontology engineeRing Methodology, combining the principals of the methodology of ontological realism and the ontology engineering best practices of NeOn methodology. We hope that ReBORM serve as guide to increasing the knowledge and ability of both researches beginners and potential ontologists.

Indeed, OntONeo ontology is not completed; OntONeo building needs more development iterations to reach the complete requirements defined at the concept phase. We verified that due the broad of knowledge on the obstetric and neonatal domain, our thesis timeline was not sufficient to cover the complete requirements.

Overall, this study serves as support to validate the use of formal ontology to foster semantic interoperability between information systems by filling the semantic gap of knowledge representation on data exchange. In addition, concretely, due my own almost 20 years of experience and expertise in database and data models in different knowledge domains, we defend that OntONeo ontology can serve as a canonical model in a service-oriented architecture. We also can argue that the OntONeo ontology can serve as a federated conceptual model between information systems that deal with data of obstetric and neonatal care.

To close, we hope that this research may contribute to the improvement of healthcare for the mother and the newborn, as well as the healthcare continuity by proposing an interoperability solution among medical records coming from the medical specialty of obstetrics. Moreover, this research reinforces the importance of interoperability in the field of healthcare and analyzes what has been done, what needs to be done, and how Information Science field can contribute to it.

6.1 Thesis contributions

According to previous chapters, our findings suggest a set of contributions listed below.

First, in section 2.4 of chapter 2, this thesis *consolidated the spread of studies on ontology building* issues, forming a consolidated reference to the community.

Second, in section 4.2 of chapter 4, this thesis established *the methodology to build ontology labelled ReBORM* (**Re**alism-**B**ased **O**ntology enginee**R**ing **M**ethodology), that defines "*what to do*", "*when to do*" and "*how to do*" in ontology building.

Third, the *OntONeo development process using ReBORM*. In section 4.3 of chapter 4, we describe how the ReBORM phases and activities were carried out during the first iterations of OntONeo development. We describe the tools used in each activity, the project decisions, and challenges encountered.

Fourth, *OntONeo fills the lack of a formal representation of the knowledge on obstetric and neonatal domain*. We highlight that our ontology consists in a *bilingual representation*. During the ontology development, the obstetrical and neonatal knowledge was represented in both English and Portuguese languages. Ontologies elements such classes and relations had their label, definitions and alternative label expressed in the two languages mentioned. Notice that not only the ontologies elements defined in OntONeo were submitted to the ontology localization activity, but also the elements imported from reusing ontologies.

Fifth, *the documentation about the implementation of the OBO Foundry principles*. To publish OntONeo as an OBO Foundry library we had to adhere to a set of principles defined by that community. However, such information, when documented, was spread and poorly organized in the OBO Foundry website and OBO discussion group. In subsection 4.4, we describe how OntONeo fits each one of OBO Foundry principles, serving as a reference for other researchers.

Sixth, *the demonstration of ontologies uses as a semantic reference on data exchange*. The proof of concept described in subsection 4.5 allowed visualizing how ontologies can be applied in the architecture of a system both to promote the semantic interoperability serving as a formal unambiguous representation and to be employed as the conceptual model of the information system.

Finally, an interesting side finding was that *there are some inconsistencies of definition among the existing ontologies in the OBO Foundry Portal*. These inconsistencies can lead to erroneous definitions when reused without any analysis. Our suggestion, in this case, is that even if it requires some extra effort in the analysis of inconsistencies, reuse should be privileged, thus avoiding the emergence or aggravation of inconsistencies. We also suggest that when some inconsistency is identified, the ontologist communicates to those responsible for the ontologies involved and the OBO Foundry community as a whole.

6.2 Directions for Future Research

This section outlines future research for further exploring the potential of the formal ontology in both semantic interoperability and data exchange perspectives.

First of all, we should highlight the continuity of the OntONeo development to reach all requirements defined at the concept phase.

Then, we going to further investigate others possibilities of using OntONeo to retrieval data from different databases. For this, we going to create some others proof of concept similar and supplement to the one presented in subsections 3.3.5 and 4.5.

We planned a second proof of concept aiming evaluation of the OntONeo use as a conceptual model for query multiples databases that use different database management systems and data model, as presented in subsection 5.4 and illustrated in Figure 98 and Figure 99. In addition, we identified the necessity of perform the proof of concept for querying NoSQL databases, such Triple Database Management Systems (third proof of concept). Consequently, we should evaluate our ontology as a point of interconnection between relational and NoSQL databases (fourth proof of concept). These future evaluations become relevant when we think about the evolution of applications to semantic web, linked data, and big data technologies.

There is also a planning for apply OntONeo for natural processing language (NLP) applications within UFMG hospital. Additionally, considering that OntONeo is a bilingual ontology, we also intended to verify how the ontology behaves in retrieving and integrating data originated from systems that deal with records, in Portuguese and in English. Thus, we could verify the semantic integration capacity regardless the language representing the data.

Finally, we suggest demonstrating in practice how OntONeo could work as a federated conceptual model and as a canonical model.

6.3 Publications originated from this thesis

During the development of this research, the author published some partial thesis' results in conferences and journals, which are:

- Farinelli, F.; Silva, S. M.; Almeida, M. B. O papel das ontologias na interoperabilidade de sistemas de informação: reflexões na esfera governamental. In: XIV ENCONTRO NACIONAL DE PESQUISA EM PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 2013, Florianópolis, Brasil.
- Silva, S. M.; Farinelli, F.; Almeida, M. B. Um roteiro para modelagem conceitual de sistemas de informação baseada em princípios ontológicos. In: XV ENCONTRO NACIONAL DE PESQUISA EM PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 2014, Belo Horizonte, Brasil. p. 4072-4095
- Souza, A. D.; Farinelli, F; Almeida, M. B. A informação em oncologia na era do big data. In: XVI ENCONTRO NACIONAL DE PESQUISA EM PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 2015, João Pessoa, Brasil.
- Farinelli, F.; Almeida, M. B.; Souza, Y. L. Linked Health Data: how linked data can help provide better health decisions. In: MEDINFO - STUDIES IN HEALTH TECHNOLOGY AND INFORMATICS, 2015, v. 216, p. 1122.
- Farinelli, F.; Almeida, M. B.; Elkin, P. L.; Smith, B. OntONeo: The Obstetric and Neonatal Ontology. In: 7th INTERNATIONAL CONFERENCE ON BIOLOGICAL ONTOLOGY (ICBO), 2016, Oregon State University, Corvallis, Oregon, USA.
- Farinelli, F.; Almeida, M. B.; Elkin, P. L.; Smith, B. Dealing with elements of medical encounters: an approach based on ontological realism. In: 7th

INTERNATIONAL CONFERENCE ON BIOLOGICAL ONTOLOGY (ICBO), 2016, Oregon State University, Corvallis, Oregon, USA.

- Farinelli, F.; Almeida, M. B.; Elkin, P. L.; Smith, B. Medical encounters in the obstetric and neonatal domain: an approach based on ontological realism. In: 7th INTERNATIONAL CONFERENCE ON BIOLOGICAL ONTOLOGY (ICBO), 2016, Oregon State University, Corvallis, Oregon, USA.
- Farinelli, F.; Almeida, M. B.; Elkin, P. L.; Smith, B. Interoperability among Prenatal EHRs: A Formal Ontology Approach. In: AMIA Annual Symposium, 2016, Chicago, USA
- Almeida, M. B.; Farinelli, F. Ontologies for the representation of electronic medical records: The obstetric and neonatal ontology. ***Journal of the Association for Information Science and Technology***, v. 68, n. 11, p. 2529-2542, 2017, doi:10.1002/asi.23900.
- Farinelli, Fernanda; Elkin, Peter L. Construção de ontologia na prática: um estudo de caso aplicado ao domínio obstétrico. **Ciência da Informação**, v. 46, n. 1, 2017. DOI: 10.18225/ci.inf..v46i1.4018.

REFERENCES

AAP. **Pediatrics 101:A Resource Guide from the American Academy of Pediatrics**. American Academy of Pediatrics, 2011. p.32.

ACOG. **PROLOG Gynecology and surgery: Question Book**. 6th Edition. American College of Obstetricians and Gynecologists, 2009.

______. Antepartum Record and Postpartum Form. In: Beckmann, C. R. B. *et al* (Ed.). **Obstetrics and Gynecology**. 6th. Philadelphia: Lippincott Williams & Wilkins, 2010a. cap. Apendice C, p.459-472. ISBN: 978-0781788076.

______. Woman's Health Record. In: Beckmann, C. R. B. *et al* (Ed.). **Obstetrics and Gynecology**. 6th. Philadelphia: Lippincott Williams & Wilkins, 2010b. cap. Apendice A, p.447-464. ISBN: 978-0781788076.

______. **Well-Woman Visit: Committee opinion** ACOG website: The American College of Obstetricians and Gynecologists. 534**:** 4 p. 2012. Disponível online em: <https://www.acog.org/wellwoman >.

______. **PROLOG Obstetrics: Critique book**. 7th Edition. American College of Obstetricians and Gynecologists, 2013.

______. **Well-Woman Recommendations**. ACOG website: The American College of Obstetricians and Gynecologists 2017. Disponível online em: <https://www.acog.org/wellwoman >.

ACSQHC. **Electronic Discharge Summary Systems Literature Scan**: Australian Commission on Safety and Quality in Health Care 2010. Disponível online em: <https://www.safetyandquality.gov.au/wp-content/uploads/2012/02/e-DischargeLiteratureScan.pdf >.

______. **Electronic Discharge Summary Self-Evaluation Toolkit**: Australian Commission on Safety and Quality in Health Care 2011a. Disponível online em: <https://www.safetyandquality.gov.au/wp-content/uploads/2012/01/EDS-self-eval-toolkit-sept2011.pdf >.

______. **Safety and Quality Evaluation of Electronic Discharge Summary Systems**. Australian Commission on Safety and Quality in Health CareAugust 2011, p.44. 2011b Disponível online em: < https://www.safetyandquality.gov.au/wp-content/uploads/2012/02/EDS-Evaluation-Final-Report-August-2011.pdf >.

AHRQ. **Children's Electronic Health Record Format**: Agency for Healthcare Research and Quality 2013. Disponível online em: <https://healthit.ahrq.gov/health-it-tools-and-resources/childrens-electronic-health-record-ehr-format >.

ALMEIDA, MAURICIO BARCELLOS; BARBOSA, RICARDO RODRIGUES. Ontologies in knowledge management support: A case study. **Journal of the American Society for Information Science and Technology,** v. 60, n. 10, p. 2032-2047, 2009. DOI: 10.1002/asi.21120

ALMEIDA, MAURICIO BARCELLOS DE. **Um modelo baseado em ontologias para representação da memória organizacional.** 2006. (Doutorado). Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil. *Published in* Portuguese.

______. Revisiting ontologies: A necessary clarification. **Journal of the American Society for Information Science and Technology,** v. 64, n. 8, p. 1682-1693, 2013. DOI: 10.1002/asi.22861

ALMEIDA, MAURICIO BARCELLOS DE; SOUZA, RENATO ROCHA; FONSECA, FREDERICO. Semantics in the Semantic Web: A Critical Evaluation. **Knowledge organization,** v. 38, n. 3, p. 187-203, 2011.

AMB. **Classificação Brasileira Hierarquizada de Procedimentos Médicos**. Associação Médica Brasileira. São Paulo, p.210. 2016. (978-85-7868-285-9).

ARANGUREN, MIKEL EGAÑA *et al.* Ontology Design Patterns for bio-ontologies: a case study on the Cell Cycle Ontology. **BMC bioinformatics,** v. 9, n. Suppl 5, p. S1-S1, 04/29 2008. DOI: 10.1186/1471-2105-9-S5-S1

ARANGUREN, MIKEL EGAÑA *et al.* Applying Ontology Design Patterns in Bio-ontologies. In: Gangemi, A. and Euzenat, J. (Ed.). **Knowledge Engineering: Practice and Patterns: 16th International Conference, EKAW 2008, Acitrezza, Italy, September 29 - October 2, 2008. Proceedings**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. p.7-16. ISBN: 978-3-540-87696-0.

ARMS, WILLIAM Y. Thoughts about Interoperability in the NSDL: Draft for discussion. 2000.

ARMS, WILLIAM Y. *et al.* A spectrum of interoperability: the site for science for prototype for the NSDL. 2002. **D-Lib Magazine,** v. 8, n. 1, 2002.

ARP, ROBERT; SMITH, BARRY. Function, Role, and Disposition in Basic Formal Ontology. The 11th Annual Bio-Ontologies Meeting, 2008, Toronto, Canada. Citeseer, 2008. p.1.

ARP, ROBERT; SMITH, BARRY; SPEAR, ANDREW D. **Building Ontologies with Basic Formal Ontology**. Cambridge, Massachusetts: The Mit Press, 2015. 220.

BARD, JONATHAN. A new ontology (structured hierarchy) of human developmental anatomy for the first 7 weeks (Carnegie stages 1–20). **Journal of anatomy,** v. 221, n. 5, p. 406-416, 2012.

BECHHOFER, SEAN; GOBLE, CAROLE. Thesaurus construction through knowledge representation. **Data & Knowledge Engineering,** v. 37, n. 1, p. 25-45, 2001.

BECKMANN, CHARLES R. B. *et al.* **Obstetrics and Gynecology**. 6th. Philadelphia: Lippincott Williams & Wilkins, 2010a. 528.

______. The woman's health examination. In: Beckmann, C. R. B. *et al* (Ed.). **Obstetrics and Gynecology**. 6th. Philadelphia: Lippincott Williams & Wilkins, 2010b. cap. 1, p.1-13. ISBN: 978-0781788076
0781788072.

BERNERS-LEE, TIM; HENDLER, JAMES; LASSILA, ORA. The semantic web. **Scientific american,** v. 284, n. 5, p. 28-37, Apr 26 2001. DOI: 10.1038/35074206

BEZERRA, SELENE MARIA. Prontuário Eletrônico do Paciente: uma ferramenta para aprimorar a qualidade dos serviços de saúde. **Revista Meta: Avaliação,** v. 1, n. 1, p. 73-82, 2009.

BIOLCHINI, JORGE CALMON DE ALMEIDA *et al.* Scientific research ontology to support systematic review in software engineering. **Advanced Engineering Informatics,** v. 21, n. 2, p. 133–151, April 2007 2007. DOI: 10.1016/j.aei.2006.11.006

BIOPPORTAL. NCBO Bioportal repository. 2017. Disponível online em: < [http://bioportal.bioontology.org] >. Accesso em: 03/12/2017.

BISHR, YASER. Overcoming the semantic and other barriers to GIS interoperability. **International Journal of Geographical Information Science,** v. 12, n. 4, p. 299-314, 1998.

BITTNER, THOMAS; DONNELLY, MAUREEN; WINTER, STEPHAN. Ontology and semantic interoperability. **Large-scale 3D data integration: Challenges and Opportunities**, p. 139-160, 2005.

BITTNER, THOMAS; SMITH, BARRY. Formal ontologies of space and time. **IFOMIS, Department of Philosophy. Leipzig, Buffalo, University of Leipzig, University at Buffalo and NCGIA,** v. 17, 2003.

______. Normalizing medical ontologies using basic formal ontology. 2004.

BLOMQVIST, EVA; SANDKUHL, KURT. Patterns in Ontology Engineering: Classification of Ontology Patterns. 7th International Conference on Enterprise Information Systems, 2005, Miami, Florida, USA. 2005. p.413-416.

BODENREIDER, OLIVIER. The unified medical language system (UMLS): integrating biomedical terminology. **Nucleic acids research,** v. 32, n. suppl 1, p. D267-D270, 2004.

BORKO, HAROLD. Information science: what is it? **American documentation,** v. 19, n. 1, p. 3-5, 1968. DOI: 10.1002/asi.5090190103

BOTTOMLEY, CECILIA; RYMER, JANICE. **100 cases in obstetrics and gynaecology**. CRC Press, 2014.

BRASIL. Saúde, M. D. Portaria nº 1.459. from 24 de Junho 2011. Institui, no âmbito do Sistema Único de Saúde - SUS - a Rede Cegonha. Diário Oficial da União, Brasília, 27 de Junho de 2011, Seção 1, p. 109. *Published in* Portuguese.

______. Saúde, M. D. Portaria nº 2.073. from 31 de Agosto de 2011. Regulamenta o uso de padrões de interoperabilidade e informação em saúde para sistemas de informação em saúde no âmbito do Sistema Único de Saúde, nos níveis Municipal, Distrital, Estadual e Federal, e para os sistemas privados e do setor de saúde suplementar. Diário Oficial da União, Brasília, 1º de Setembro de 2011, Seção 1, p. 63. *Published in* Portuguese.

______. Saúde, M. D. Portaria nº 1.412. from 10 de julho de 2013. Institui o Sistema de Informação em Saúde para a Atenção Básica. Diário Oficial da União, Brasília, 11 de Julho de 2013, Seção 1, p. 294. *Published in* Portuguese.

______. Saúde, M. D. Portaria nº 1.653. from 02 de Outubrode 2015. Acrescenta o art. 2º-A à Portaria nº 1.412/GM/MS, de 10 de julho de 2013, que Institui o Sistema de Informação em Saúde para a Atenção Básica (SISAB), com o objetivo de garantir a transição entre o Sistema de Registro das Ações Ambulatoriais de Saúde (RAAS) e o SISAB. Diário Oficial da União, Brasília, 5 de Outubro de 2015, Seção 1, p. 673. *Published in* Portuguese.

______. Saúde, M. D. Portaria nº 76. from 22 de Janeiro de 2016. Institui os prazos para o envio da produção da Atenção Básica para o Sistema de Informação em Saúde para a Atenção Básica referente às competências de janeiro a dezembro de 2016. Diário Oficial da União, Brasília, 25 de Janeiro de 2016, Seção 1, p. 35. *Published in* Portuguese.

BRASIL, MINISTÉRIO DA SAÚDE DO. **Pré-natal e Puerpério: atenção qualificada e humanizada**. Manual técnico do Ministério da Saúde do Brasil. Mulher., S. D. a. À. S. D. D. a. P. E. Á. T. D. S. D. Brasília: Ministério da Saúde. Caderno nº 5**:** 163 p. 2006.

______. **Entendendo o SUS**. Ministério da Saúde do Brasil. Brasília, DF, Brasil, p.28. 2007 Disponível online em: < http://portalarquivos.saude.gov.br/images/pdf/2013/agosto/28/cartilha-entendendo-o-sus-2007.pdf >.

______. Brasil, M. D. S. D. Nota técnica conjunta da Rede Cegonha.

______. **Caderneta da Gestante**. Ministério Da Saúde, B., Brasil. Brasília, Brasil: SUS 2014a. Disponível online em: <http://189.28.128.100/dab/docs/portaldab/documentos/caderneta_gestante.pdf >.

______. **FICHA PERINATAL – Ambulatório**. Ministério Da Saúde, B., Brasil. Brasília, Brasil: SUS 2014b. Disponível online em: <http://189.28.128.100/dab/docs/portaldab/documentos/ficha_perinatal_ambulatorio.pdf >.

______. **Protocolos da Atenção Básica : Saúde das Mulheres**. Brasília: Ministério da Saúde, 2016. p.230. ISBN: 978-85-334-2360-2.

BRASIL, MINISTÉRIO DA SAÚDE DO; SAÚDE, SECRETARIA DE ATENÇÃO À; ESTRATÉGICAS, DEPARTAMENTO DE AÇÕES PROGRAMÁTICAS. **Política nacional de atenção integral à saúde da mulher: princípios e diretrizes**. Brasília, DF, Brasil: Ministério da Saúde, 2004. 82.

BROWN, ELLIOT G.; WOOD, LOUISE; WOOD, SUE. The medical dictionary for regulatory activities (MedDRA). **Drug safety,** v. 20, n. 2, p. 109-117, 1999.

BROWNING, PEGGY ELAINE. Medical Records. In: Longe, J. L. (Ed.). **The Gale Encyclopedia of Nursing and Allied Health**. 2nd ed. Detroit: Gale, v.3, 2006. p.1723-1724.

CALLEN, JOANNE L.; ALDERTON, MELANIE; MCINTOSH, JEAN. Evaluation of electronic discharge summaries: A comparison of documentation in electronic and handwritten discharge summaries. **International Journal of Medical Informatics,** v. 77, n. 9, p. 613-620, 9// 2008. DOI: http://doi.org/10.1016/j.ijmedinf.2007.12.002

CALVANESE, DIEGO; DE GIACOMO, GIUSEPPE; LENZERINI, MAURIZIO. A framework for ontology integration. 2001, CEUR-WS. org, 2001. p.303-316.

CEUSTERS, WERNER; SMITH, BARRY; FLANAGAN, JIM. Ontology and medical terminology: Why description logics are not enough. **Towards an Electronic Patient Record (TEPR 2003), Boston, MA**, 2003.

CFM. Medicina, C. F. D. Resolução nº 1.638 from 9 de agosto de 2002. Define prontuário médico e torna obrigatória a criação da Comissão de Revisão de Prontuários nas instituições de saúde. Diário Oficial União, Brasília, DF, Brasil, 10 de julho de 2002, Seção 1, p. 184-185.

CFM; SBIS. **Cartilha sobre Prontuário Eletrônico - A Certificação de Sistemas de Registro Eletrônico de Saúde**. Conselho Federal de Medicina
Sociedade Brasileira de Informática em Saúde, p.20. 2012 Disponível online em: < http://portal.cfm.org.br/crmdigital/Cartilha_SBIS_CFM_Prontuario_Eletronico_fev_2012.pdf >.

CHACON, SCOTT; STRAUB, BEN. **Pro git**. Apress, 2014. ISBN: 978-1-4842-0077-3 (Print) 978-1-4842-0076-6 (Online).

CHO, SUNYOUNG; MATHIASSEN, LARS; NILSSON, AGNETA. Contextual dynamics during health information systems implementation: an event-based actor-network approach. **European Journal of Information Systems,** v. 17, n. 6, p. 614-630, 2008.

CHOI, NAMYOUN; SONG, IL-YEOL; HAN, HYOIL. A survey on ontology mapping. **ACM Sigmod Record,** v. 35, n. 3, p. 34-41, 2006.

COLLIS, JILL; HUSSEY, ROGER. **Business research: A practical guide for undergraduate and postgraduate students**. 4th Edition. Palgrave macmillan, 2014. 376.

CORCHO, OSCAR; FERNÁNDEZ-LÓPEZ, MARIANO; GÓMEZ-PÉREZ, ASUNCIÓN. Methodologies, tools and languages for building ontologies. Where is their meeting point? **Data & Knowledge Engineering,** v. 46, n. 1, p. 41-64, 7// 2003. DOI: http://dx.doi.org/10.1016/S0169-023X(02)00195-7

CÔTÉ, RICHARD G. *et al.* The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. **BMC bioinformatics,** v. 7, n. 1, p. 97, 2006. DOI: 10.1186/1471-2105-7-97

CÔTÉ, RICHARD G. *et al.* The Ontology Lookup Service: more data and better tools for controlled vocabulary queries. **Nucleic acids research,** v. 36, n. suppl_2, p. W372-W376, 2008. DOI: https://doi.org/10.1093/nar/gkn252

CÔTÉ, RICHARD *et al.* The ontology lookup service: bigger and better. **Nucleic acids research,** v. 38, n. suppl_2, p. W155-W160, 2010. DOI: https://doi.org/10.1093/nar/gkq331

COURTOT, MÉLANIE *et al.* MIREOT: The minimum information to reference an external ontology term. **Applied ontology,** v. 6, n. 1, p. 23-33, 2011. DOI: 10.3233/Ao-2011-0087

CRESWELL, JOHN W. **Research design: qualitative, quantitative, and mixed methods approaches**. 3rd edition. SAGE Publications, 2009.

______. **Research design: qualitative, quantitative, and mixed methods approaches**. 4th edition. Sage, 2013.

CUNNINGHAM, F. GARY *et al.* Embryogenesis and Fetal Morphological Development. In: (Ed.). **Williams Obstetrics, 24e**. New York, NY: McGraw-Hill Education, 2013.

______. **Williams Obstetrics**. 24th edition. New York, US: McGraw-Hill Education, 2014. 1376.

D'AQUIN, MATHIEU *et al.* Watson: Supporting next generation semantic web applications. 2007.

D'AQUIN, MATHIEU; MOTTA, ENRICO. Watson, more than a semantic web search engine. **Semantic Web,** v. 2, n. 1, p. 55-63, 2011.

D'AQUIN, MATHIEU *et al.* Watson: A gateway for the semantic web. 2007.

DAHLBERG, INGETRAUT. Knowledge organization: its scope and possibilits. **Knowledge organization,** v. 20, n. 4, p. 211-222, 1993.

______. Current trends in knowledge organization. **Actas del I Encuentro ISKO-España. Organización del Conocimiento en Sistemas de Información y Documentación,** v. 1, p. 7-25, 1995.

______. Knowledge organization: a new science? **Knowledge organization,** v. 33, n. 1, p. 11-19, 2006.

DAMA, DATA MANAGEMENT ASSOCIATION INTERNATIONAL. **DAMA-DMBOK: Data management body of knowledge** 2ª. Bas King Ridge, New Jersey, USA: Technics Publications LLC, 2017. p.624. ISBN: 978-1-6346223-6-3.

DE NICOLA, ANTONIO; MISSIKOFF, MICHELE; NAVIGLI, ROBERTO. A proposal for a unified process for ontology building: UPON. In: K.V., A.;J., D. and R., W., Database and Expert Systems Applications. DEXA 2005, 2005, Springer, Berlin, Heidelberg. p.655-664.

______. A software engineering approach to ontology building. **Information Systems,** v. 34, n. 2, p. 258-275, 2009.

DE SOUZA, ANDRÉIA CRISTINA. **Identificaçao do Conteúdo Padronizado do Sumário de Alta**. 2012. Master dissertation Escola Politécnica, Pontifícia Universidade Católica do Paraná, Curitiba, Paraná, Brasil. *Published in* Portuguese.

DING, LI *et al.* Swoogle: a search and metadata engine for the semantic web. 13th ACM international conference on Information and knowledge management - CIKM'04, 2004, Washington, DC, USA. ACM New York, NY, USA ©2004, November 08-13. p.652-659.

DING, LI *et al.* Finding and ranking knowledge on the semantic web. The Semantic Web – ISWC 2005, 2005, Springer, Berlin, Heidelberg, 2005. p.156-170.

DUBOSE, TERRY J. Sex, Heart Rate and Age. **OBGYN. net," Arkiverade kopian". Arkiverad från originalet den,** v. 15, 2012.

DUCHARME, BOB. **Learning Sparql: Querying and Updating with SPARQL 1.1**. 1st Edition. O'Reilly Media, Inc., 2011. 235.

EHRIG, MARC. **Ontology alignment: bridging the semantic gap**. Springer Science & Business Media, 2006. p.247. ISBN: 038736501X
e-ISBN-10: 0-387-36501-X

e-ISBN-13: 978-0-387-36501-5

EHRIG, MARC; SURE, YORK. Ontology mapping-an integrated approach. 2004, Springer, 2004. p.76-91.

ELKIN, PETER L. *et al.* Guideline and quality indicators for development, purchase and use of controlled health vocabularies. **International Journal of Medical Informatics,** v. 68, n. 1, p. 175-186, 2002/12/18/ 2002. DOI: https://doi.org/10.1016/S1386-5056(02)00075-8

ELKIN, PETER L.; TUTTLE, MARK SAMUEL. Introduction. In: Elkin, P. L. (Ed.). **Terminology and terminological systems**. London: Springer-Verlag, 2012. cap. 1, p.1-4. ISBN: 1447128168.

ESPINOZA, MAURICIO; MONTIEL-PONSODA, ELENA; GÓMEZ-PÉREZ, ASUNCIÓN. Ontology localization. In: Suárez-Figueroa, M. C. *et al*, Fifth international conference on Knowledge Capture (KCAP 2009), 2009, Redondo Beach, CA, USA. ACM, September 01 - 04. p.33-40.

EUZENAT, JÉRÔME. Towards a principled approach to semantic interoperability. Proc. IJCAI 2001 workshop on ontology and information sharing, 2001, No commercial editor. p.19-25.

EUZENAT, JÉRÔME *et al.* Ontologies and semantic interoperability. ECAI-02 Workshop on Ontologies and Semantic Interoperability, 2002, Lyon, France.

FALBO, RICARDO DE ALMEIDA. SABiO: Systematic Approach for Building Ontologies. In: Guizzardi, G. *et al*, ONTO-COM-ODISE 2014- Ontologies in Conceptual Modeling and Information Systems Engineering, 2014, Rio de Janeiro, Brazil. CEUR Workshop Proceedings, September 21.

FEIGENBAUM, LEE. **SPARQL By Example: A tutorial**: Word Wide Web Consortium 2009. Disponível online em: <https://www.w3.org/2009/Talks/0615-qbe/ >. Accesso em: 7/4/2017.

FERNÁNDEZ-LÓPEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. Methontology: from ontological art towards ontological engineering. Proceedings of the Ontological Engineering AAAI-97 Spring Symposium Series, 1997, Stanford University, EEUU. American Asociation for Artificial Intelligence, 24-26 March

FERNÁNDEZ-LÓPEZ, MARIANO. Overview of methodologies for building ontologies. IJCAI-99 workshop on Ontologies and Problem-Solving Methods, 1999, Stockholm, Sweden. CEUR Workshop Proceedings (CEUR-WS.org), 1999. p.1-13.

FERNÁNDEZ-LÓPEZ, MARIANO; GÓMEZ-PÉREZ, ASUNCIÓN; OACUTE. Overview and analysis of methodologies for building ontologies. **The Knowledge Engineering Review,** v. 17, n. 02, p. 129-156, 2002. DOI: doi:10.1017/S0269888902000462

FERREIRA, ALVARO A T *et al.* PROPOSIÇÃO DE UM SUMÁRIO DE ALTA OBSTÉTRICO VISANDO À TROCA DE INFORMAÇÕES, EM PADRÃO OPENEHR, PARA CONTINUIDADE DO CUIDADO MATERNO-INFANTIL. **Revista da Faculdade de Medicina de Ribeirão Preto e do Hospital das Clínicas da FMRP Universidade de São Paulo,** v. 47, n. Supl 1, p. 59-66, 2014.

FESCHAREK, REINHARD *et al.* Medical Dictionary for Regulatory Activities (MedDRA). **International journal of pharmaceutical medicine,** v. 18, n. 5, p. 259-269, 2004.

FININ, TIM *et al.* Swoogle: Searching for knowledge on the Semantic Web. Twentieth National Conference on Artificial Intelligence AAAI'05, 2005, Pittsburgh, Pennsylvania, USA. 4: Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, July 9–13. p.1682.

FONSECA, FREDERICO. The double role of ontologies in information science research. **Journal of the American Society for Information Science and Technology,** v. 58, n. 6, p. 786-793, 2007.

FREITAS, FRED; SCHULZ, STEFAN; MORAES, EDUARDO. Survey of current terminologies and ontologies in biology and medicine. **Reciis,** v. 3, p. 7-18, 2009.

GALVÃO, MARIA CRISTIANE BARBOSA; RICARTE, IVAN LUIZ MARQUES. **Prontuário do paciente**. Rio de Janeiro: Guanabara Koogan, 2012. 322.

GANGEMI, ALDO. Ontology design patterns for semantic web content. In: Gil, Y. *et al*, 4th international conference on The Semantic Web - ISWC'05, 2005, Galway, Ireland. Springer-Verlag Berlin, November 06 - 10. p.262-276.

GANGEMI, ALDO *et al.* A theoretical framework for ontology evaluation and validation. In: Bouquet, P. and Tummarello, G., Semantic Web Applications and Perspectives - SWAP, 2005, Trento, Italy. CEUR Workshop Proceedings.

______. Modelling Ontology Evaluation and Validation. In: Sure, Y. and Domingue, J. (Ed.). **The Semantic Web: Research and Applications: 3rd European Semantic Web Conference, ESWC 2006 Budva, Montenegro, June 11-14, 2006 Proceedings**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. p.140-154. ISBN: 978-3-540-34545-9.

GANGEMI, ALDO *et al.* A formal ontological framework for semantic interoperability in the fishery domain. In: Euzenat, J. *et al*, European Conference on Artificial Intelligence (ECAI'02) - Workshop on Ontologies and Semantic Interoperability, 2002a, Lyon, France. CEUR Workshop Proceedings, July 22. p.16-30.

______. A formal ontological framework for semantic interoperability in the fishery domain. In: Euzenat, J. *et al*, ECAI-02 Workshop on Ontologies and Semantic Interoperability, 2002b, Lyon, France. 2002. p.16-30.

GANGEMI, ALDO *et al.* Towards a catalog of owl-based ontology design patterns. 2007.

GANGEMI, ALDO *et al.* Sweetening ontologies with DOLCE. EKAW '02 - 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, 2002, Siguenza, Spain. Springer-Verlag, October 01-04. p.166-181.

GANGEMI, ALDO; PRESUTTI, VALENTINA. Ontology design patterns. In: (Ed.). **Handbook on ontologies**: Springer, 2009. p.221-243. ISBN: 3540709991.

GARCIA, ALEXANDER *et al.* CMAPS supporting the development of OWL ontologies. 2007 International Conference on Posters and Demonstrations, 2007, CEUR-WS. org, 2008. p.58-59.

GIBBONS, MICHAEL *et al.* **The new production of knowledge: The dynamics of science and research in contemporary societies**. Sage, 1994.

GIL, ANTONIO CARLOS. **Métodos e técnicas de pesquisa social**. 6. São Paulo: Atlas, 2008.

GÓMEZ-PÉREZ, ASUNCIÓN. Evaluation of taxonomic knowledge in ontologies and knowledge bases. In: Gaines, B. R. and Musen, M., Banff Knowledge Acquisition for Knowledge-Based Systems (KAW'99), 1999, Banff, Alberta, Canadá. University of Calgary, October 16-21. p.6.1.1-6.1.18.

______. Ontology evaluation. In: (Ed.). **Handbook on ontologies**: Springer, 2004. p.251-273.

GÓMEZ-PÉREZ, ASUNCIÓN; FERNÁNDEZ-LÓPEZ, MARIANO; CORCHO, OSCAR. Methodologies and methods for building ontologies. **Ontological Engineering: with examples from the áreas of Knowledge Management, e-Commerce and the Semantic Web**, p. 107-197, 2004.

GÓMEZ-PÉREZ, ASUNCIÓN. Evaluation of ontologies. **International Journal of intelligent systems,** v. 16, n. 3, p. 391-409, 2001. DOI: 10.1002/1098-111X(200103)16:3<391::AID-INT1014>3.0.CO;2-2

GREER, RAQUEL C. *et al.* Hospital discharge communications during care transitions for patients with acute kidney injury: a cross-sectional study. **BMC Health Services Research,** v. 16, n. 1, p. 449, 2016// 2016. DOI: 10.1186/s12913-016-1697-7

GRENON, PIERRE. **BFO in a Nutshell: A Bi-categorial Axiomatization of BFO and Comparison with DOLCE**. Ifomis, 2003a.

______. Knowledge Management from the ontological Standpoint. In: Freyberg, K.;Petsche, H.-J. and Klein, B., Workshop on Knowledge Management and Philosophy, 2003b, Luzern, Switzerland. CEUR Workshop Proceedings, 2003. p.415-416.

______. Knowledge Management from the ontological Standpoint. Wissensmanagement, 2003c. p.415-416.

______. Nuts in BFO's nutshell: Revisions to the bi-categorial axiomatization of BFO. 2003d.

GRENON, PIERRE; SMITH, BARRY. SNAP and SPAN: Towards Dynamic Spatial Ontology. **http://dx.doi.org/10.1207/s15427633scc0401_5,** v. 4, n. 1, p. 69-104, 2004. DOI: Spatial Cognition and Computation, Vol. 4, No. 1, March 2004, pp. 69–104

GRENON, PIERRE; SMITH, BARRY; GOLDBERG, LOUIS. Biodynamic ontology: applying BFO in the biomedical domain. **Studies in health technology and informatics**, p. 20-38, 2004.

GRUBER, THOMAS R. A translation approach to portable ontology specifications. **Knowledge Acquisition,** v. 5, n. 2, p. 199-220, 1993a. DOI: http://dx.doi.org/10.1006/knac.1993.1008

______. What is an Ontology? , 1993b. Disponível online em: < http://www-ksl.stanford.edu/kst/what-is-an-ontology.html >. Accesso em: 01/31.

______. Toward principles for the design of ontologies used for knowledge sharing. **International journal of human-computer studies,** v. 43, n. 5–6, p. 907-928, 1995. DOI: http://dx.doi.org/10.1006/ijhc.1995.1081

______. Ontology. In: Liu, L. and Özsu, M. T. (Ed.). **Encyclopedia of Database Systems**: Springer-Verlag, 2008.

GRÜNINGER, MICHAEL; FOX, MARK S. Methodology for the Design and Evaluation of Ontologies. 1995.

GUARINO, NICOLA. Formal ontology, conceptual analysis and knowledge representation. **International journal of human-computer studies,** v. 43, n. 5-6, p. 625-640, 1995.

______. Semantic matching: Formal ontological distinctions for information organization, extraction, and integration. In: Pazienza, M. T. (Ed.). **Information Extraction A Multidisciplinary Approach to an Emerging Information Technology**: Springer Berlin Heidelberg, v.1299, 1997a. p.139-170. (Lecture Notes in Computer Science). ISBN: 978-3-540-63438-6
978-3-540-69548-6.

______. Understanding, building and using ontologies. **International journal of human-computer studies,** v. 46, n. 2–3, p. 293-310, 2// 1997b. DOI: http://dx.doi.org/10.1006/ijhc.1996.0091

______. Formal ontology in information systems In: Guarino, N., First Formal Ontology in Information Systems, 1998a, Trento, Italy. IOS press, June 6-8. p.81-97.

______. Some ontological principles for designing upper level lexical resources. **arXiv preprint cmp-lg/9809002**, 1998b.

GUARINO, NICOLA; GIARETTA, PIERDANIELE. Ontologies and knowledge bases: towards a terminological clarification. In: (Ed.). **Towards very large knowledge bases: knowledge building & knowledge sharing**, v.25, 1995. cap. 25, p.25-32.

GUARINO, NICOLA; OBERLE, DANIEL; STAAB, STEFFEN. What is an Ontology? In: (Ed.). **Handbook on ontologies**: Springer, 2009. p.1-17.

GUARINO, NICOLA; WELTY, CHRISTOPHER. A formal ontology of properties. In: Dieng, R. and Corby, O., 12th International Conference on Knowledge Engineering and Knowledge Management, 2000, Juan-les-Pins, France. Springer Berlin Heidelberg, October 2–6. p.97-112.

GUIZZARDI, GIANCARLO; WAGNER, GERD. Towards ontological foundations for agent modelling concepts using the unified fundational ontology (UFO). In: (Ed.). **Agent-Oriented Information Systems II**: Springer, 2005. p.110-124.

______. Using the unified foundational ontology (UFO) as a foundation for general conceptual modeling languages. In: (Ed.). **Theory and Applications of Ontology: Computer Applications**: Springer, 2010. p.175-196. ISBN: 9048188466.

GUY, MARIEKE. **Interoperability Focus: Looking at Interoperability** 2005. Disponível online em: <http://www.ukoln.ac.uk/interop-focus/about/leaflet.html >.

HAAS, LAURA M.; LIN, EILEEN TIEN; ROTH, MARY A. Data integration through database federation. **IBM Systems Journal,** v. 41, n. 4, p. 578-596, 2002.

HAENDEL, MELISSA A. *et al.* Unification of multi-species vertebrate anatomy ontologies for comparative biology in Uberon. **Journal of Biomedical Semantics,** v. 5, n. 1, p. 21, 2014. DOI: https://doi.org/10.1186/2041-1480-5-21

HAENDEL, MELISSA A. *et al.* Uberon: towards a comprehensive multi-species anatomy ontology. 2009.

HAENDEL, MELISSA A. *et al.* CARO: the common anatomy reference ontology. In: Burger, A.;Davidson, D. and Baldock, R. (Ed.). **Anatomy Ontologies for Bioinformatics**. New York: Springer, 2008. p.327-349.

HAFNER, NATALYA FRIDMAN NOY; CAROLE, D. The State of the Art in Ontology Design: A Survey and Comparative Review. **18,** v. 3, p. 53-74, 2008-01-29 1997. DOI: http://www.aaai.org/ojs/index.php/aimagazine/article/view/1306

HARTH, ANDREAS; JANIK, MACIEJ; STAAB, STEFFEN. Semantic Web Architecture. In: Domingue, J.;Fensel, D. and Hendler, J. A. (Ed.). **Handbook of Semantic Web Technologies**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. p.43-75. ISBN 978-3-540-92913-0.

HARVEY, FRANCIS *et al.* Semantic interoperability: A central issue for sharing geographic information. **The Annals of Regional Science,** v. 33, n. 2, p. 213-232, 1999.

HAUX, REINHOLD. Health information systems – past, present, future. **International Journal of Medical Informatics,** v. 75, n. 3, p. 268-281, 2006/03/01/ 2006. DOI: http://dx.doi.org/10.1016/j.ijmedinf.2005.08.002

HAWKINS, DONALD T. Information science abstracts: tracking the literature of information science. Part 1: definition and map. **Journal of the American Society for Information Science and Technology,** v. 52, n. 1, p. 44-53, 2001.

HAWKINS, DONALD T.; LARSON, SIGNE E.; CATON, BARI Q. Information science abstracts: tracking the literature of information science. Part 2: a new taxonomy for information science. **Journal of the American Society for Information Science and Technology,** v. 54, n. 8, p. 771-781, 2003.

HAY, WILLIAM W *et al.* **CURRENT Diagnosis & Treatment: Pediatrics**. 22thvedition. New York: McGraw-Hill Education, 2014. p.1328. ISBN: 007182734X
978-0071827348.

HAYES, PAT *et al.* COE: Tools for collaborative ontology development and reuse. 2005, 2005.

HAYES, PAT; SAAVEDRA, RAUL; REICHHERZER, THOMAS. A collaborative development environment for ontologies (CODE). 2003, 2003. p.139.

HÄYRINEN, KRISTIINA; SARANTO, KAIJA; NYKÄNEN, PIRKKO. Definition, structure, content, use and impacts of electronic health records: A review of the research literature. **International Journal of Medical Informatics,** v. 77, n. 5, p. 291-304, 5// 2008. DOI: http://doi.org/10.1016/j.ijmedinf.2007.09.001

HEALTHIT.GOV. Better Information Means Safer Health Care. 2017. Disponível online em: < https://www.healthit.gov/patients-families/better-information-means-safer-health-care >. Accesso em: 6/21/2017.

HEGDE, DINESH S. **Essays on Research Methodology**. New Delhi, INDIA: Springer India, 2015.

HELMS, MARILYN M.; MOORE, RITA; AHMADI, MOHAMMAD. Information technology (IT) and the healthcare industry: A SWOT analysis. **Medical Informatics: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications,** v. 134, 2008.

HENNIG, BORIS. What is formal ontology. In: Munn, K. and Smith, B. (Ed.). **Applied Ontology. An Introduction**. Frankfurt, Germany: Ontos Verlag, 2008. p.39-56.

HEPP, MARTIN. Ontologies: State of the art, business potential, and grand challenges. In: (Ed.). **Ontology Management**: Springer, 2008. p.3-22.

HERRE, HEINRICH. General Formal Ontology (GFO): A foundational ontology for conceptual modelling. In: (Ed.). **Theory and applications of ontology: computer applications**: Springer, 2010. p.297-345.

HERRE, HEINRICH *et al.* **General formal ontology (GFO): A foundational ontology integrating objects and processes**. Institute of Medical Informatics, Statistics and Epidemiology (IMISE) at University of Leipzig. Leipzig, Germany, p.75. 2006 Disponível online em: < http://www.onto-med.de/Archiv/ontomed2002/en/publications/scientific-reports/om-report-no8.pdf >.

HHS, UNITED STATES DEPARTMENT OF HEALTH AND HUMAN SERVICES HHS; HRSA, HEALTH RESOURCES AND SERVICES ADMINISTRATION. **Women's Health USA 2011**. U.S. Department of Health and Human Services. Rockville, Maryland, USA, p.80. 2011 Disponível online em: < https://mchb.hrsa.gov/whusa11/more/downloads/pdf/w11.pdf >.

HIQA, IRELAND HEALTH INFORMATION AND QUALITY AUTHORITY. **National Standard for Patient Discharge Summary Information**. Ireland**:** 42 p. 2013. Disponível online em: <https://www.hiqa.ie/sites/default/files/2017-01/National-Standard-Patient-Discharge-Summary.pdf >.

HJØRLAND, BIRGER. Nine principles of knowledge organization. In: Albrechtsen, H. and Oernager, S., Third International Isko Conference, 1994, Copenhagen. Frankfurt/Main: Indeks Verlag, 20-24 June 1994. p.91-100.

______. Fundamentals of knowledge organization. **Knowledge organization,** v. 30, n. 2, p. 87-111, 2003.

______. Semantics and knowledge organization. **Annual review of information science and technology,** v. 41, n. 1, p. 367-405, 2007.

______. What is knowledge organization (KO)? **Knowledge organization. International journal devoted to concept theory, classification, indexing and knowledge representation**, 2008.

HODGE, GAIL. **Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files**. ERIC, 2000.

HOFFMAN, BARBARA *et al.* **Williams Gynecology**. 2nd edition. New York: McGraw Hill Professional, 2012. 1352.

HUNTER, AMY *et al.* An ontology of human developmental anatomy. **Journal of anatomy,** v. 203, n. 4, p. 347-355, 2003.

JAMOOM, ERIC *et al.* Physician adoption of electronic health record systems: United States, 2011. **NCHS data brief,** v. 98, n. July, 2012.

JANSEN, LUDGER. Categories: The Top-Level Ontology. In: Munn, K. and Smith, B. (Ed.). **Applied Ontology: An Introduction**: Ontos Verlag, 2009a. p.173-196. ISBN 978-3938793985.

______. Classifications. In: Munn, K. and Smith, B. (Ed.). **Applied Ontology: An Introduction**: Ontos Verlag, 2009b. p.159-172. ISBN 978-3938793985.

JASPER, ROBERT; USCHOLD, MIKE. A framework for understanding and classifying ontology applications. 1999, 1999. p.16-21.

JONES, DEAN; BENCH-CAPON, TREVOR; VISSER, PEPIJN. Methodologies for ontology development. 1998.

JONES, M. L. Application of systematic review methods to qualitative research: practical issues. **Journal of Advanced Nursing,** v. 48, n. 3, p. 271-278 8p, 2004. DOI: 10.1111/j.1365-2648.2004.03196.x

KALFOGLOU, YANNIS; SCHORLEMMER, MARCO. Ontology mapping: the state of the art. **The Knowledge Engineering Review,** v. 18, n. 1, p. 1-31, 2003.

KITCHENHAM, BARBARA. Procedures for Performing Systematic Reviews. **(July 2004)**, 2009-01-24T18:45:28-00:00 2009.

KLÜCK, MARIZA MACHADO; GUIMARÃES, JOSÉ RICARDO. Sumário eletrônico de alta: garantindo a continuidade da assistência ao paciente através da informação. **Informática Pública**, p. 123-137, 1999.

KUMAR, RANJIT. **Research methodology: a step-by-step guide for beginners**. 3rd edition. SAGE Publications, 2011.

LAPELLE, NANCY R. *et al.* Identifying strategies to improve access to credible and relevant information for public health professionals: a qualitative study. **BMC Public Health,** v. 6, n. 1, p. 89, 2006.

LARSEN, WILLIAM J. **Human embriology**. third. Churchill Livingstone, 2001.

LIYANAGE, HARSHANA; KRAUSE, PAUL; DE LUSIGNAN, SIMON. Using ontologies to improve semantic interoperability in health data. **Journal of Innovation in Health Informatics,** v. 22, n. 2, p. 309-315, 2015.

LOWE, E. JONATHAN. **The four-category ontology: A metaphysical foundation for natural science**. Oxford University Press, 2006.

LOWE, HENRY J.; BARNETT, G. OCTO. Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. **Jama,** v. 271, n. 14, p. 1103-1108, 1994.

LUDWICK, D. A.; DOUCETTE, JOHN. Adopting electronic medical records in primary care: Lessons learned from health information systems implementation experience in seven countries. **International Journal of Medical Informatics,** v. 78, n. 1, p. 22-31, 1// 2009. DOI: http://doi.org/10.1016/j.ijmedinf.2008.06.005

MACKENZIE, NOELLA; KNIPE, SALLY. Research dilemmas: Paradigms, methods and methodology. **Issues in educational research,** v. 16, n. 2, p. 193-205, 2006.

MARCONI, MARINA DE ANDRADE; LAKATOS, EVA MARIA. **Fundamentos de metodologia científica**. 5ª. São Paulo: Atlas, 2003. 4150-465.

______. **Metodologia científica**. 5th Edition. São Paulo: Atlas 2007.

MARQUES, CONSUELO PENHA CASTRO *et al.* **Redes de atenção à saúde: a Rede Cegonha**. Universidade Federal do Maranhão, Universidade Aberta do SUS. São Luís, Maranhão, Brasil, p.43. 2015 Disponível online em: < http://www.multiresidencia.com.br/site/assets/uploads/kcfinder/files/REDE%20CEGONHA.pdf >.

MASCARDI, VIVIANA; CORDÌ, VALENTINA; ROSSO, PAOLO. A Comparison of Upper Ontologies. In: Baldoni, M. *et al*, WOA - 8th Workshop dagli Oggetti agli Agenti, 2007, Genova, Italy. 24-25, September. p.55-64.

MASON, JENNIFER. **Qualitative researching**. 2nd Edition. SAGE Publications, 2002.

MASSAD, EDUARDO; MARIN, HEIMAR DE FÁTIMA; NETO, RAYMUNDO SOARES DE AZEVEDO. **O prontuário eletrônico do paciente na assistência, informação e conhecimento médico**. 2003. 213.

MCDONALD, CLEM *et al.* Logical observation identifiers names and codes (LOINC®) users' guide. **Indianapolis: Regenstrief Institute**, 2004.

MCGLYNN, ELIZABETH A. *et al.* Health information systems. **Design Issues and Analytic Application, RAND Health Corporation**, 1998.

MCINERNY, THOMAS K. *et al.* **American Academy of Pediatrics textbook of pediatric care**. American Academy of Pediatrics, 2009.

MENA, EDUARDO *et al.* Domain specific ontologies for semantic information brokering on the global information infrastructure. In: Guarino, N., 1st International Conference of Formal Ontology in Information Systems (FOIS), 1998, Trento, Italy. Amsterdam: IOS Press, 1998. p.269-283.

MENACHEMI, NIR; COLLUM, TALEAH H. Benefits and drawbacks of electronic health record systems. **Risk Manag Healthc Policy,** v. 4, p. 47-55, 2011.

MENDONÇA, FABRICIO MARTINS. **Ontoforinfoscience: metodologia para construção de ontologias pelos cientistas da informação - Uma aplicação prática no desenvolvimento**

da ontologia sobre componentes do sangue humano (HEMONTO). 2015. Tese de Doutorado (Doctor in Information Science). UFMG. *Published in* Português.

MIAN, PAULA *et al.* A systematic review process for software engineering. 2nd Experimental Software Engineering Latin American Workshop, 2005, Uberlândia, MG, Brazil. 2005.

MILLER, ERIC. An introduction to the resource description framework. **Bulletin of the Association for Information Science and Technology,** v. 25, n. 1, p. 15-19, 1998.

MILLER, PAUL. **Interoperability: What Is It and Why Should I Want It?**: Ariadne 2000. Disponível online em: <http://www.ariadne.ac.uk/issue24/interoperability >. Accesso em: Issue 24.

MIZOGUCHI, RIICHIRO. YAMATO: yet another more advanced top-level ontology. In: Taylor, K.;Meyer, T. and Orgun, M., Sixth Australasian Ontology Workshop, 2010, Adelaide, Australia. December 7. p.1-16.

MONTIEL-PONSODA, ELENA; ESPINOZA, MAURICIO. Ontology Localization. In: Gómez-Pérez, A.;Motta, E. and Suárez-Figueroa, M. C. (Ed.). **NeOn Methodology in a Nutshell**: NeOn Project, 2010.

MUNGALL, CHRISTOPHER J. *et al.* Uberon, an integrative multi-species anatomy ontology. **Genome Biology,** v. 13, n. 1, p. R5, 2012. DOI: https://doi.org/10.1186/gb-2012-13-1-r5

MUSEN, M. *et al.* BioPortal: ontologies and data resources with the click of a mouse. **Nucleic acids research,** v. 37, n. suppl_2, p. W170 - W173, 2008 2009. DOI: 10.1093/nar/gkp440

MUSEN, MARK A. Foundational Model of Anatomy. In: Dubitzky, W. *et al* (Ed.). **Encyclopedia of Systems Biology**. New York, NY: Springer New York, 2013. p.757-757. ISBN: 978-1-4419-9863-7.

NEUMAN, LAWRENCE W. **Social research methods: Qualitative and quantitative approaches**. 7th Edition. Pearson Education Limited, 2014.

NOY, N. F. *et al.* BioPortal: ontologies and integrated data resources at the click of a mouse. In: (Ed.). **Nucleic acids research**, 2009.

NOY, NATALYA F; MCGUINNESS, DEBORAH L. **Ontology development 101: A guide to creating your first ontology**: Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880 2001.

NOY, NATALYA F. *et al.* Protege-2000: an open-source ontology-development and knowledge-acquisition environment. AMIA 2003 Symposium, 2003, 2003. p.953.

NOY, NATALYA FRIDMAN; MUSEN, MARK A. SMART: Automated support for ontology merging and alignment. 12th Workshop on Knowledge Acquisition, Modelling, and Management (KAW'99), 1999, Banf, Canada. 1999.

______. Algorithm and tool for automated ontology merging and alignment. 2000, 2000.

NOY, NATALYA FRIDMAN *et al.* BioPortal: A Web Repository for Biomedical Ontologies and Data Resources. In: Joshi, C. B. A., 7th International Semantic Web Conference (ISWC2008), 2008, Karlsruhe, Germany. CEUR-WS.org, October 13th, 2008.

OBO_FOUNDRY, OBO FOUNDRY TECHNICAL GROUP. The OBO Foundry. 2017a. Disponível online em: < http://www.obofoundry.org/ >. Accesso em: 03/12/2017.

______. Policy for OBO namespace and associated PURL requests. 2017b. Disponível online em: < http://obofoundry.org/docs/Policy_for_OBO_namespace_and_associated_PURL_requests.html >. Accesso em: 03/14/2017.

OBRST, LEO. Ontologies for semantically interoperable systems. Proceedings of the twelfth international conference on Information and knowledge management, 2003, ACM. p.366-369.

OBRST, LEO *et al.* The evaluation of ontologies: Toward Improved Semantic Interoperability. In: Baker, C. and Cheung, K.-H. (Ed.). **Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences**. New York: Springer Verlag,: Springer Verlag, 2006. cap. 6, p.139-158. ISBN: 0387484361.

ODP-PC, ONTOLOGY DESIGN PATTERNS PUBLIC CATALOG. Ontology Design Patterns Public Catalog 2017. Disponível online em: < http://www.gong.manchester.ac.uk/odp/html/ >. Accesso em: April 15th.

ODPA, ASSOCIATION FOR ONTOLOGY DESIGN & PATTERNS. Ontology Design Pattern Catalogue. Ontology Design Patterns Portal, 2017a. Disponível online em: < http://ontologydesignpatterns.org/wiki/Category:OntologyDesignPattern >. Accesso em: April 14th.

______. Ontology Design Pattern types. Ontology Design Patterns Portal, 2017b. Disponível online em: < http://ontologydesignpatterns.org/wiki/OPTypes >. Accesso em: April 14th.

OLIVEIRA, JACQUELINE PAWLOWSKI; ALMEIDA, MAURICIO BARCELLOS; QUINTELA, ERIKA LEITE. Uma visão geral sobre fontes de informação em saúde. I Congresso ISKO Espanha e Portugal and XI Congresso ISKO Espanha, 2013, Porto, Portugal. 2013. p.993-1008.

OLS, ONTOLOGY LOOKUP SERVICE. EMBL-EBI Ontology Lookup Service. 2017. Disponível online em: < http://www.ebi.ac.uk/ols/index >.

ONC, OFFICE OF THE NATIONAL COORDINATOR FOR HEALTH INFORMATION TECHNOLOGY. **How to Implement EHRs** 2017a. Disponível online em: <https://www.healthit.gov/providers-professionals/ehr-implementation-steps >.

______. **Learn EHR Basics** 2017b. Disponível online em: <https://www.healthit.gov/providers-professionals/learn-ehr-basics >.

______. **What are the differences between electronic medical records, electronic health records, and personal health records?** 2017c. Disponível online em: <https://www.healthit.gov/providers-professionals/faqs/what-are-differences-between-electronic-medical-records-electronic >.

ONG, EDISON *et al.* Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration. **Nucleic acids research,** v. Jan 4, n. 45(D1):D347-D352, 2017.

ONTOBEE. Ontobee linked data server. 2017. Disponível online em: < http://www.ontobee.org/ >. Accesso em: 03/12/2017.

ORGUN, BHAVNA *et al.* Approaches for semantic interoperability between domain ontologies. **Expert Systems,** v. 25, n. 3, p. 179-196, 2008. DOI: 10.1111/j.1468-0394.2008.00461.x

OUKSEL, ARIS M.; SHETH, AMIT. Semantic interoperability in global information systems. **ACM Sigmod Record,** v. 28, n. 1, p. 5-12, 1999.

PAN, JEFF Z. Resource description framework. In: (Ed.). **Handbook on ontologies**: Springer, 2009. p.71-90.

PATERSON, J. M.; ALLEGA, R. L. Improving communication between hospital and community physicians. Feasibility study of a handwritten, faxed hospital discharge summary. Discharge Summary Study Group. **Canadian Family Physician,** v. 45, p. 2893-2899, 1999.

PEASE, ADAM; NILES, IAN; LI, JOHN. The suggested upper merged ontology: A large ontology for the semantic web and its applications. 2002.

PINTO, H. SOFIA; GÓMEZ-PÉREZ, ASUNCIÓN; MARTINS, JOÃO P. Some issues on ontology integration. 1999, IJCAI and the Scandinavian AI Societies. CEUR Workshop Proceedings, 1999.

PINTO, H. SOFIA; STAAB, STEFFEN; TEMPICH, CHRISTOPH. DILIGENT: Towards a fine-grained methodology for DIstributed, Loosely-controlled and evolvInG Engineering of oNTologies. In: R., L. D. M. and L., S., 16th European Conference on Artificial Intelligence, 2004, IOS Press, 2004. p.393-397.

PINTO, HELENA SOFIA; MARTINS, JOÃO P. A methodology for ontology integration. 2001, ACM, 2001. p.131-138.

PINTO, VIRGÍNIA BENTES. Prontuário eletrônico do paciente: documento técnico de informação e comunicação do domínio da saúde. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação,** v. 11, n. 21, p. 34-48, 2007. DOI: 10.5007/1518-2924.2006

PLANTIER, MORGANE *et al.* Does adoption of electronic health records improve the quality of care management in France? Results from the French e-SI (PREPS-SIPS) study. **International Journal of Medical Informatics**, 2017.

POISSANT, LISE *et al.* The Impact of Electronic Health Records on Time Efficiency of Physicians and Nurses: A Systematic Review. **Journal of the American Medical Informatics Association,** v. 12, n. 5, p. 505-516, 9// 2005. DOI: http://doi.org/10.1197/jamia.M1700

POVEDA-VILLALÓN, MARÍA. **Ontology Evaluation: a pitfall-based approach to ontology diagnosis**. 2016. PhD Thesis (Doctoral). Departmento de Inteligencia Artificial
Escuela Técnica Superior de Ingenieros Informáticos, Universidad Politécnica de Madrid.

POVEDA-VILLALÓN, MARÍA; GÓMEZ-PÉREZ, ASUNCIÓN; SUÁREZ-FIGUEROA, MARI CARMEN. OOPS! (Ontology Pitfall Scanner!): an on-line tool for ontology evaluation. **International Journal on Semantic Web and Information Systems,** v. 10, p. 7+, 2014 April-June

// 2014.

POVEDA-VILLALÓN, MARÍA; SUÁREZ-FIGUEROA, MARI CARMEN. OOPS!–OntOlogy Pitfalls Scanner! , 2012.

POVEDA-VILLALÓN, MARÍA *et al.* OOPS!(OntOlogy Pitfall Scanner!): supporting ontology evaluation on-line. **Semantic Web Journal – Interoperability, Usability, Applicability**, p. 1-9, 2009.

POVEDA-VILLALÓN, MARÍA; SUÁREZ-FIGUEROA, MARI CARMEN; GÓMEZ-PÉREZ, ASUNCIÓN. Common pitfalls in ontology development. In: (Ed.). **Current Topics in Artificial Intelligence**: Springer, 2010a. p.91-100. ISBN: 364214263X.

______. A double classification of common pitfalls in ontologies. Workshop on Ontology Quality (OntoQual 2010), Co-located with EKAW 2010, 2010b, Lisbon, Portugal. 15th October. p.1-12.

______. Did you validate your ontology? OOPS! , 2012a.

______. Validating ontologies with oops! In: (Ed.). **Knowledge Engineering and Knowledge Management**: Springer, 2012b. p.267-281. ISBN: 3642338755.

QIAOLI, ZHU *et al.* Global ontology research progress: a bibliometric analysis. **Aslib Journal of Information Management,** v. 67, n. 1, p. 27-54, 2015/01/19 2015. DOI: 10.1108/AJIM-05-2014-0061

QUEIROGA, RODRIGO. Sumário de alta de internação obstétrica no Hospital das Clinicas da UFMG: Modelo de informação. 2014.

RAVAL, AMISH N.; MARCHIORI, GORDON E.; ARNOLD, J. M. Improving the continuity of care following discharge of patients hospitalized with heart failure: is the discharge summary adequate? **The Canadian journal of cardiology,** v. 19, n. 4, p. 365-370, 2003.

RECTOR, A. L.; NOWLAN, W. A.; KAY, S. Foundations for an electronic medical record. **Methods Inf Med,** v. 30, n. 3, p. 179-186, 1991.

RECTOR, ALAN L. Clinical terminology: why is it so hard? **Methods of information in medicine,** v. 38, n. 4/5, p. 239-252, 1999.

RECTOR, ALAN; ROGERS, JEREMY. Ontological and practical issues in using a description logic to represent medical concept systems: Experience from GALEN. In: (Ed.). **Reasoning Web**: Springer, 2006. p.197-231.

REED, STEPHEN L.; LENAT, DOUGLAS B. Mapping ontologies into Cyc. AAAI 2002 Conference Workshop on Ontologies For The Semantic Web, 2002. p.1-6.

REIS, ZILMA SILVEIRA NOGUEIRA; CORREIA, RICARDO JOÃO CRUZ; PEREIRA, ALTAMIRO DA COSTA. Sistemas eletrônicos de informação na assistência e pesquisa em saúde da mulher: para quando um maior envolvimento dos profissionais de saúde. **Rev Bras Ginecol Obstet,** v. 33, n. 3, p. 107-10, 2011.

RÖHL, JOHANNES; JANSEN, LUDGER. Representing dispositions. **Journal of Biomedical Semantics,** v. 2, n. 4, p. S4, 2011/08/09 2011. DOI: 10.1186/2041-1480-2-S4-S4

ROMM, FREDRIC J.; PUTNAM, SAMUEL M. The Validity of the Medical Record. **Medical Care,** v. 19, n. 3, p. 310-315, 1981.

ROSENBLOOM, S. TRENT *et al.* Interface Terminologies: Facilitating Direct Entry of Clinical Data into Electronic Health Record Systems. **Journal of the American Medical Informatics Association : JAMIA,** v. 13, n. 3, p. 277-288, May-Jun
08/31/received
01/09/accepted 2006. DOI: 10.1197/jamia.M1957

ROSSE, CORNELIUS; MEJINO, JOSÉ L. V. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. **Journal of biomedical informatics,** v. 36, n. 6, p. 478-500, 2003.

ROSSE, CORNELIUS; MEJINO JR, JOSÉ L. V. The foundational model of anatomy ontology. In: (Ed.). **Anatomy Ontologies for Bioinformatics**: Springer, 2008. p.59-117.

ROUKEMA, JOLT *et al.* Paper versus computer: feasibility of an electronic medical record in general pediatrics. **Pediatrics,** v. 117, n. 1, p. 15-21, 2006.

RUBIN, DANIEL L. *et al.* BioPortal: A Web Portal to Biomedical Ontologies. AAAI Spring Symposium: Symbiotic Relationships between Semantic Web and Knowledge Engineering, 2008, Palo Alto, California, USA. 4: The AAAI Press, 2008. p.74-77.

SALVADORES, MANUEL *et al.* BioPortal as a dataset of linked biomedical ontologies and terminologies in RDF. **Semantic Web,** v. 4, n. 3, p. 277-284, 2013. DOI: 10.3233/Sw-2012-0086

SARACEVIC, TEFKO. Interdisciplinary nature of information science. **Ciência da Informação,** v. 24, n. 1, p. 36-44, 1995. DOI: 10.18225/ci.inf..v24i1.608

______. Information science. **Journal of the American Society for Information Science,** v. 50, n. 12, p. 1051-1063, 1999. DOI: 10.1002/(SICI)1097-4571(1999)50:12<1051::AID-ASI2>3.0.CO;2-Z

SCHOBER, DANIEL *et al.* Survey-based naming conventions for use in OBO Foundry ontology development. **BMC bioinformatics,** v. 10, n. 1, p. 125, 2009// 2009. DOI: 10.1186/1471-2105-10-125

SCHULZ, S. *et al.* **Guideline on Developing Good Ontologies in the Biomedical Domain with Description Logics**. December 11th, 2012, p.85. 2012 Disponível online em: < http://www.iph.uni-rostock.de/fileadmin/PHF_Philosophie/media/goodod/GoodOD-Guideline_v1_2012.pdf >.

SCHULZ, STEFAN; LÓPEZ-GARCÍA, PABLO. Big Data, medizinische Sprache und biomedizinische Ordnungssysteme. **Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz,** v. 58, n. 8, p. 844–852, 2015. DOI: 10.1007/s00103-015-2190-x

SHADBOLT, NIGEL; BERNERS-LEE, TIM; HALL, WENDY. The semantic web revisited. **IEEE Intelligent systems,** v. 21, n. 3, p. 96-101, 2006.

SHETH, AMIT P. Changing focus on interoperability in information systems: from system, syntax, structure to semantics. In: (Ed.). **Interoperating geographic information systems**: Springer, 1999. p.5-29.

SILVA, DANIELA LUCAS DA. **Ontologias para representação de documentos multimídia: análise e modelagem**. 2014. (Doctor in Information Science). Federal University of Minas Gerais at Brazil.

SILVA, DANIELA LUCAS DA; SOUZA, RENATO ROCHA; ALMEIDA, MAURÍCIO BARCELLOS. Ontologies and controlled vocabulary: comparison of building methodologies. In: Khosrow-Pour, M. (Ed.). **Computer Engineering: Concepts, Methodologies, Tools and Applications**: IGI Global, v.1, 2011. cap. 4, p.46-60. ISBN: 1613504578.

SIMON, JONATHAN; FIELDING, JAMES MATTHEW; SMITH, BARRY. Using Philosophy to Improve the Coherence and Interoperability of Applications Ontologies: A Field Report on the Collaboration of IFOMIS and L&C. In: Büchel, G.;Klein, B. and Roth-Berghofer, T., First International Workshop on Philosophy and Informatics - WSPI '04, 2004, Cologne, Germany. CEUR Workshop Proceedings, March 31 - April 1.

SIRINTRAPUN, S. JOSEPH; ARTZ, DAVID R. Health Information Systems. **Clinics in Laboratory Medicine,** v. 36, n. 1, p. 133-152, 2016/03/01/ 2016. DOI: http://dx.doi.org/10.1016/j.cll.2015.09.012

SMITH, BARRY. An essay in formal ontology. **Grazer Philosophische Studien,** v. 6, p. 39-62, 1978.

______. Logic and formal ontology. In: (Ed.). **Husserl's Phenomenology: A Textbook**: University Press of America, v.23, 1989.

______. Basic Concepts of Formal Ontology. In: Guarino, N. (Ed.). **Formal Ontology in Information Systems**: IOS Press, 1998a. p.19-28.

______. An Introduction to Ontology. In: Peuquet, D.;Smith, B. and Brogaard, B. O. (Ed.). **The Ontology of Fields: Report of the Specialist Meeting held under the auspices of the Varenius Project**. Bar Harbor, Maine: National Center for Geographic Information and Analysis, 1998b. cap. 4, p.9-14.

______. The logic of biological classification and the foundations of biomedical ontology. Invited Papers from the 10th International Conference in Logic Methodology and Philosophy of Science, Oviedo, Spain, 2003a. p.19-25.

______. Ontology. In: Floridi, L. (Ed.). **The Blackwell Guide to the Philosophy of Computing and Information**. Oxford: Blackwell, 2003b. cap. 11, p.155-166. ISBN: 0631229183
0470756764
9780470756768.

______. **Ontology and Information Systems**. p.93. 2003c

______. Beyond concepts: ontology as reality representation. Proceedings of the third international conference on formal ontology in information systems (FOIS 2004), 2004. p.73-84.

______. New desiderata for biomedical terminologies. In: Munn, K. and Smith, B. (Ed.). **Applied Ontology: An Introduction**: Ontos Verlag, 2008a. p.83-107. ISBN ISBN-13: 978-3938793985
ISBN-10: 3938793988.

______. Ontology (Science). FOIS, 2008b. p.21-35.

______. On Classifying Material Entities in Basic Formal Ontology. 2012.

SMITH, BARRY *et al.* **Basic Formal Ontology 2.0: Specification and User's Guide**. June 26th, 2015, p.97. 2015 Disponível online em: < https://github.com/BFO-ontology/BFO/raw/master/docs/bfo2-reference/BFO2-Reference.pdf >.

SMITH, BARRY *et al.* The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. **Nature Biotechnology,** v. 25, 2007// 2007a. DOI: 10.1038/nbt1346

______. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. **Nature Biotechnology,** v. 25, n. 11, p. 1251-1255, 11/07/2007 2007b. DOI: doi:10.1038/nbt1346

SMITH, BARRY; BROCHHAUSEN, MATHIAS. Putting biomedical ontologies to work. **Methods of information in medicine,** v. 49, n. 2, p. 135, 2010.

SMITH, BARRY; BROGAARD, BERIT. A unified theory of truth and reference. **Logique et Analyse**, p. 49-93, 2000.

SMITH, BARRY; CEUSTERS, WERNER. Ontological realism: A methodology for coordinated evolution of scientific ontologies. **Applied ontology,** v. 5, n. 3-4, p. 139-188, November 15th, 2010 2010. DOI: 10.3233/AO-2010-0079

SMITH, BARRY *et al.* Relations in biomedical ontologies. **Genome Biology,** v. 6, n. 5, 28/04/2005 2005. DOI: info:pmid/15892874

SMITH, BARRY; GRENON, PIERRE. The Cornucopia of Formal-Ontological Relations. **Dialectica,** v. 58, n. 3, p. 279-296, 2004.

SMITH, BARRY; KUMAR, ANAND; BITTNER, THOMAS. Basic formal ontology for bioinformatics. **Journal of Information Systems**, p. 1-16, 2005.

SMITH, BARRY *et al.* Towards a Reference Terminology for Ontology Research and Development in the Biomedical Domain. In: Bodenreider, O., 2nd International Workshop on Formal Biomedical Knowledge Representation: "Biomedical Ontology in Action" (KR-MED 2006), 2006, Baltimore, Maryland, USA. CEUR Workshop Proceedings, November 8. p.57-65.

SMITH, BARRY; MULLIGAN, KEVIN. Framework for formal ontology. **Topoi,** v. 2, n. 1, p. 73-85, 1983.

SMITH, BARRY; ROSSE, CORNELIUS. The role of foundational relations in the alignment of biomedical ontologies. **Medinfo,** v. 11, n. Pt 1, p. 444-8, 2004.

SMITH, BARRY; WELTY, CHRISTOPHER. Ontology: Towards a new synthesis. 2nd International Conference on Formal Ontology in Information Systems (FOIS 2001), 2001, Ogunquit, Maine, USA. ACM Press, October 17-19. p.3-9.

SØERGEL, DAGOBERT. **Organizing information: principles of data base and retrieval systems**. Elsevier, 1985.

______. The rise of ontologies or the reinvention of classification. **Journal of the Association for Information Science and Technology,** v. 50, n. 12, p. 1119, 1999.

______. **Thesauri and ontologies in digital libraries: tutorial**. 2002

______. Digital libraries and knowledge organization. In: (Ed.). **Semantic digital libraries**: Springer, 2009a. p.9-39.

______. Knowledge Organization Systems: Overview. 2009b.

SOUZA, RENATO ROCHA; TUDHOPE, DOUGLAS; ALMEIDA, MAURICIO BARCELLOS DE. The KOS spectra: A tentative typology of knowledge organization systems. In: Gnoli, C. and Mazzocchi, F., 11th International ISKO Conference, 2010, Rome, Italy. Ergon Verlag. p.122-128.

______. Towards a taxonomy of KOS: Dimensions for classifying Knowledge Organization Systems. **Knowledge organization,** v. 39, n. 3, p. 179-192, 2012.

STADTMÜLLER, STEFFEN; HARTH, ANDREAS; GROBELNIK, MARKO. Accessing information about linked data vocabularies with vocab. cc. In: Juanzi Li *et al* (Ed.). **Semantic Web and Web Science**: Springer, 2013. p.391-396.

STAUSBERG, JÜRGEN *et al.* Comparing Paper-based with Electronic Patient Records: Lessons Learned during a Study on Diagnosis and Procedure Codes. **Journal of the American Medical Informatics Association : JAMIA,** v. 10, n. 5, p. 470-477, Sep-Oct
11/11/received
04/19/accepted 2003. DOI: 10.1197/jamia.M1290

STEARNS, MICHAEL Q. *et al.* SNOMED clinical terms: overview of the development process and project status. AMIA Symposium 2001, 2001, Washington DC, USA. American Medical Informatics Association, November 3-7. p.662-666.

STOUTENBURG, SUZETTE KRUGER. **Advancing ontology alignment: new methods for biomedical ontology alignment using non equivalence relations**. 2009. 202 Dissertation (Ph.D. in Engineering, Focus in Computer Science). Department of Computer Science, University of Colorado at Colorado Springs.

SUÁREZ-FIGUEROA, MARI CARMEN. **NeOn Methodology for building ontology networks: specification, scheduling and reuse**. 2010a. 268 (PhD thesis). Departamento de Inteligência Artificial, Universidad Politécnica de Madrid, Madri. *Published in* Inglês.

______. **NeOn Methodology for building ontology networks: specification, scheduling and reuse**. 2010b. 268 (Doctoral thesis). Inteligência Artificial, Universidad Politécnica de Madrid, Madri. *Published in* English.

SUÁREZ-FIGUEROA, MARI CARMEN; GÓMEZ-PÉREZ, ASUNCIÓN; FERNÁNDEZ-LÓPEZ, MARIANO. The NeOn methodology for ontology engineering. In: Suárez-Figueroa, M. C. *et al* (Ed.). **Ontology engineering in a networked world**: Springer Berlin Heidelberg, 2012. cap. 2, p.9-34. ISBN: ISBN: 978-3-642-24793-4 (Print) 978-3-642-24794-1 (Online).

______. The NeOn Methodology framework: A scenario-based methodology for ontology development. **Applied ontology,** v. 10, n. 2, p. 107-145, 2015.

SUÁREZ-FIGUEROA, MARI CARMEN; KAMEL, MOUNA; POVEDA-VILLALÓN, MARÍA. Benefits of natural language techniques in ontology evaluation: The OOPS! case. 2013.

SUÁREZ-FIGUEROA, MARICARMEN; GÓMEZ-PÉREZ, ASUNCIÓN; VILLAZÓN-TERRAZAS, BORIS. How to Write and Use the Ontology Requirements Specification Document. In: Meersman, R.;Dillon, T. and Herrero, P. (Ed.). **On the Move to Meaningful Internet Systems: OTM 2009**: Springer Berlin Heidelberg, v.5871, 2009. cap. 16, p.966-982. (Lecture Notes in Computer Science). ISBN: 978-3-642-05150-0.

SURE, YORK; STAAB, STEFFEN; STUDER, RUDI. On-to-knowledge methodology (OTKM). In: Staab, S. and Studer, R. (Ed.). **Handbook on Ontologies**. Berlin: Springer Berlin Heidelberg, 2004. cap. 6, p.117-132. (International Handbooks on Information Systems). ISBN: Print ISBN: 978-3-662-11957-0
Online ISBN: 978-3-540-24750-0.

SWOOGLE. Swoogle Semantic Web Search Engine. 2017. Disponível online em: < http://swoogle.umbc.edu/2006/ >.

TUFO, HENRY M.; SPEIDEL, JOSEPH J. Problems with Medical Records. **Medical Care,** v. 9, n. 6, p. 509-517, 1971.

UKOLN. **Interoperability Focus: About** 1999. Disponível online em: <http://www.ukoln.ac.uk/interop-focus/about/ >.

______. **Interoperability Focus: Looking at Interoperability** 2005. Disponível online em: <http://www.ukoln.ac.uk/interop-focus/about/flyer-interoperability.pdf >.

UNICODE_CONSORTIUM, THE UNICODE CONSORTIUM. What is Unicode? , 2017. Disponível online em: < http://www.unicode.org/standard/WhatIsUnicode.html >. Accesso em: 7/4/2017.

USCHOLD, MICHAEL; GRUNINGER, MICHAEL. Ontologies and semantics for seamless connectivity. **ACM Sigmod Record,** v. 33, n. 4, p. 58-64, 2004. DOI: 10.1145/1041410.1041420

USCHOLD, MICHAEL; KING, MARTIN. **Towards a methodology for building ontologies**. Citeseer, 1995.

USCHOLD, MIKE; GRUNINGER, MICHAEL. Ontologies: Principles, methods and applications. **The Knowledge Engineering Review,** v. 11, n. 02, p. 93-136, 1996. DOI: doi:10.1017/S0269888900007797

VEZINA, BRAD. Universals and Particulars: Aristotle's Ontological Theory and Criticism of the Platonic Forms. **Undergraduate Review,** v. 3, n. 1, p. 101-103, 2007.

VICKERY, BRIAN C. Ontologies. **Journal of Information Science,** v. 23, n. 4, p. 277-286, 1997. DOI: 10.1177/016555159702300402

VIGO, MARKEL *et al.* Overcoming the pitfalls of ontology authoring: Strategies and implications for tool design. **International journal of human-computer studies,** v. 72, n. 12, p. 835-845, 2014. DOI: http://doi.org/10.1016/j.ijhcs.2014.07.005

VOCAB.CC. vocab.cc - RDF vocabulary search and lookup. 2017.

VRANDEČIĆ, DENNY. **Ontology evaluation**. Springer, 2009.

W3C. **XML Schema**: World Wide Web Consortium 2000. Disponível online em: <https://www.w3.org/XML/Schema >.

______. **OWL Web Ontology Language Guide**. Smith, M. K.;Welty, C. and Mcguinness, D. L.: World Wide Web Consortium OWL Working Group. 2004 2004a. Disponível online em: <https://www.w3.org/TR/2004/REC-owl-guide-20040210/ >. Accesso em: February 10th.

______. **OWL Web Ontology Language Reference**. Bechhofer, S. *et al*: World Wide Web Consortium OWL Working Group. 2017 2004b. Disponível online em: <http://www.w3.org/TR/2004/REC-owl-ref-20040210/ >.

______. **Resource Description Framework (RDF): Concepts and Abstract Syntax**. Klyne, G.;Carroll, J. J. and Mcbride, B.: World Wide Web Consortium, OWL Working Group 2004c. Disponível online em: <https://www.w3.org/TR/2004/REC-rdf-concepts-20040210/ >.

______. **Namespaces in XML 1.1 (Second Edition)**: 11/10/2016 2006. Disponível online em: <https://www.w3.org/TR/xml-names11/ >.

______. **SPARQL Query Language for RDF**: Word Wide Web Consortium. 2017 2008. Disponível online em: <https://www.w3.org/TR/rdf-sparql-query/ >. Accesso em: 7/4/2017.

______. **A Direct Mapping of Relational Data to RDF**: Word Wide Web Consortium. 2016 2012a. Disponível online em: <https://www.w3.org/TR/rdb-direct-mapping/ >. Accesso em: June 17th, 2016.

______. **OWL 2 Web Ontology Language: Document Overview**. Group, W. C. O. W.: World Wide Web Consortium OWL Working Group. 2016 2012b. Disponível online em: <https://www.w3.org/TR/2012/REC-owl2-overview-20121211/ >. Accesso em: September 28th.

______. **OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax**. Motik, B.;Patel-Schneider, P. F. and Parsia, B.: World Wide Web Consortium OWL Working Group. 2016 2012c. Disponível online em: <http://www.w3.org/TR/2012/REC-owl2-syntax-20121211/ >. Accesso em: September 28th.

______. **SPARQL 1.1 Overview**: Word Wide Web Consortium. 2016 2013a. Disponível online em: <https://www.w3.org/TR/sparql11-overview/ >. Accesso em: June 17th, 2016.

______. **SPARQL 1.1 Query Language**: Word Wide Web Consortium. 2016 2013b. Disponível online em: <https://www.w3.org/TR/sparql11-query/ >. Accesso em: June 17th, 2016.

______. W3C Semantic web activity. 11th December, 2013 2013c. Disponível online em: < https://www.w3.org/2001/sw/ >. Accesso em: 6/23/2017.

______. **RDF 1.1 Concepts and Abstract Syntax**. Cyganiak, R.;Wood, D. and Lanthaler, M.: World Wide Web Consortium, OWL Working Group 2014a. Disponível online em: <https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/ >.

______. **RDF 1.1 Semantics**. Hayes, P. J. and Patel-Schneider, P. F.: W3C Recommendation 2014b. Disponível online em: <https://www.w3.org/TR/rdf11-mt/ >.

______. **RDF 1.1 Turtle**. Beckett, D. *et al*: W3C Recommendation 2014c. Disponível online em: <http://www.w3.org/TR/turtle/ >.

______. **RDF 1.1 XML Syntax**. Gandon, F. and Schreiber, G.: W3C Recommendation 2014d. Disponível online em: <http://www.w3.org/TR/rdf-syntax-grammar/ >.

______. **RDF 1.1: On Semantics of RDF Datasets**. Zimmermann, A.: W3C Working Group Note 2014e. Disponível online em: <https://www.w3.org/TR/rdf11-datasets/ >.

______. **RDF Schema 1.1**. Cyganiak, R.;Wood, D. and Lanthaler, M.: World Wide Web Consortium 2014f. Disponível online em: <https://www.w3.org/TR/rdf-schema/ >.

______. **Extensible Markup Language (XML)**: World Wide Web Consortium 2016. Disponível online em: <https://www.w3.org/XML/ >.

WALD JS, RIZK S, WEBB JR., HAQUE S, ET AL. . **Children's EHR Format Enhancement: Final Recommendation Report**. 2015 Disponível online em: < https://healthit.ahrq.gov/sites/default/files/docs/page/children-ehr-format-enhancement-final-recommendation-report.pdf >.

WAND, YAIR; WEBER, RON. Mario Bunge's Ontology as a formal foundation for information systems concepts. **Studies on Mario Bunge's Treatise, Rodopi, Atlanta**, p. 123-149, 1990.

WANG, SAMUEL J. *et al.* A cost-benefit analysis of electronic medical records in primary care. **The American Journal of Medicine,** v. 114, n. 5, p. 397-403, 4/1/ 2003. DOI: http://doi.org/10.1016/S0002-9343(03)00057-3

WATSON. Watson Search Engine. 2017. Disponível online em: < http://watson.kmi.open.ac.uk/WatsonWUI/ >.

WEAVER, WARREN. The mathematical theory of communication. **Scientific american,** v. 181, n. 1, p. 11-15, 1949.

______. Recent contributions to the mathematical theory of communication. **ETC: A Review of General Semantics,** v. 10, n. 4, p. 261-281, 07/1953 1953.

WELTY, CHRISTOPHER; GUARINO, NICOLA. Supporting ontological analysis of taxonomic relationships. **Data & Knowledge Engineering,** v. 39, n. 1, p. 51-74, 2001.

WHETZEL, PATRICIA L. *et al.* BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. **Nucleic acids research,** v. 39, n. suppl 2, p. W541-W545, July, 2011 2011. DOI: 10.1093/nar/gkr469

WIEDERHOLD, GIO; JANNINK, JAN. Composing diverse ontologies. 8th Working Conference on Database Semantics (DS-8), 1999, Rotorua, New Zealand.

WRIGHT, RICK W. *et al.* How to Write a Systematic Review. **Clinical Orthopaedics and Related Research,** v. 455, 2007.

XIANG, ZUOSHUANG *et al.* OntoFox: web-based support for ontology reuse. **BMC research notes,** v. 3, n. 1, p. 175, 2010.

XIANG, ZUOSHUANG *et al.* Ontobee: A Linked Data Server and Browser for Ontology Terms, Ontology. 2nd International Conference on Biomedical Ontology (ICBO2011), 2011, Buffalo, NY, USA. CEUR-WS.org, 2011. p.279-281.

______. Ontobee: A Linked Data Server that publishes RDF and HTML data simultaneously. **Semantic Web – Interoperability, Usability, Applicability an IOS Press Journal,** v. 261, 2013.

YIN, ROBERT K. **Case study research : design and methods**. 4th Edition. Los Angeles, CA, US: Sage Publications, 2009. 240.

APPENDIX 1: Catalog of pitfalls in ontology building

Table 20 - Catalog of ontology pitfalls

	Pitfall	Pitfall definition
1	Creating polysemous elements	An ontology element (class, object property or datatype property) whose identifier has different senses is included in the ontology to represent more than one conceptual idea or property.
2	Creating synonyms as classes	Several classes whose identifiers are synonyms are created and defined as equivalent (owl:equivalentClass) in the same namespace
3	Creating the relationship “is” instead of using "rdfs:subClassOf", "rdf:type" or "owl:sameAs"	The relationship “is” is created in the ontology instead of using OWL primitives for representing the subclass relationship (rdfs:subClassOf), class membership (rdf:type), or the equality between instances (owl:sameAs).
4	Creating unconnected ontology elements	Ontology elements (classes, object properties and datatype properties) are created isolated, with no relation to the rest of the ontology.
5	Defining wrong inverse relationships	Ontology elements (classes, object properties and datatype properties) are created isolated, with no relation to the rest of the ontology.
6	Including cycles in the hierarchy	Ontology elements (classes, object properties and datatype properties) P0are created isolated, with no relation to the rest of the ontology.
7	Merging different concepts in the same class	A class whose name refers to two or more different concepts is created.
8	Missing annotations	This pitfall consists in creating an ontology element and failing to provide human readable annotations attached to it. Consequently, ontology elements lack annotation properties that label them (e.g. rdfs:label, lemon:LexicalEntry, skos:prefLabel or skos:altLabel) or that define them (e.g. rdfs:comment or dc:description)
9	Missing basic information	Part of the information needed for modelling the intended domain is not included in the ontology. This pitfall may be related to (a) the requirements included in the Ontology Requirement Specification Document (ORSD) that are not covered by the ontology, or (b) to the lack of knowledge that can be added to the ontology to make it more complete
10	Missing disjointness	The ontology lacks disjoint axioms between classes or between properties that should be defined as disjoint
11	Missing domain or range in properties	Object and/or datatype properties without domain or range (or none of them) are included in the ontology
12	Missing equivalent properties	The ontology lacks information about equivalent properties (owl:equivalentProperty) in the cases of duplicated relationships and/or attributes
13	Missing inverse relationships	This pitfall appears when any relationship (except for those that are defined as symmetric properties using owl:SymmetricProperty) does not have an inverse relationship (owl:inverseOf) defined within the ontology
14	Misusing "owl:allValuesFrom"	This pitfall consists in using the universal restriction (owl:allValuesFrom) as the default qualifier instead of the existential restriction (owl:someValuesFrom).

15	Misusing "not some" and "some not"	The pitfall consists in using a "some not" structure when a "not some" is required. This is due to the misplacement of the existential quantifier (owl:someValuesFrom) and the negative operator (owl:complementOf).
16	Misusing primitive and defined classes	This pitfall implies creating a primitive class rather than a defined one in case automatic classification of individuals is intended
17	Overspecializing a hierarchy	The hierarchy in the ontology is specialized in such a way that the final leaves are defined as classes and these classes will not have instances.
18	Overspecializing the domain or range	This pitfall consists in defining a domain or range not general enough for a property, i.e, no considering all the individuals or datatypes that might be involved in such a domain or range.
19	Defining multiple domains or ranges in properties Or Swapping intersection and union	The domain or range (or both) of a property (relationships and attributes) is defined by stating more than one rdfs:domain or rdfs:range statements. In OWL multiple rdfs:domain or rdfs:range axioms are allowed, but they are interpreted as conjunction, being, therefore, equivalent to the construct owl:intersectionOf.
20	Misusing ontology annotations	The contents of some annotation properties are swapped or misused. This pitfall might affect annotation properties related to natural language information (for example, annotations for naming such as rdfs:label or for providing descriptions such as rdfs:comment). Other types of annotation could also be affected as temporal, versioning information, among others.
21	Using a miscellaneous class	This pitfall refers to the creation of a class with the only goal of classifying the instances that do not belong to any of its sibling classes (classes with which the miscellaneous problematic class shares a common direct ancestor).
22	Using different naming criteria in the ontology	The ontology elements are not named following the same convention (for example CamelCase or use of delimiters as "-" or "_").
23	Duplicating a datatype already provided by the implementation language	A class and its corresponding individuals are created to represent existing datatypes in the implementation language.
24	Using recursive definition	An ontology element (a class, an object property or a datatype property) is used in its own definition.
25	Defining a relationship inverse to itself	A relationship is defined as inverse of itself. In this case, this relationship could have been defined as owl:SymmetricProperty instead.
26	Defining inverse relationships for a symmetric one	A symmetric object property (owl:SymmetricProperty) is defined as inverse of another object property (using owl:inverseOf).
27	Defining wrong equivalent relationships	Two object properties or two datatype properties are defined as equivalent, using owl:equivalentProperty, even though they do not have the same semantics.
28	Defining wrong symmetric relationships	A relationship is defined as symmetric, using owl:SymmetricProperty, when the relationship is not necessarily symmetric.
29	Defining wrong transitive relationships	A relationship is defined as transitive, using owl:TransitiveProperty, when the relationship is not necessarily transitive.
30	Missing equivalent classes	This pitfall consists in missing the definition of equivalent classes (owl:equivalentClass) in case of duplicated concepts. When an ontology reuses terms from other ontologies, classes

		that have the same meaning should be defined as equivalent in order to benefit the interoperability between both ontologies.
31	Defining wrong equivalent classes	Two classes are defined as equivalent, using owl:equivalentClass, when they are not necessarily equivalent.
32	Several classes with the same label	Two or more classes have the same content for natural language annotations for naming, for example, the rdfs:label annotation. This pitfall might involve lack of accuracy when defining terms.
33	Creating a property chain with just one property	This pitfall consists in creating a property chain (owl:propertyChainAxiom) that includes only one property in the antecedent part.
34	Untyped class	An ontology element is used as a class without having been explicitly declared as such using the primitives owl:Class or rdfs:Class.
35	Untyped property	An ontology element is used as a property without having been explicitly declared as such using the primitives rdf:Property, owl:ObjectProperty or owl:DatatypeProperty.
36	URI contains file extension	This pitfall occurs if file extensions such as ".owl", ".rdf", ".ttl", ".n3" and ".rdfxml" are included in an ontology URI.
37	Ontology not available	This pitfall occurs when the ontology code (OWL encoding) or its documentation (HTML document) is missing when looking up its URI.
38	No OWL ontology declaration	This pitfall consists in not declaring the owl:Ontology tag, which provides the ontology metadata. The owl:Ontology tag aims at gathering metadata about a given ontology such as version information, license, provenance, creation date, and so on. It is also used to declare the inclusion of other ontologies.
39	Ambiguous namespace	This pitfall consists in declaring neither the ontology URI nor the xml:base namespace. If this is the case, the ontology namespace is matched to the file location.
40	Namespace hijacking	It refers to reusing or referring to terms from another namespace that are not defined in such namespace.
41	No license declared	The ontology metadata omits information about the license that applies to the ontology.

Source: By thesis' researcher based on (Poveda-Villalón, 2016) and (Poveda-Villalón; Gómez-Pérez; Suárez-Figueroa, 2014)

APPENDIX 2: Main parts of systematic review protocol

Systematic review protocol: methods and techniques to build ontologies

1. Sources control group:

In a systematic review, it usually defined a control group of papers to later process with the search. These studies help the researcher to make decisions under each study should compose the base of study (Biolchini *et al.*, 2007, Mian *et al.*, 2005). The list bellow refers to title and respective reference to the control group of papers used in this SR. These studies were indicated by researcher's supervisor, found during the literature review main period, or found by an informal research on google scholar. It consists of previous studies on general concepts of ontologies and methodologies for developing ontologies.

- Methodologies, tools, and languages for building ontologies. Where is their meeting point? (Corcho; Fernández-López; Gómez-Pérez, 2003).
- Overview of methodologies for building ontologies. (Fernández-López, 1999).
- Overview and analysis of methodologies for building ontologies. (Fernández-López; Gómez-Pérez; Oacute, 2002).
- Methodologies and methods for building ontologies. (Gómez-Pérez; Fernández-López; Corcho, 2004).
- Understanding, building, and using ontologies. (Guarino, 1997b).
- The State of the Art in Ontology Design. (Hafner; Carole, 1997).
- Methodologies for ontology development. (Jones; Bench-Capon; Visser, 1998).
- Ontologies: Principles, methods and applications. (Uschold; Gruninger, 1996)
- Ontological realism: A methodology for coordinated evolution of scientific ontologies. (Smith; Ceusters, 2010).
- NeOn Methodology for building ontology networks: specification, scheduling and reuse. (Suárez-Figueroa, 2010b).
- Ontologies and controlled vocabulary: comparison of building methodologies. (Silva; Souza; Almeida, 2011).

2. Sources Selection Criteria Definition:

Search primary studies using electronic databases indexed that provide full text of studies. All databases selected must be open or provided for academic research by CAPES[94] or University at Buffalo Library. The sources cover the following areas of expertise: Information Science, computer science, philosophy, health sciences and medicine. The databases selected are IEEE Xplore, ACM Digital Library, INSPEC (Information Services for Physics, Electronics, and Computing), Web of Science, PubMed, EBSCOhost Research Databases, and CAPES Journals Portal

The preferably language of the sources were ***English***, by being the language most internationally accepted for writing scientific papers. ***Portuguese***, first due to be the native language of the researcher, and secondly because some previous studies that are known by the researcher were written in this language. English was the language used to compose the search strings.

This systematic review analyzed studies available in formats such as papers published in journals, conference proceedings, and books with a collection of papers. This review analyzed just one book due to researcher familiarity with it.

3. Sources Search Methods:

This SR performed two types of search in databases, an initial search identified a list of consolidated existing methodologies to building ontologies, and a second search recovered main studies describing each methodology from the previously list.

a. Search 1: Identifying methodologies to build ontologies

This was a preliminary search done to identify a set of methodologies to build domain ontologies. At this moment, the full-text language selection criteria were not relevant. The first search process included 1st) a search by studies on methodologies for ontology building; 2nd) analysis of results applying some filters to select the main methodologies found; and 3rd) documentation of selected methodologies.

[94] Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). In English: Coordination for the Improvement of Higher Education Personnel. CAPES is a Foundation within the Ministry of Education in Brazil whose central purpose is to coordinate efforts to improve the quality of Brazil's faculty and staff in higher education through grant programs. CAPES is particularly concerned with the training of Doctoral candidates, Pre-doctoral short-term researchers, and Post-doctoral Scholars.

First, it defined a string used to search. Remark that each data source has its own search mechanism, becoming necessary to adapt the search string to each source. The basic search string used in this preliminary search was:

```
(methodology OR methodologies OR method OR methods OR
techniques OR technique OR process OR processes OR
procedure OR procedures OR proceeding OR proceedings)
AND
(building OR build OR engineering OR engineer OR
developing OR develop OR designing OR design OR
compound OR conception OR make OR making OR fabricating
OR fabrication OR creation OR create)
AND
 (ontology OR ontologies OR ontological)
AND
 (knowledge
          AND (organization OR representation))
```

Then, it performed a filtering on the set of documents originated from searches just by reading the title, abstract and keywords looking for methodologies names. This filtering also helped to reject documents in disagreement with search goal (identify methodologies to build ontology). Sometimes, just this first filtering is not enough to identify the methodologies names, thus, it performed some reading of the both introduction and conclusion. All documents found were stored in a repository of research (EndNote X7, a software for managing references). Finally, it documented a list of methodologies for building ontologies.

b. Search 2: Analyzing methodologies to build ontologies

The goal of this second search was the identification of the main references of each methodology listed that are able to provide descriptions and explanations of the methodology. Typically, these studies are primary sources, however, some secondary and tertiary sources helps to reach the goal of this search. In fact, this is not a single search, but a search for each methodology identified.

4. Studies Selection Criteria

For the selection of the studies were defined types of studies, besides inclusion and exclusion criteria as described below. The scale used for each inclusion and exclusion criteria is nominal involving two options: *yes* or *no.*

Studies Types: Conference and journals papers, thesis, dissertation, books.

<u>Studies Priority Criteria:</u> Studies wrote by the own methodology authors.
<u>Studies Inclusion Criteria:</u> This SR selected studies that…

- must be published and available work fully in scientific databases or printed versions.
- must be written in English or Portuguese.
- must be recent work (published since 2005) or a consolidated oldest work.
- must already have approval by the scientific community.
- must present methods and techniques to build ontologies that has already been used at least once.
- present methodologies and methods most discussed in the literature. This will be defined by an occurrence frequency table that will be built.
- present a comparison between methods and techniques to build ontologies if properly present the researched references.
- must present methods and techniques to build a formal domain ontolog.

<u>Studies Exclusion Criteria:</u> This research did not select studies that …

- describing methodologies to build top level ontologies.
- published as short papers or posters.
- present ontology construction methodology without showing their practical application.
- not present domain ontology construction.
- present methodologies and methods weakly discussed in the literature. This will be defined by an occurrence frequency table that will be built.
- their titles, keyword and abstract are not available in English.
- did not fit to Inclusion Criteria

5. Quality criteria of studies:

- In the case of papers or articles, the study must have been published in a journal or conference proceedings with peer review.

- In case of dissertation or thesis, they must have been approved by the examining board.
- Prioritize articles following criteria below: population considered in the evaluation and statistical methods.
    - number of citations received.
    - number of citations received by the authors.
    - ranking of source or journal that the study was found according SJR[95].

SR results consisted of a qualitative analysis of each methodology draw up in order to determine the suitability of the methodology to ontologies building, identifying advantages and disadvantages of each methodology. Next, it presented some features evaluated:

- Activities and activities workflow.
- Inputs and outputs activities.Development life cycle.
- Ontologies built and its area or field of application.
- Principles and patterns defined.
- Tools used and recommended.
- Adherence to OBO foundry principals.

6. Defining the Ontology Scope

The existing literature on ontology development emphasizes that the development of an ontology starts by defining its coverage domain and scope (Arp; Smith; Spear, 2015, Suárez-Figueroa, 2010b). In addition to determining the domain of knowledge that will be covered by the ontology, Noy and Mcguinness (2001) states that this step should identify who are its users, what are its applications (utility), and what questions intend to answers.

Given the research problem of this study and the case study design described, in general, the target ontology aims to cover the lack of formal representation of obstetric and neonatal knowledge. This ontology intends to provide the exchange of patient data stored in

[95] SJR-Scientific Journal Rankings (URL: http://www.scimagojr.com/)

multiples EHR systems, in which patients are undergoing on prenatal, childbirth and postnatal care.

Notwithstanding the foregoing, ontology scope definition is one of the initial activities of ontology development. Ontology scope definition usually carried out together with identification of ontology requirements. Considering that the ontology requirements identification activity is part of the ontology development methodology, this activity description is part the chapter 4.3.1.

APPENDIX 3: Ontology Requirements Specification Document

Obstetric and Neonatal Ontology Requirements Specification Document

<table>
<tr><th>1</th><th>Purpose</th></tr>
<tr><td colspan="2">The purpose of building the Obstetric and Neonatal Ontology is to provide a formal representation of the data from health records involved in the care of the pregnant woman, and of her baby. This ontology will cover data from different medical specialties involved in care both of the woman and her child.</td></tr>
<tr><th>2</th><th>Scope</th></tr>
<tr><td colspan="2">The ontology should cover the data found on medical records involved in clinical care activities given to the mother, the developing fetus, and the newborn child during in the prenatal, intrapartum and postnatal periods of care. The ontology also encompasses the knowledge of the obstetric and neonatal domain necessary to understand the medical records as well as the clinical care of the care periods.</td></tr>
<tr><th>3</th><th>Implementation Language</th></tr>
<tr><td colspan="2">OWL 2 Web Ontology Language is the implementation language chosen.</td></tr>
<tr><th>4</th><th>Intended End-Users</th></tr>
<tr><td colspan="2">User 1: User of the electronic health record (EHR) system who wants to model a new EHR.
User 2: User who administrates the EHR system.
User 3: Ontologists that wants to extend or reuse the terms of this ontology.
User 4: Software developments that want to build interoperable EHR systems.
User 5: Government analysts that study the situation of the obstetric and neonatal care and prepare reports to support the development of public health policies, improvements in the application of health resources, and provide better quality of maternal and newborn care.
User 6: Organizations analysts that want to analyze the statistics of the obstetric and neonatal care and improve their services.
User 7: Healthcare institutions, public (federal, state, district or municipal) and private that want to provide data and information on obstetric and neonatal care, open and structured format.
User 8: Organizations and entities of society that relate in some way with health institutions providing care to pregnant women and newborn.
User 9: Researchers who want to use the ontology to improve their research by the common representation and understanding of the obstetrical and neonatal EHR data.
User 10: Institutions in general that want to publish their data by open data linked solutions.</td></tr>
<tr><th>5</th><th>Intended Uses</th></tr>
<tr><td colspan="2">Use 1: To provide a consensus representation of EHR data for both mother and child understandable by humans as well as by computers.
Use 2: To serve as semantic interoperability solution among information systems.
Use 3: To connect heterogeneous clinical information models.
Use 4: To manage data from different EHRs by defining a common representation of data.
Use 5: To improve the clinical translational research by providing enhanced access to EHRs data.
Use 6: To increase the reasoning with EHRs data by providing better inference among the data.
Use 7: To serve as a conceptual model of obstetric and neonatal domain data.
Use 8: To provide data on obstetric and neonatal in open and semanticized format.</td></tr>
<tr><th>6</th><th>Ontology Requirements</th></tr>
<tr><th colspan="2">a. Non-Functional Requirements (NFR)</th></tr>
<tr><td colspan="2">NFR1. The ontology must support a multilingual scenario at least the following languages: English and Portuguese.
NFR2. The ontology should follow the principles of the OBO Foundry[96].
NFR 3. The implementation language must be OWL 2 Web Ontology Language.</td></tr>
</table>

[96] Available at: <http://www.obofoundry.org/principles/fp-000-summary.html>

NFR4. The ontology should use as top-level ontology Basic Formal Ontology version 2.0. NFR5. The ontology should reuse other ontologies already accepted by the OBO Foundry whenever possible. NFR6. The ontology should prioritize the use open source tools. NFR7. The ontology should use the modularization strategy in order to promote reuse flexibility of thematic modules.
b. Functional Requirements: Groups of Competency Questions
Group 1 of Competency Questions (G1CQ): Basic Demographic information G1CQ-1: What is the educational level, occupation, age group, ethnicity of the pregnant? G1CQ-2: What is the educational level, occupation, age group, ethnicity of the baby's father? Group 2 of Competency Questions(G2CQ): Basic health information and vitals G2CQ-1: What is the height, the blood group, the Rh factor of the pregnant? G2CQ-2: What is the blood group, the Rh factor of the fetus? G2CQ-3: What is the blood group, the Rh factor of the baby's father? Group 3 of Competency Questions(G3CQ): Evolution of the basic health information and vital signs previously and during pregnancy G3CQ-1: How is the evolution of the pregnant weight during the pregnancy? G3CQ-2: How is the evolution of the body temperature, blood pressure, pulse (heart rate), and breathing rate (respiratory rate) of the pregnant during the pregnancy? G3CQ-3: How is the evolution of the abdominal circumference and the fundus of uterus (uterine height) measurement of the pregnant during the pregnancy? G3CQ-4: How is the evolution of the fetal heart beats auscultation during the pregnancy? G3CQ-5: When the first period (menstrual period) of the patient X happens? (age of patient) G3CQ-6: How long the menstrual period of the patient X usually last? G3CQ-7: What is the frequency of the menstrual period of the patient X? (Number of days between periods). G3CQ-8: Does the patient X have heavy bleeding? G3CQ-9: Does the patient X have irregular periods? G3CQ-10: How long the menstrual cycle of the patient X last? Group 4 of Competency Questions(G4CQ): Evolution of the basic health information and vital signs during childhood G4CQ-1: How is the evolution of the newborn weight and height during the first year of life? G4CQ-2: How is the evolution of the baby weight and height during the second year of life? G4CQ-3: Does the newborn physical examination detect any anomaly? Group 5 of Competency Questions(G5CQ): Information about pregnant and prenatal conducted G5CQ-1: What are the results of prenatal tests that the pregnant performed? G5CQ-2: When the pregnant performed the prenatal exams? G5CQ-3: Does the patient have any report of placenta previa? G5CQ-4: Does the patient have any report of gestational diabetes? G5CQ-5: Does the patient have any report of gestational hypertension? G5CQ-6: Does the patient have any relevant report of family history such as sown syndrome or cystic fibrosis? Group 6 of Competency Questions(G6CQ): Prenatal monitoring information G6CQ-1: What are the congenital diseases that the fetus has provision to develop? G6CQ-2: What are the occupational hazards that pregnant this subject? G6CQ-3: What are the pregnant's symptoms and signs during the pregnancy? Group 7 of Competency Questions(G7CQ): Information of intrapartum G7CQ-1: How many previous pregnancies the patient X had? G7CQ-2: How many previous C-section the patient X had? G7CQ-3: How many previous vaginal births the patient X had? G7CQ-4: How many previous spontaneous abortion the patient X had? G7CQ-5: How many previous inducted abortion the patient X had?

Group 8 of Competency Questions(G8CQ): Profile of prenatal, intrapartum and postpartum
G8CQ-1: How many childbirth happened at institution X?
G8CQ-2: How many C-section happened at institution X?
G8CQ-3: How many vaginal births happened at institution X?
G8CQ-4: How many hospitalizations happened as consequence of inducted abortion?
G8CQ-5: How many (abortion or C-section) by education level happened?
G8CQ-6: How many (abortion or C-section) by age group happened
G8CQ-7: How many (abortion or C-section) by marital status happened

7	**Pre-Glossary of Terms**	
	a. Terms from Competency Questions and possible answers	
	Vital sign, body temperature, blood pressure, pulse (heart rate), breathing rate (respiratory rate). Abdominal circumference measurement. Fundus of uterus (uterine height) measurement. Fetal heart beats auscultation. Weight (Pregnant and Newborn). Height (Pregnant and Newborn). Fertilization, Ovulation, Implantation, Menstrual cycle, menstrual period. Cramps, symptoms, signs. Occupational hazards.	Pregnancy, prenatal, intrapartum, postnatal, puerperium. Pregnant, fetus, embryo, child, newborn. Delivery, childbirth, labor, birth. Vaginal birth, Cesarean (C-section). Disease, congenital disease, diagnosis. Educational level, occupation, age group, ethnicity, gender, name, phone number. Patient, husband, emergence contact, responsible. Family history, immunizations. Gestacional diabetes, hypertension, placenta previa.
	b. Objects or Instances Examples	

APPENDIX 4: Iteration 1 Requirements Specification

Iteration 1: Woman anatomy and physical examination	
1	**Purpose of iteration**

The purpose of iteration 1 is to provide a formal representation of the knowledge of the anatomy of woman body and the physical examination issues.

2	**Scope of iteration**

This piece of ontology should cover knowledge of the anatomy of the female reproductive system. In this iteration, the ontology also should represent information related to the physical examination and general systems examination.

3	**Ontology Iteration Requirements**

a. Non-Functional Requirements (NFR)

NFR1. The ontology must support a multilingual scenario at least the following languages: English and Portuguese.
NFR2. The ontology should follow the principles of the OBO Foundry[97].
NFR3. The ontology should use as top-level ontology Basic Formal Ontology version 2.0.
NFR4. OntONeo must prioritize the reuse of entities from OBO Foundry ontologies and others instead to create new ones.

b. Functional Requirements

The ontology should represent knowledge of …

FR1. female breast anatomy including and all breast parts.
FR2. female breast examination.
FR3. female reproductive system anatomy including all parts.
FR4. female abdomen anatomy and its parts.
FR5. woman physical examination.
FR6. female pelvic examination.
FR6. female pelvic examination.
FR7. pap smear examination.

The ontology should covers data of …

FR8. physical examination section of medical record.
FR9. examination of all body systems of medical record.

c. Competency Questions

CQ1: What are the clinical findings of the patient physical examination?
CQ2: What is the weight of patient measured at the last medical encounter?
CQ3: Had the patient presented weight gain or loss since the previous medical encounter?
CQ4: What is the height of patient measured at the last medical encounter?
CQ5: What are the organs of the female reproductive system?
CQ6: What are the parts of the female breast?
CQ7: What are the symptoms identified during the physical examination?

---

[97] Available at: <http://www.obofoundry.org/principles/fp-000-summary.html>

4	Ontology Iteration Requirements Elicitation

FR1: The ontology should represent knowledge of female breast anatomy including all breast parts.
FR5. The ontology should represent knowledge of woman physical examination.

Figure 1 showed the female breast anatomy

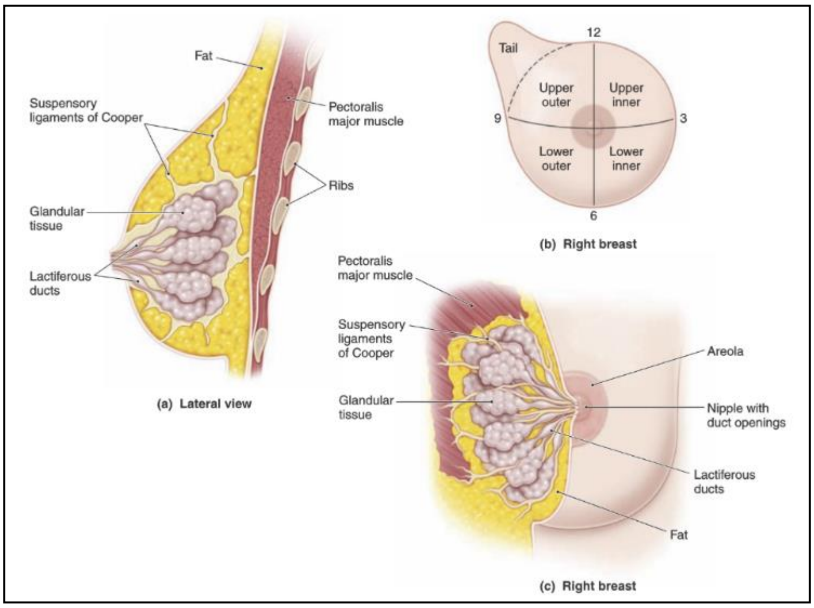


Figure 1: Clinical anatomy and associated examination schema of the breast (Beckmann et al., 2010b, p. 7).

FR2. The ontology should represent knowledge of female breast examination.

Figure 2 showed some knowledge on female breast examination.

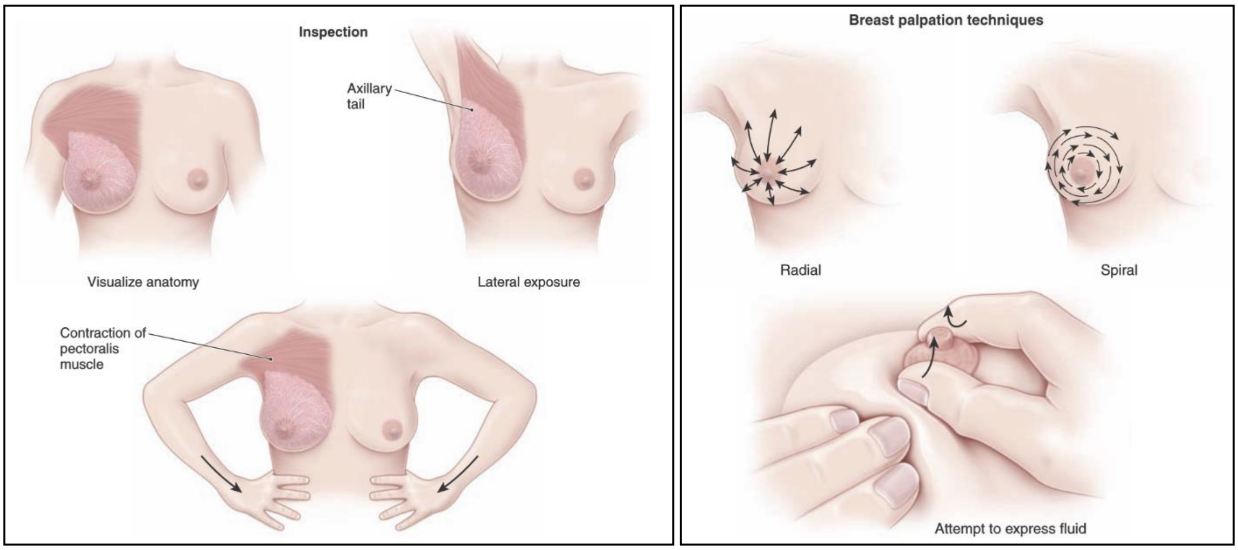


Figure 2: Clinical anatomy and associated examination schema of the breast (Beckmann et al., 2010b, p. 8).

FR3. The ontology should represent knowledge of female reproductive system anatomy including all parts.

Figure 3 showed some knowledge on female external genitalia.

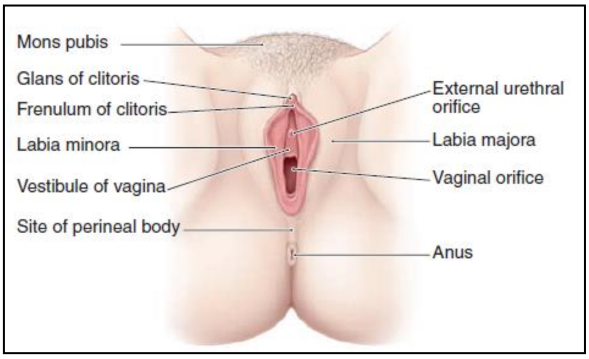


Figure 3: Female external genitalia parts (Beckmann et al., 2010a, p. 36).

Next, Figure 4 showed some knowledge on female internal genitalia.

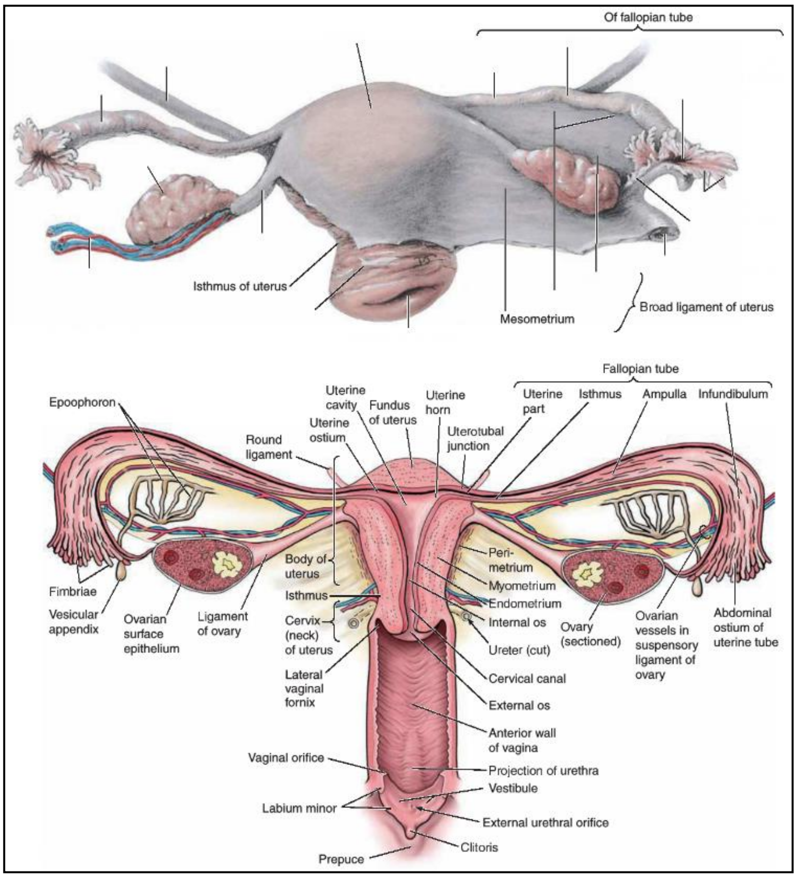
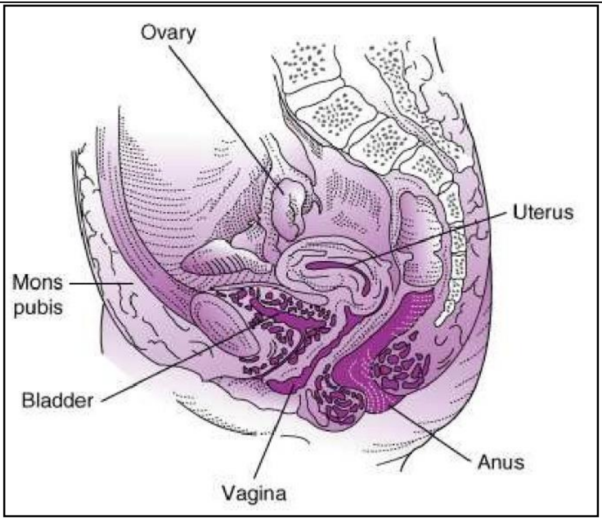


Figure 4: Female internal genitalia parts (Beckmann et al., 2010a, p. 38).

Figure 5 showed an overview on female reproductive system.



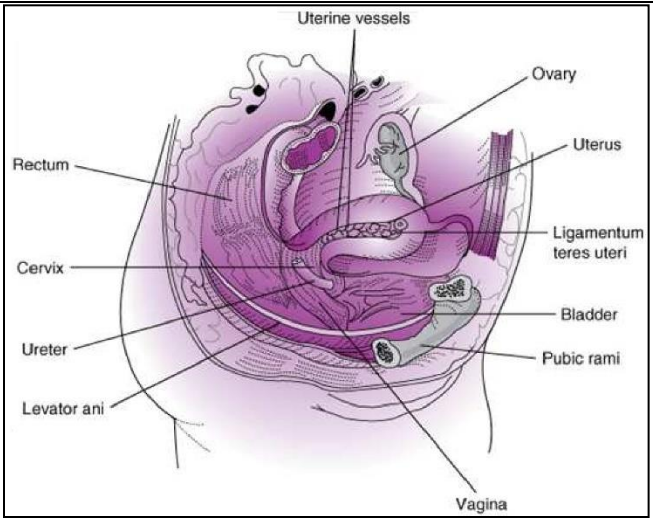


Figure 5: Female reproductive system overview (Heitmann, 2013, Fig. 20 and 21 of Chapter 1).

FR4. The ontology should represent knowledge of female abdomen anatomy and its parts.

Figure 6 showed an overview on female abdomen and its parts.

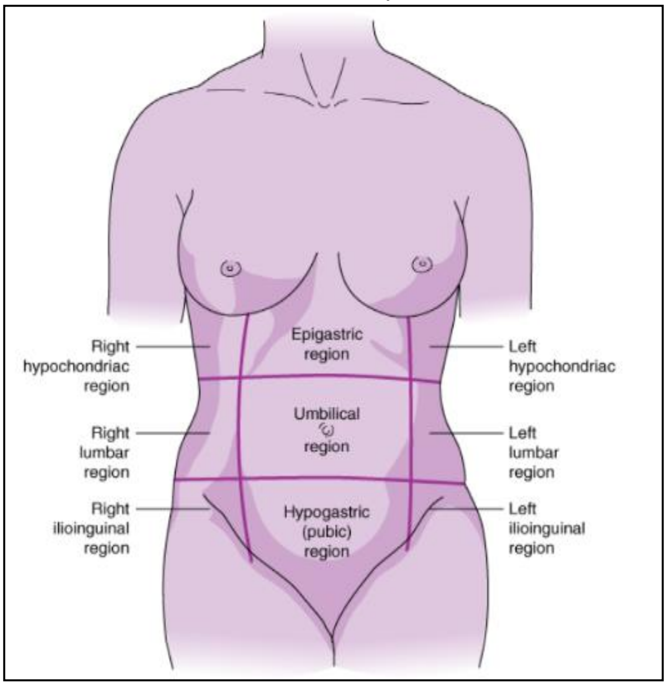


Figure 6: Female abdomen parts (Heitmann, 2013).

FR6. The ontology should represent knowledge of female pelvic examination.
FR5. The ontology should represent knowledge of woman physical examination.

Figures 7 and 8 presents some knowledge of female pelvic examination.

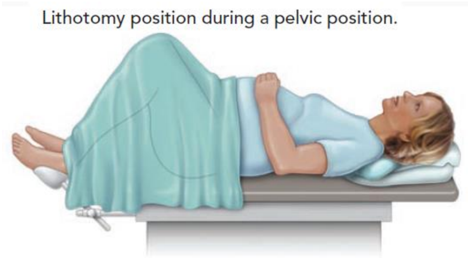


Figure 7: Position during Female pelvic examination *(Beckmann et al., 2010a, p. 9).*

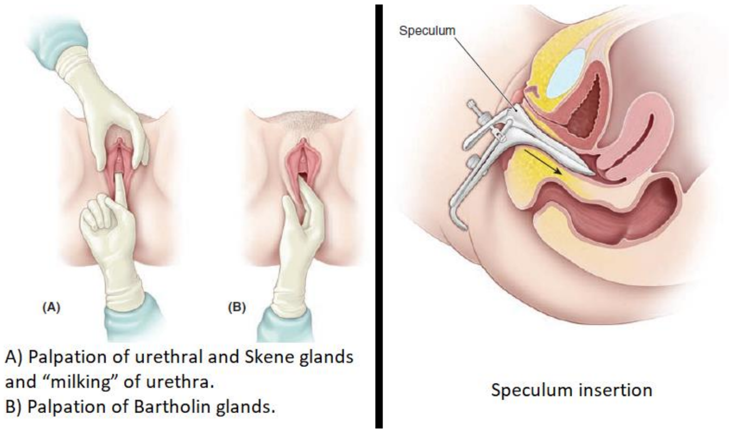


Figure 8: Details of Female pelvic examination *(Beckmann et al., 2010a, pp. 10-11).*

FR7. The ontology should represent knowledge of pap smear examination.
Figure 9 showed some details of pap smear examination.

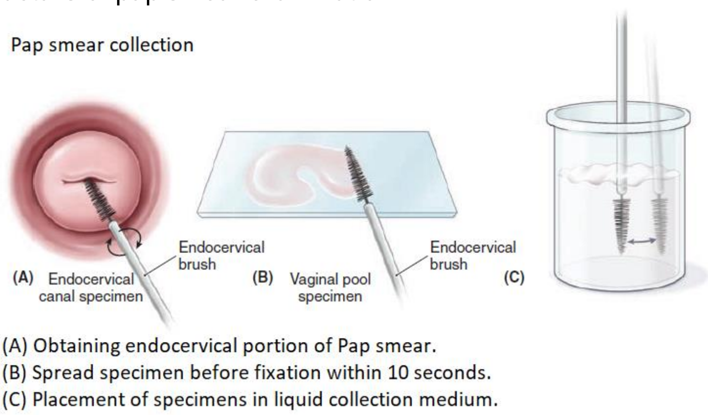


Figure 9: Pap smear examination (Beckmann et al., 2010a, p. 11)

FR9. The ontology should covers data of examination of all body systems of medical record.

Figure 10 presents the review of systems section of the ACOG Woman's Health Record.

REVIEW OF SYSTEMS (ROS)

1. CONSTITUTIONAL	☐ NEGATIVE ☐ WEIGHT LOSS ☐ WEIGHT GAIN ☐ FEVER ☐ FATIGUE ☐ OTHER TALLEST HEIGHT__________
2. EYES	☐ NEGATIVE ☐ VISION CHANGE ☐ GLASSES/CONTACTS ☐ OTHER
3. EAR, NOSE, AND THROAT	☐ NEGATIVE ☐ ULCERS ☐ SINUSITIS ☐ HEADACHE ☐ HEARING LOSS ☐ OTHER
4. CARDIOVASCULAR	☐ NEGATIVE ☐ ORTHOPNEA ☐ CHEST PAIN ☐ DIFFICULTY BREATHING ON EXERTION ☐ EDEMA ☐ PALPITATION ☐ OTHER
5. RESPIRATORY	☐ NEGATIVE ☐ WHEEZING ☐ HEMOPTYSIS ☐ SHORTNESS OF BREATH ☐ COUGH ☐ OTHER
6. GASTROINTESTINAL	☐ NEGATIVE ☐ DIARRHEA ☐ BLOODY STOOL ☐ NAUSEA/VOMITING/INDIGESTION ☐ CONSTIPATION ☐ FLATULENCE ☐ PAIN ☐ FECAL INCONTINENCE ☐ OTHER
7. GENITOURINARY	☐ NEGATIVE ☐ HEMATURIA ☐ DYSURIA ☐ URGENCY ☐ FREQUENCY ☐ INCOMPLETE EMPTYING ☐ INCONTINENCE ☐ DYSPAREUNIA ☐ ABNORMAL OR PAINFUL PERIODS ☐ PMS ☐ ABNORMAL VAGINAL BLEEDING ☐ ABNORMALVAGINAL DISCHARGE ☐ OTHER
8. MUSCULOSKELETAL	☐ NEGATIVE ☐ MUSCLE WEAKNESS ☐ MUSCLE OR JOINT PAIN ☐ OTHER
9a. SKIN	☐ NEGATIVE ☐ RASH ☐ ULCERS ☐ DRY SKIN ☐ PIGMENTED LESIONS ☐ OTHER
9b. BREAST	☐ NEGATIVE ☐ MASTALGIA ☐ DISCHARGE ☐ MASSES ☐ OTHER
10. NEUROLOGIC	☐ NEGATIVE ☐ SYNCOPE ☐ SEIZURES ☐ NUMBNESS ☐ TROUBLE WALKING ☐ SEVERE MEMORY PROBLEMS ☐ OTHER
11. PSYCHIATRIC	☐ NEGATIVE ☐ DEPRESSION ☐ CRYING ☐ SEVERE ANXIETY ☐ OTHER
12. ENDOCRINE	☐ NEGATIVE ☐ DIABETES ☐ HYPOTHYROID ☐ HYPERTHYROID ☐ HOT FLASHES ☐ HAIR LOSS ☐ HEAT/COLD INTOLERANCE ☐ OTHER
13. HEMATOLOGIC/LYMPHATIC	☐ NEGATIVE ☐ BRUISES ☐ BLEEDING ☐ ADENOPATHY ☐ OTHER
14. ALLERGIC/IMMUNOLOGIC	(SEEFIRST PAGE)

Figure 10: Review of systems section of ACOG Woman's Health Record (ACOG, 2010)

FR8. The ontology should covers data of physical examination section of medical record.

Figure 11 presents the review of systems section of the ACOG Woman's Health Record.

PHYSICAL EXAMINATION

PATIENT NAME: | BIRTH DATE: / / | ID NO.: | DATE: / /

CONSTITUTIONAL

• VITAL SIGNS (RECORD ≥ 3 VITAL SIGNS):

HEIGHT: _______ WEIGHT: _______ BMI: _______ BLOOD PRESSURE (SITTING): _______ TEMPERATURE: _______ PULSE: _______ RESPIRATION: _______

• GENERAL APPEARANCE (NOTE ALL THAT APPLY):

☐ WELL-DEVELOPED ☐ OTHER ☐ NO DEFORMITIES ☐ OTHER

☐ WELL-NOURISHED ☐ OTHER ☐ WELL-GROOMED ☐ OTHER

☐ NORMAL HABITUS ☐ OBESE ☐ OTHER

NECK

• NECK ☐ NORMAL ☐ ABNORMAL

• THYROID ☐ NORMAL ☐ ABNORMAL

RESPIRATORY

• RESPIRATORY EFFORT ☐ NORMAL ☐ ABNORMAL

• AUSCULTATED LUNGS ☐ NORMAL ☐ ABNORMAL

CARDIOVASCULAR

• AUSCULTATED HEART

SOUNDS ☐ NORMAL ☐ ABNORMAL

MURMURS ☐ NORMAL ☐ ABNORMAL

• PERIPHERAL VASCULAR ☐ NORMAL ☐ ABNORMAL

GASTROINTESTINAL

• ABDOMEN ☐ NORMAL ☐ ABNORMAL

• HERNIA ☐ NONE ☐ PRESENT

• LIVER/SPLEEN

LIVER ☐ NORMAL ☐ ABNORMAL

SPLEEN ☐ NORMAL ☐ ABNORMAL

• STOOL GUAIAC, IF INDICATED ☐ POSITIVE ☐ NEGATIVE

LYMPHATIC

• PALPATION OF NODES (CHOOSE ALL THAT ARE APPLICABLE)

NECK ☐ NORMAL ☐ ABNORMAL

AXILLA ☐ NORMAL ☐ ABNORMAL

GROIN ☐ NORMAL ☐ ABNORMAL

OTHER SITE ☐ NORMAL ☐ ABNORMAL

SKIN

• INSPECTED/PALPATED ☐ NORMAL ☐ ABNORMAL

NEUROLOGIC/PSYCHIATRIC

• ORIENTATION ☐ TIME ☐ PLACE ☐ PERSON ☐ COMMENTS

• MOOD AND AFFECT ☐ NORMAL ☐ DEPRESSED ☐ ANXIOUS ☐ AGITATED ☐ OTHER

GYNECOLOGIC (AT LEAST 7)

• BREASTS ☐ NORMAL ☐ ABNORMAL

• EXTERNAL GENITALIA ☐ NORMAL ☐ ABNORMAL

• URETHRAL MEATUS ☐ NORMAL ☐ ABNORMAL

• URETHRA ☐ NORMAL ☐ ABNORMAL

• BLADDER ☐ NORMAL ☐ ABNORMAL

• VAGINA/PELVIC SUPPORT ☐ NORMAL ☐ ABNORMAL

• CERVIX ☐ NORMAL ☐ ABNORMAL

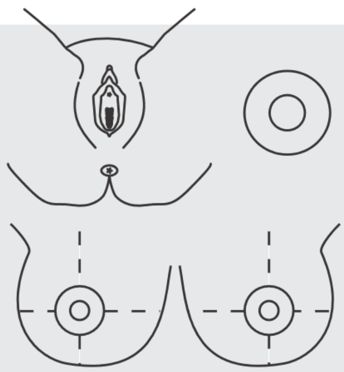
• UTERUS ☐ NORMAL ☐ ABNORMAL

• ADNEXA/PARAMETRIA ☐ NORMAL ☐ ABNORMAL

• ANUS/PERINEUM ☐ NORMAL ☐ ABNORMAL

• RECTAL ☐ NORMAL ☐ ABNORMAL

(SEE ALSO "STOOL GUAIAC" ABOVE)



• TOTAL NUMBER OF BULLETED (•) ELEMENTS EXAMINED:

Figure 11: Physical examination section of ACOG Woman's Health Record (ACOG, 2010)

3	Intended Uses

Use 1: To provide a consensus representation of knowledge of woman anatomy and physiology.

Use 2: To serve as semantic interoperability solution among information systems that deal with woman anatomy domain.
Use 3: To connect heterogeneous clinical information models.
Use 4: To manage data from different EHRs by defining a common representation of data.
Use 5: To improve the clinical translational research by providing enhanced access to EHRs data.
Use 6: To increase the reasoning with EHRs data by providing better inference among the data.
Use 7: To serve as a conceptual model of obstetric and neonatal domain data.
Use 8: To provide data on obstetric and neonatal in open and semanticized format.

5 Pre-Glossary of Terms

a. List of general terms

- Body, female body, body weight, body height.
- medical encounter
- patient, woman, physician,
- weight gain or weight loss
- Female breast, breast, etc.
- clinical findings, physical examination, pelvic examination, breast examination, pap smear examination
- last medical encounter (date)
- organ, organ system, body system, female reproductive system, reproductive system
- symptom

b. Objects or Instances Examples

- Woman 1
- Woman 2
- Woman 3

6 Ontologies and terminologies candidates to reuse

- Ontology for General Medical Science (OGMS)
- Ontology for Biomedical Investigations (OBI)
- Ontology of Medically Related Social Entities (OMRSE)
- Information Artifact Ontology (IAO)
- Ontology of Document Acts (d-acts)
- Informed Consent Ontology (ICO)
- Common Anatomy Reference Ontology (CARO)
- Foundational Model of Anatomy (FMA)
- Uberon multi-species anatomy ontology (UBERON)
- Cell Ontology (CL)
- Human phenotype ontology (HP)

- Phenotypic quality (PaTO)
- Genotype Ontology (GENO)
- Gene Ontology (GO) - biological process
- Symptom Ontology (SYMP)
- Clinical measurement ontology (CMO)
- Human Disease Ontology (DOID)
- Infectious Disease Ontology (IDO)
- Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT)
- National Cancer Institute Thesaurus (NCIT)
- Health Level Seven (HL7)
- Medical Dictionary for Regulatory Activities (MEDRA)
- Medical Subject Headings (MESH)
- International Classification of Diseases (ICD)
- Logical Observation Identifier Names and Codes (LOINC)

7 Ontology Design Patterns candidates to reuse

- Presentation Ontology Design Patterns (Presentation OPs)
    - Annotation Ontology Design Patterns (Annotation OPs)
    - Naming Ontology Design Patterns (Naming OPs)
- Content Ontology Design Patterns (CPs)
    - All subcategory of patterns
- Lexico-syntactic Ontology Design Patterns (Lexico-syntactic OPs)
    - All subcategory of patterns
- Correspondence Ontology Design Patterns (Correspondence OPs)
    - Alignment Ontology Design Patterns (Alignment OPs)
    - Reengineering Ontology Design Patterns (Reengineering OPs)
- Structural Ontology Design Patterns (Structural OPs)
    - Logical Ontology Design Patterns (Logical OPs)

8 Literature review

ACOG, American College of Obstetricians and Gynecologists. (2010). ACOG Woman's Health Record. In Charles R. B. Beckmann, Frank W. Ling, Roger P. Smith, Barbara M. Barzansky, William N.P. Herbert, & Douglas W. Laube (Eds.), Obstetrics and Gynecology (6th ed., pp. 447-464). Philadelphia: Lippincott Williams & Wilkins.

Beckmann, Charles R. B., Ling, Frank W., Herbert, William N. P. , Laube, Douglas W. , & Smith, Roger P. . (2010a). Embryology and anatomy. In Charles R. B. Beckmann, Frank W. Ling, Roger P. Smith, Barbara M. Barzansky, William N.P. Herbert, & Douglas W. Laube (Eds.), Obstetrics and Gynecology (6th ed., pp. 29-41). Philadelphia: Lippincott Williams & Wilkins.

Beckmann, Charles R. B., Ling, Frank W., Herbert, William N. P. , Laube, Douglas W. , & Smith, Roger P. . (2010b). The woman's health examination. In Charles R. B. Beckmann, Frank W. Ling, Roger P. Smith, Barbara M. Barzansky, William N.P. Herbert, & Douglas W. Laube (Eds.), Obstetrics and Gynecology (6th ed., pp. 1-13). Philadelphia: Lippincott Williams & Wilkins.

Heitmann, Ryan J. (2013). Anatomy of the Female Reproductive System. In Alan H. DeCherney, Lauren Nathan, Neri Laufer, & Ashley S. Roman (Eds.), CURRENT Diagnosis & Treatment: Obstetrics & Gynecology. New York, NY: The McGraw-Hill Companies.

Hoffman, Barbara L., Schorge, John O., Schaffer, Joseph I., Halvorson, Lisa M., Bradshaw, Karen D., Cunningham, F. Gary, & Calver, Lewis E. (2012). Anatomy Williams Gynecology, 2e (pp. 918-947). New York, NY: The McGraw-Hill Companies.

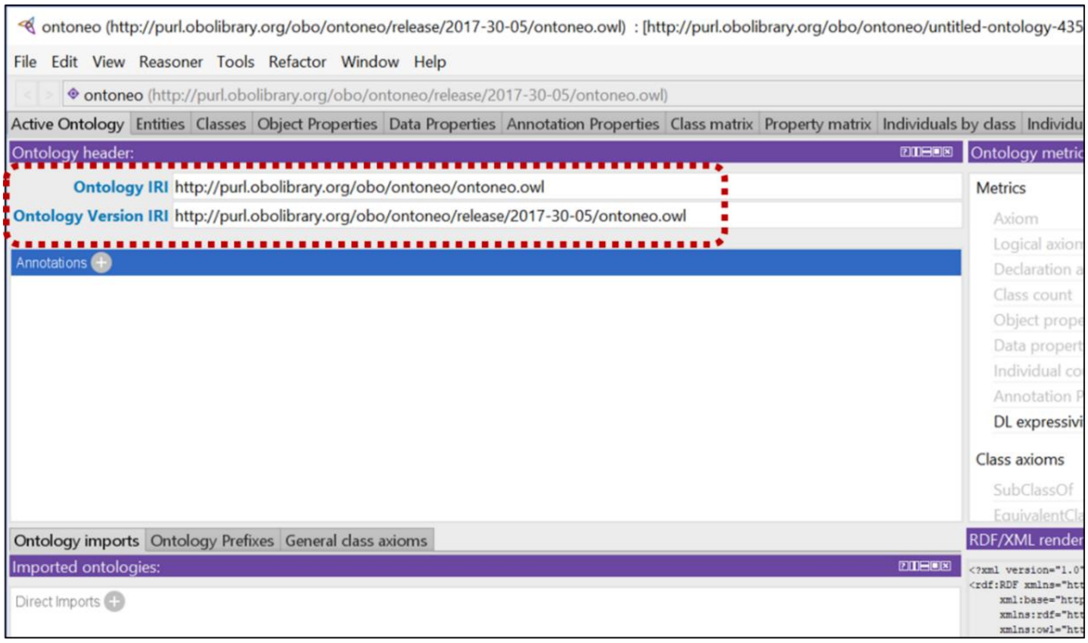
APPENDIX 5: Architecture Specification Document of OntONeo

Obstetric and Neonatal Ontology Architecture Specification Document		
1	**Top-level ontology to extend**	Basic Formal Ontology version 2.0
2	**Ontology namespace**	http://purl.obolibrary.org/obo/ontoneo/
	PREFIX or IDSPACE	Ontoneo
	Ontology IRI	http://purl.obolibrary.org/obo/ontoneo/ontoneo.owl
	Default purl	http://purl.obolibrary.org/obo/ontoneo/
	Local identifier	http://purl.obolibrary.org/obo/ontoneo/<PREFIX>_<99999999> Ex: http://purl.obolibrary.org/obo/ontoneo/ONTONEO_00000001
3	**Ontology project website**	https://ontoneo.com/
4	**Development repository**	https://github.com/ontoneo-project/Ontoneo
5	**Development tools**	Protégé version 5.2 Ontobee ontology search engine Ontofox system to get pieces of ontologies
6	**Coding language**	OWL 2 Web Ontology Language
7	**Default language to ontology elements**	English
8	**Alternative language to ontology elements**	Portuguese
9	**License**	Creative Commons CC-BY license version 4.0
10	**Minimal ontology documentation**	
Each new element must be documented with: term creator date of term creation label both in english and portuguese definition both in english and portuguese Each reuse element must be documented with: source ontology		
11	Additional information	
The ontology should follow the principles of the OBO Foundry. The ontology should reuse other ontologies already accepted by the OBO Foundry whenever possible. The ontology should prioritize the use open source tools. The ontology should use the modularization strategy in order to promote reuse flexibility of thematic modules.		

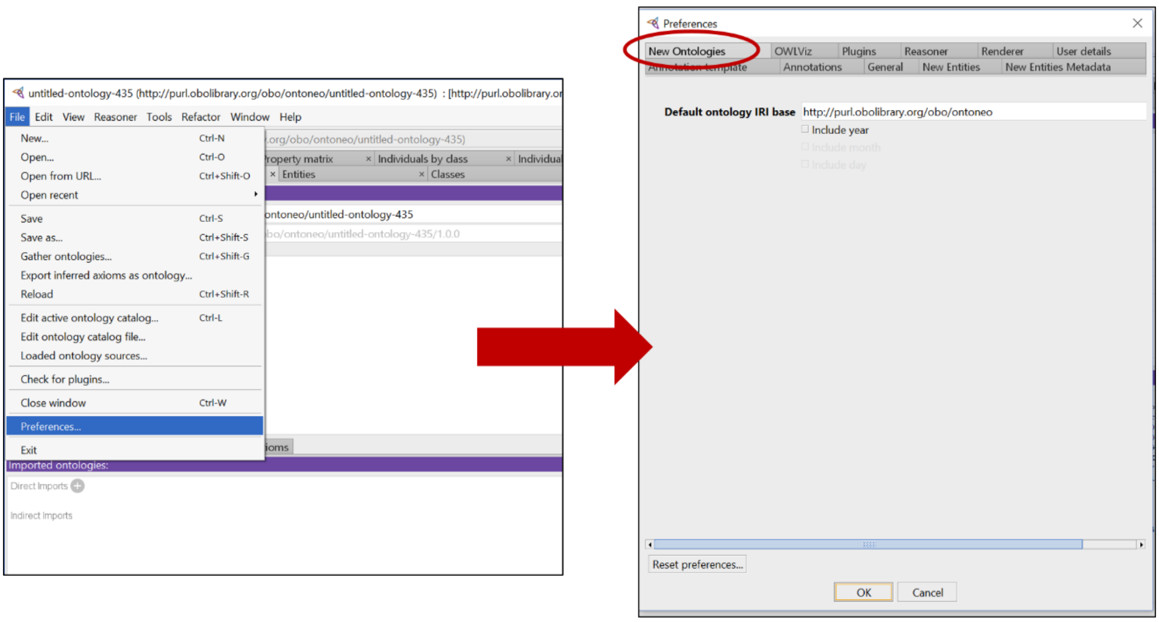
APPENDIX 6: Configuration of Protégé desktop

<u>**Step-by-step to configure the Protégé desktop for the ontology development**</u>

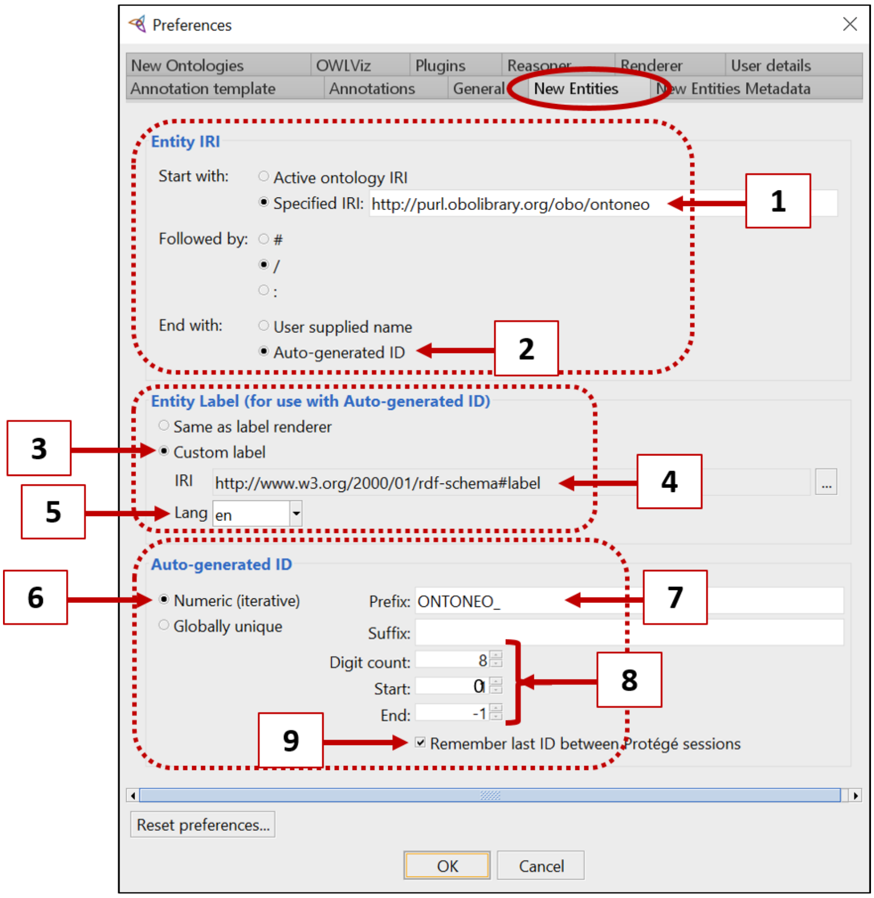
1) Configure the URI / IRI of both the ontology and the versioning of the ontology.



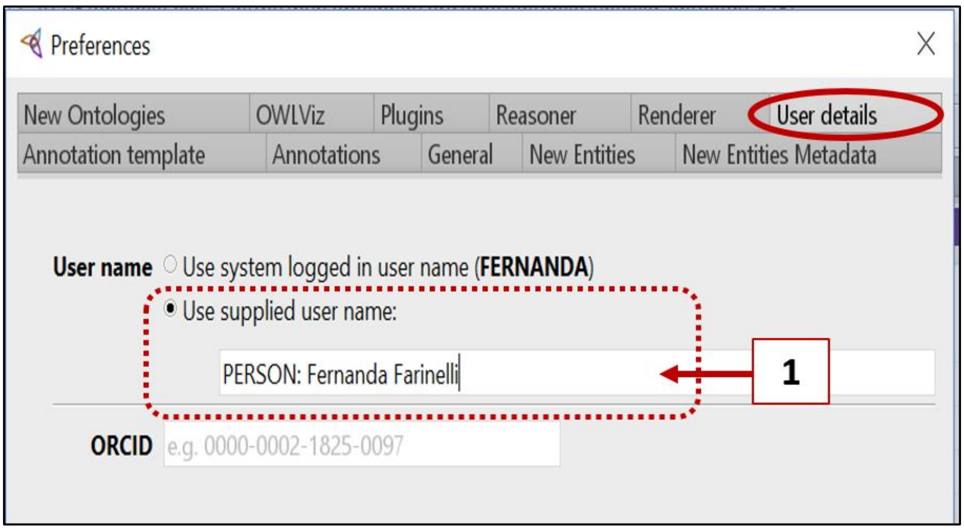
2) Configure base URI / IRI for ontology elements.



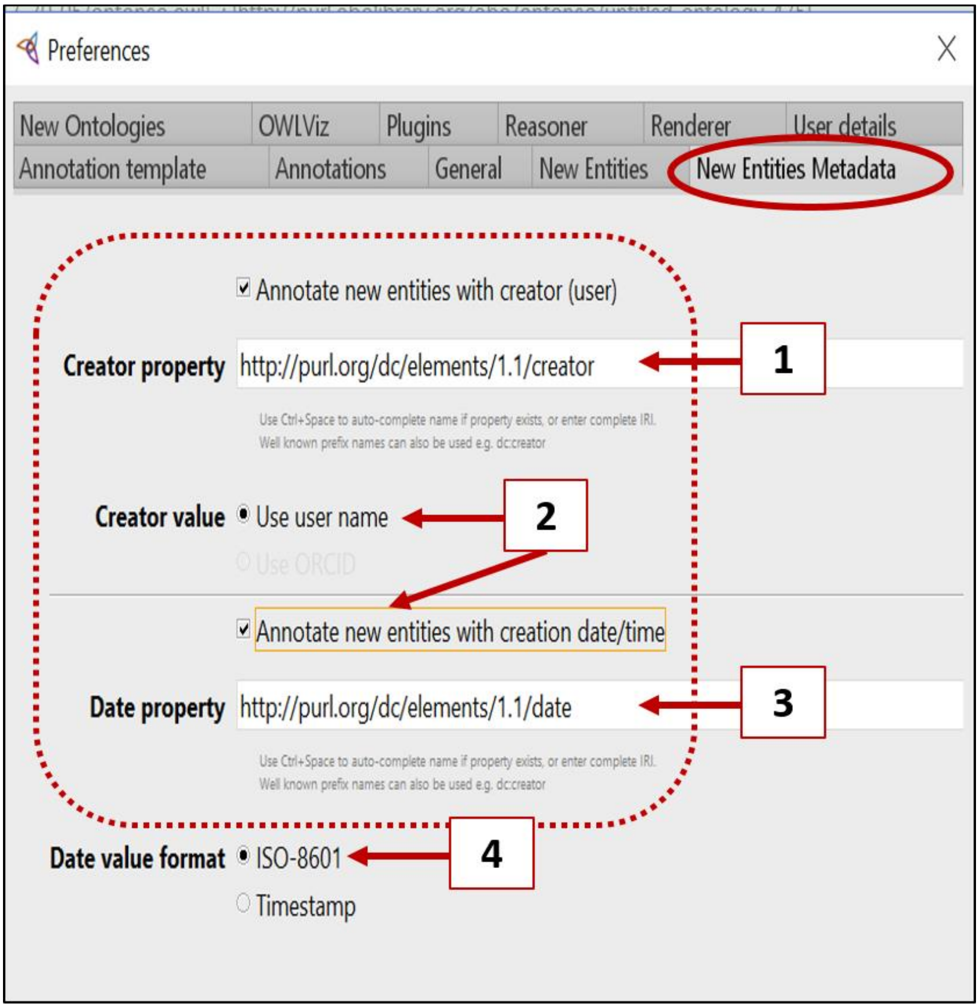
3) Configure URI / IRI for the entities of the ontology and the form of generation of the identifier of the elements of the ontology.



4) Configure the name of the creator or user that will work with the ontology in the respective tool (Protégé Desktop)

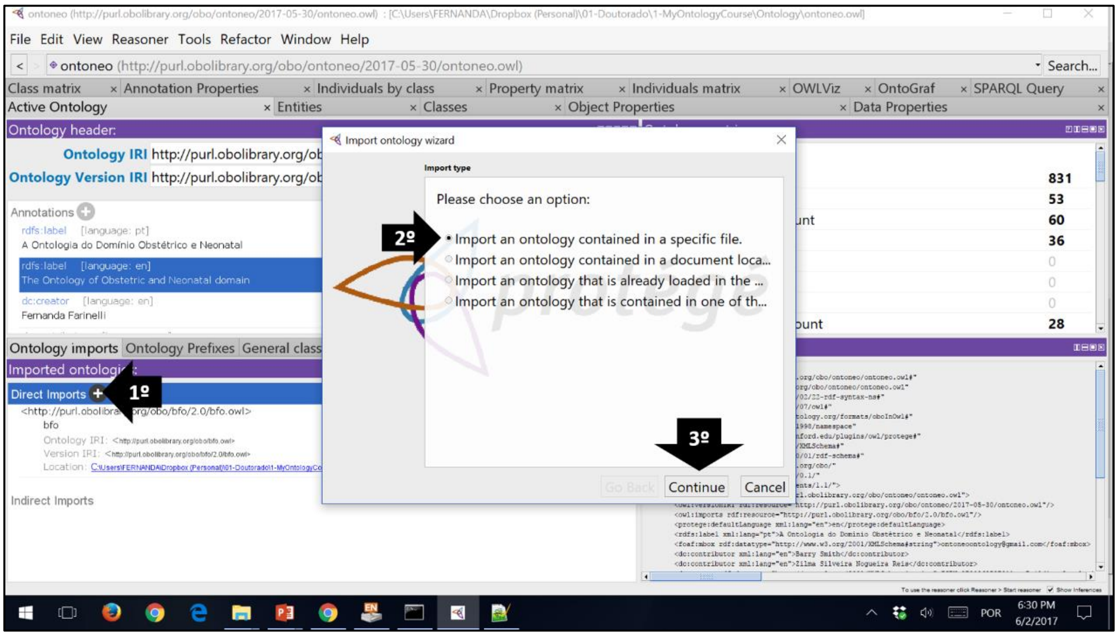


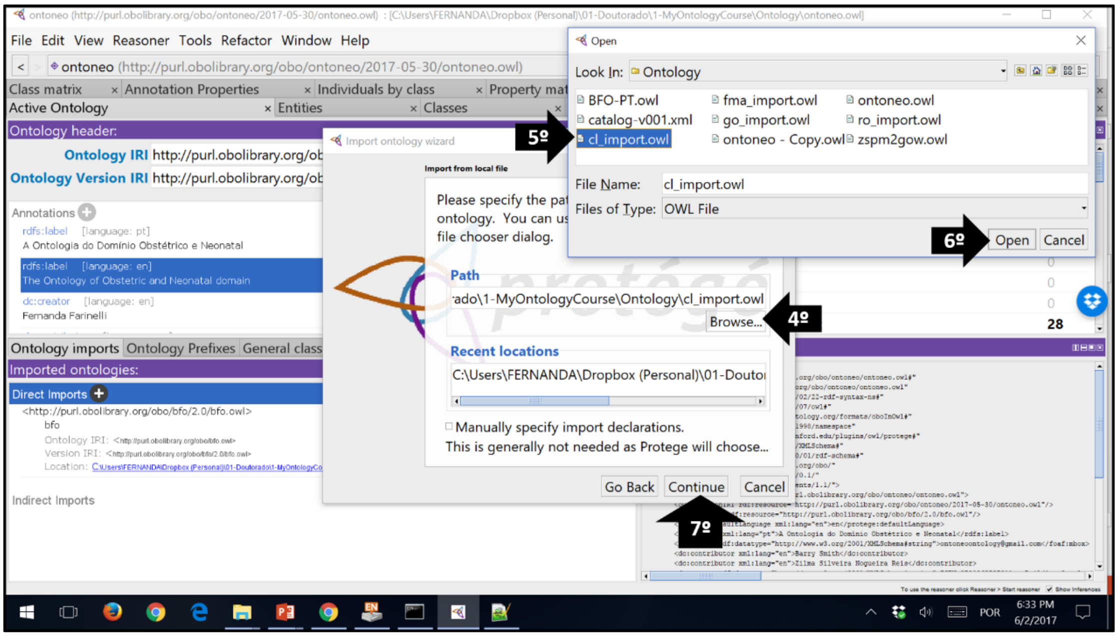
5) Determine the metadata that indicates the creator and the creation date of each new ontology element.

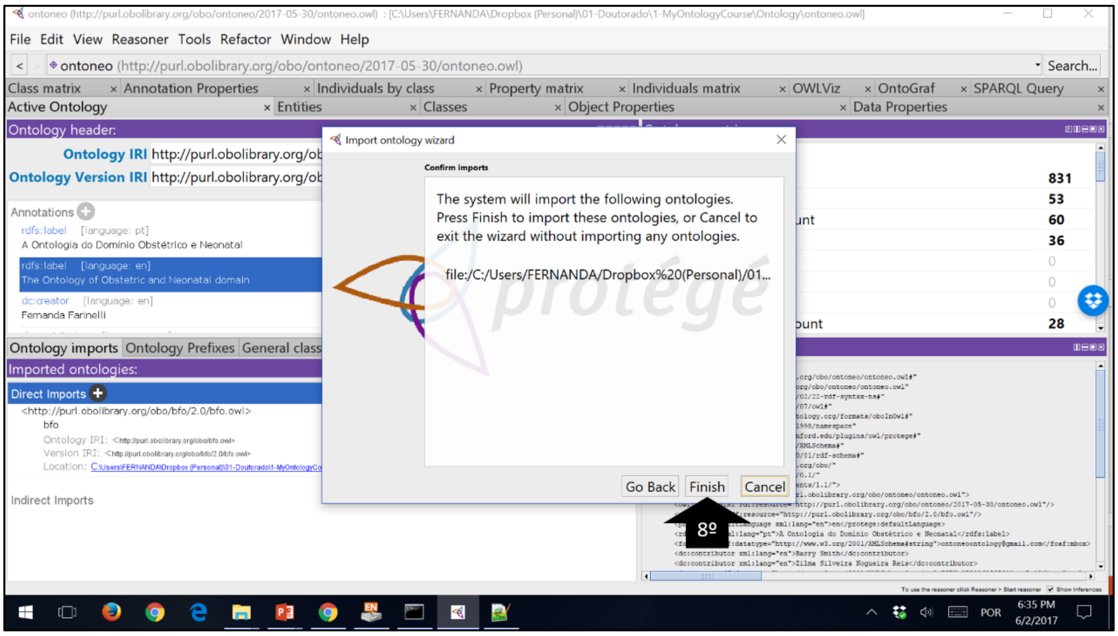

Preferences
New Ontologies
OWLViz
Plugins
Reasoner
Renderer
User details
Annotation template
Annotations
General
New Entities
New Entities Metadata
Annotate new entities with creator (user)
Creator property
http://purl.org/dc/elements/1.1/creator
1
Use Ctrl+Space to auto-complete name if property exists, or enter complete IRI.
Well known prefix names can also be used e.g. dc:creator
Creator value
Use user name
2
Use ORCID
Annotate new entities with creation date/time
Date property
http://purl.org/dc/elements/1.1/date
3
Use Ctrl+Space to auto-complete name if property exists, or enter complete IRI.
Well known prefix names can also be used e.g. dc:creator
Date value format
ISO-8601
4
Timestamp

APPENDIX 7: Importing ontologies on Protégé desktop

Step-by-step of how to import ontologies to Protégé desktop for the ontology development





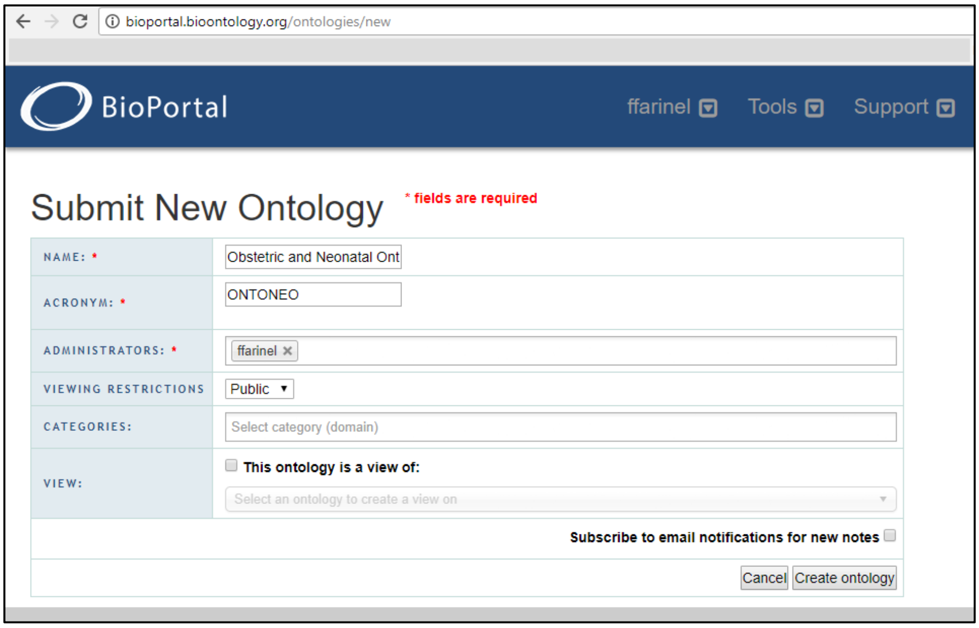

Import ontology wizard
Confirm imports
The system will import the following ontologies. Press Finish to import these ontologies, or Cancel to exit the wizard without importing any ontologies.
file:/C:/Users/FERNANDA/Dropbox%20(Personal)/01...
Go Back
Finish
Cancel
8º
File Edit View Reasoner Tools Refactor Window Help
ontoneo (http://purl.obolibrary.org/obo/ontoneo/2017-05-30/ontoneo.owl)
Ontology header:
Ontology IRI
Ontology Version IRI
Annotations
A Ontologia do Domínio Obstétrico e Neonatal
The Ontology of Obstetric and Neonatal domain
Fernanda Farinelli
Ontology imports
Ontology Prefixes
Imported ontologies:
Direct Imports
<http://purl.obolibrary.org/obo/bfo/2.0/bfo.owl>
Indirect Imports

APPENDIX 8: Publishing OntONeo at Bioportal

In order to publish OntONeo at BioPortal we followed the orientations available at the portal[98]. Figure 103 presents the initial submission form of OntONeo ontology, and Figure 83 presents a completed entry of OntONeo at BioPortal.

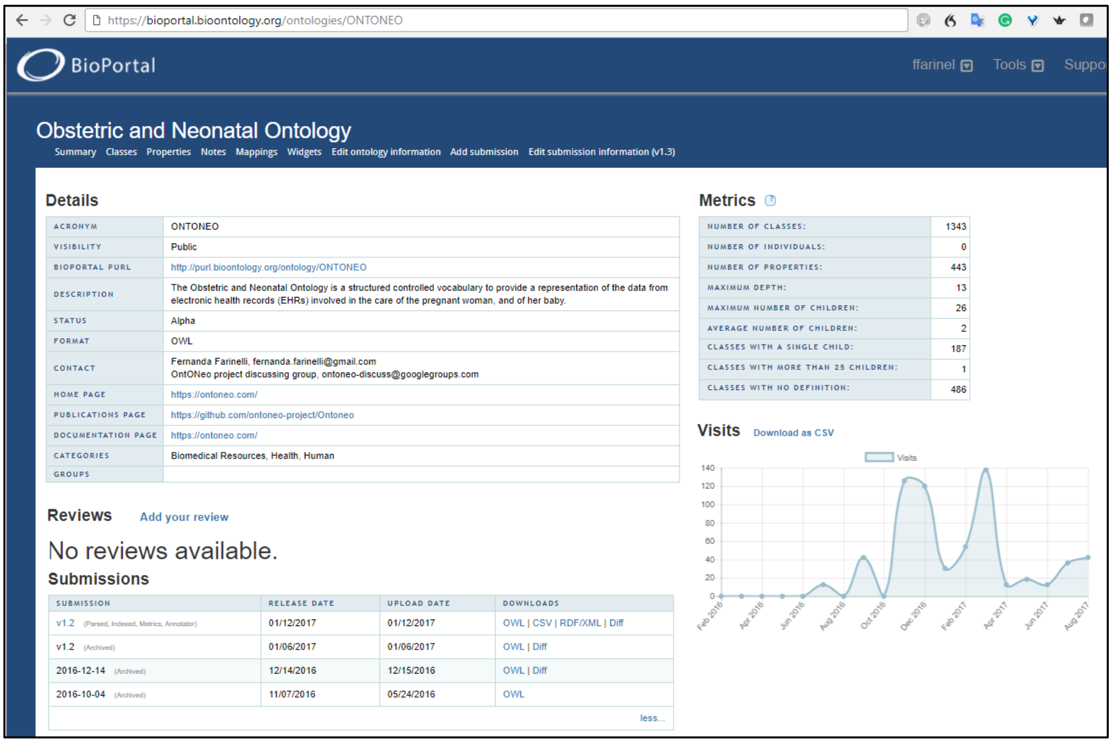
OntONeo BioPortal URL address is https://bioportal.bioontology.org/ontologies/ONTONEO.

Figure 103: Submission form of OntONeo ontology to BioPortal.



[98] How to submit an ontology to BioPortal: https://bioportal.bioontology.org/help#Submitting_an_ontology

Figure 104: OntONeo entry at Biportal (retrieved in January/2017).



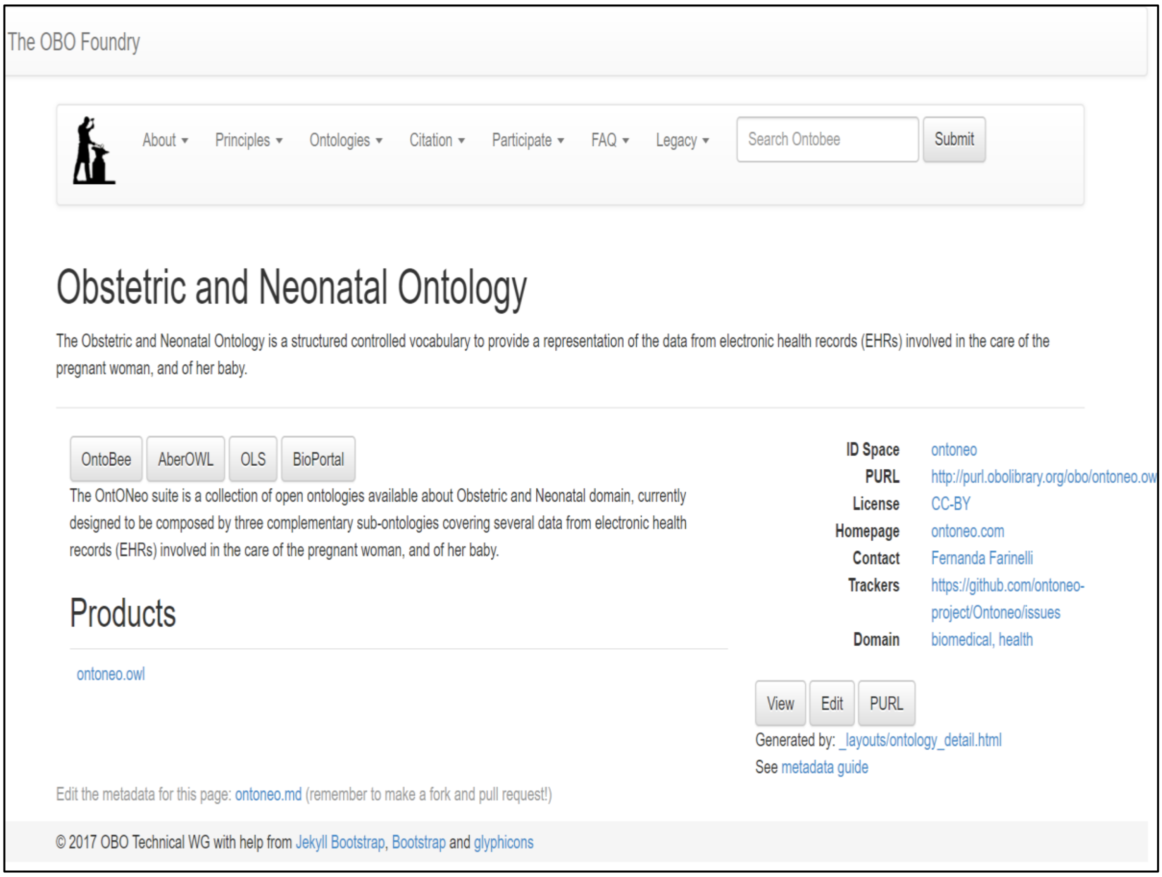
APPENDIX 9: Publishing OntONeo at OBO Foundry Portal

In the case of OBO Foundry Portal, we followed the the OBO Foudry guidelines[99] to register OntONeo as an ontology of OBO Foundry library.

OntONeo OBO Foundry address is: http://www.obofoundry.org/ontology/ontoneo.html

In addition, to be part of OBO Foundry (Figure 105) an ontology should follow the thirteen principles established by OBO Foundry Consortium. We demonstrates how OntONeo implements each one of these thirteen principles at section 4.4.

Figure 105: Reference of OntONeo at OBO Foundry Portal



[99] How to register an ontology at OBO Foundry Portal: http://www.obofoundry.org/faq/how-do-i-register-my-ontology.html