

**A terminological approach to knowledge organization
within the scope of endometriosis: the EndoTerm project**

Sara Micaela Pereira Carvalho

PhD in Linguistics

Specialization in Lexicology, Lexicography and Terminology

September 2018

Thesis submitted to fulfill the requirements for obtaining the doctorate degree

in Linguistics

Specialization in Lexicology, Lexicography and Terminology

Developed under the supervision of

Professor Rute Costa

and

Professor Christophe Roche

This thesis has been developed within the scope of a *co-tutelle* agreement between the Universidade NOVA de Lisboa and the Communauté Université Grenoble Alpes

A journey of a thousand miles

begins with a single step.

(Lao Tzu)

ACKNOWLEDGEMENTS

“If you want to go fast, go alone. If you want to go far, go together.”

(African proverb)

At the end of this journey, I would like, first and foremost, to express my deepest gratitude to my supervisors: Professor Rute Costa, you have been a mentor for me, encouraging me always to go one step further and providing me with countless opportunities to grow as a researcher and as a person. Professor Christophe Roche, thank you for introducing me into the world of ontologies and Knowledge Engineering, for your most valuable input, and for the trust you have put in me and in my work.

I would also like to thank Professor Teresa Roberto, for always having the right word of encouragement at the right time. My sincere thanks to Professors Teresa Lino and Raquel Amaro, as well as to Raquel Silva, for all their support and availability. I am also profoundly grateful to the late Professor Francisca Athayde. Her passion for Linguistics was a major influence in the academic path I have been following ever since that first year at university. Her wit and friendship will never be forgotten. Thank you to my PhD colleagues Margarida and Bruno, for the encouragement and the fruitful discussions concerning this amazing world of Terminology.

A special thanks to all my colleagues at the Águeda School of Technology and Management – University of Aveiro, especially to the current and previous Directors, for providing me with all the necessary conditions to pursue and achieve this goal.

My utmost appreciation and respect also go to the experts who have taught me so much about endometriosis and also to the numerous healthcare professionals worldwide who strive to fight this condition every day. To my fellow Endosisters, especially the MulherEndo ‘musketeers’ Susana, Dália, and Fátima: you are an inspiration to me, and I know that together, we will continue to change many lives and open people’s minds about endometriosis. Especially this past year, our motto became even more meaningful to me: “You never know how strong you are until being strong is your only choice.”

I would also like to thank all of my friends for their precious support and for understanding my long absences. A special thanks to my wonderful family, for being so supportive and caring, particularly my loving mother and father, who taught me and my brother the value of hard work, and that education is the best investment parents can make in their children’s future. I dedicate this thesis to my grandparents, our family’s heart and soul, with whom I learned that who you are is not defined by what you own.

Last but not least, I would like to thank my soul mate Bruno, for being my rock not only during the many rough days of this journey but throughout the 20 years that we have been together, always moving forward, side by side. Words cannot even begin to express my gratitude for everything you have done for our family and for me. Finally, I also dedicate this thesis to my sweet prince Miguel, who reminded me that life’s most precious gifts often lie in the smallest and most ordinary things. Remember, my son: “You—you alone will have the stars as no one else has them... In one of the stars, I shall be living. In one of them, I shall be laughing. And so it will be as if all the stars were laughing when you look at the sky at night... You—only you—will have stars that can laugh.” (Antoine de Saint-Exupéry, *The Little Prince*)

Uma abordagem terminológica à organização do conhecimento no âmbito da endometriose: o projeto EndoTerm

Sara Micaela Pereira Carvalho

RESUMO

A presente tese de doutoramento visa descrever a conceção, o desenvolvimento e a implementação do recurso EndoTerm, um produto terminológico dedicado à endometriose e que visa contribuir para uma organização e partilha mais eficazes do conhecimento especializado relativo a esta patologia.

Globalmente, os fundamentos teóricos e metodológicos do recurso terminológico em questão encontram-se ancorados nas sinergias entre a Terminologia – na dupla dimensão que a caracteriza – e as ontologias enquanto artefactos informáticos. A estrutura nuclear do produto EndoTerm reflete essa abordagem, já que este é composto por dois módulos autónomos que, no entanto, se articulam entre si: [1] o módulo conceptual, que tem por base princípios oriundos de recursos terminológicos biomédicos, tais como o SNOMED CT ou o Unified Medical Language System (UMLS), e que serviram de base à construção de uma ontologia formal; e [2] o módulo linguístico, cujo objetivo é enriquecer o recurso terminológico através da inclusão de termos, definições em língua natural e outra informação linguística que possa ser considerada relevante. Acresce que o conteúdo deste segundo módulo resulta da análise do EndoCorpus, um corpus monolíngue (em língua inglesa), constituído por textos especializados do domínio da endometriose e construído especificamente no âmbito do presente projeto.

Considera-se que um dos principais contributos do trabalho de investigação que aqui se apresenta está relacionado com uma proposta de otimização de redação de definições em língua natural. Mais concretamente, a proposta centra-se no papel decisivo das definições formais dos conceitos nessa tarefa de redação, nomeadamente através da criação de modelos (*templates*) de estruturas definicionais, resultantes do trabalho terminológico. Em suma, pensa-se que a metodologia proposta poderá contribuir para uma abordagem mais consistente e eficaz à redação de definições em língua natural em contextos comunicacionais especializados e multilíngues.

Por último, o produto EndoTerm poderá ainda servir de base ao desenvolvimento de recursos terminológicos e/ou lexicográficos relativos ao domínio da endometriose, destinados quer a utilizadores humanos, quer a não humanos, já que o conteúdo desta base de conhecimento pode ser operacionalizável, isto é, manipulável e partilhável em termos informáticos. Deste modo, os princípios teóricos e metodológicos que subjazem à criação deste recurso serão ilustrados através de um estudo de caso centrado num procedimento cirúrgico comumente realizado em pacientes com endometriose.

PALAVRAS-CHAVE: EndoTerm, endometriose, modelo de estrutura definicional em Terminologia, interoperabilidade, Terminologia, ontologias, Informática em Saúde.

A terminological approach to knowledge organization within the scope of endometriosis: the EndoTerm project

Sara Micaela Pereira Carvalho

ABSTRACT

This PhD thesis aims to describe the conception, development and implementation of EndoTerm, a knowledge-based terminological resource focused on endometriosis which seeks to contribute to a more effective way of organizing and sharing the current knowledge regarding this particular subject field.

Overall, EndoTerm's theoretical and methodological principles are anchored in the synergies between Terminology - with its double-dimensional nature - and ontologies. This is also illustrated by EndoTerm's core structure, composed of two modules that function in an autonomous, yet interrelated, way: [1] the conceptual module, supported by existing biomedical terminological resources, such as the SNOMED CT or the Unified Medical Language System (UMLS), and which constitutes the basis of a formal ontology; and [2] the linguistic module, whose goal is to enrich the knowledge base with terms, natural language definitions and other relevant linguistic information, and which results from the analysis of the EndoCorpus, a monolingual (English) corpus specifically built within the scope of this project and containing specialized texts, written by subject field experts, regarding endometriosis.

It is believed that one of the key contributions of this research work is its focus on formal conceptual definitions and on the drafting of natural language definitions, as well as on how the former can help ground the latter, especially via a set of definitional templates. Such a template-based methodology may provide a more consistent approach to the drafting of natural language definitions in more specialized – and hence, more restricted – communicative settings, in addition to contributing towards a more effective way of devising such natural language definitions in multiple languages.

Ultimately, EndoTerm can constitute a valuable foundation for existing or future terminological and lexicographic resources concerning the subject field of endometriosis and which can be targeted either at human or machine users, since EndoTerm content also allows operationalization, i.e. it can be manipulated and shared by a computer. For that reason, a case study involving a common surgical procedure in cases of endometriosis will be provided and explored, to show how EndoTerm content can be exported and, thus, shared, in both human- and machine-readable formats, respectively.

KEYWORDS: EndoTerm, endometriosis, definitional templates, interoperability, Terminology, ontologies, Health Informatics.

Une approche terminologique pour l'organisation des connaissances du domaine de l'endométriose: le projet EndoTerm

Sara Micaela Pereira Carvalho

RÉSUMÉ

Cette thèse de doctorat se propose de décrire la conception, le développement et la mise en œuvre d'EndoTerm, un outil terminologique consacré à l'endométriose et qui prétend contribuer à une organisation et à un partage plus efficaces du savoir spécialisé concernant cette pathologie.

Généralement, les principes théoriques et méthodologiques d'EndoTerm se trouvent ancrés dans les synergies entre la Terminologie - en sa nature bidimensionnelle - et les ontologies en tant qu'objets informatiques. Ceci se confirme dans la structure de base d'EndoTerm, qui est composée par deux modules autonomes, s'articulant néanmoins: [1] le module conceptuel, fondé sur des ressources terminologiques biomédicales existantes, telles que SNOMED CT ou Unified Medical Language System (UMLS), qui sont à la base d'une ontologie formelle; et [2] le module linguistique, dont le but est d'enrichir la ressource terminologique grâce à l'introduction de termes, de définitions en langue naturelle et d'autres informations linguistiques pertinentes. Le contenu de ce deuxième module découle de l'analyse d'EndoCorpus, un corpus monolingue (anglais) constitué par des textes spécialisés dans le domaine de l'endométriose, spécifiquement élaboré dans le cadre de ce projet.

L'un des principaux apports de ce travail de recherche concerne son inflexion sur les définitions conceptuelles formelles et sur la rédaction des définitions en langue naturelle, ainsi que sur la façon dont la première peut aider la deuxième, en particulier à travers un ensemble de modèles définitionnels. Une telle méthodologie peut fournir une approche plus cohérente de la rédaction des définitions en langue naturelle dans des contextes de communication plus spécialisés - et donc plus restreints -, en plus de contribuer à une conception plus efficace de ces définitions en plusieurs langues.

Finalement, EndoTerm peut constituer une base importante pour les ressources terminologiques et / ou lexicographiques existantes ou futures concernant le domaine de l'endométriose et qui peuvent être dirigées soit à des utilisateurs humains, soit à des machines, puisque le contenu d'EndoTerm peut être opérationnalisé à partir d'un ordinateur. Nous présenterons, donc, une étude de cas impliquant une procédure chirurgicale commune dans les cas d'endométriose, afin de montrer comment le contenu d'EndoTerm peut être exporté et, par conséquent, partagé, dans des formats lisibles par l'homme et par les machines, respectivement.

MOTS-CLÉ: EndoTerm, endométriose, modèle définitionnel en Terminologie, interopérabilité, Terminologie, ontologies, Informatique de Santé

LIST OF ABBREVIATIONS

AI:	Artificial Intelligence
ART:	Assisted Reproductive Technologies
ASRM:	American Society for Reproductive Medicine
AUI:	Atom Unique Identifier
BFO:	Basic Formal Ontology
CAST:	Categorical Structure
CCAM:	Classification Commune des Actes Médicaux
CCI:	Canadian Classification of Health Interventions
CUI:	Concept Unique Identifier
DO:	Disease Ontology
eEIF:	eHealth European Interoperability Framework
EFI:	Endometriosis Fertility Index
EHR:	Electronic Health Record
EMR:	Electronic Medical Record
EPHect:	Endometriosis Phenome and Biobanking Harmonisation Project
EUPATI:	European Patient's Academy
FMA:	Foundational Model of Anatomy
FOAF:	Friend Of A Friend
GALEN:	Generalized Architecture for Languages, Encyclopaedias and Nomenclatures in medicine
GRAIL:	GALEN Representation and Integration Language
HTML:	Hypertext Markup Language
HTTP:	Hypertext Transfer Protocol
ICD-10-PCS:	International Classification of Diseases - 10th edition - Procedure Coding System
ICD:	International Classification of Diseases
ICHI:	International Classification of Health Interventions
IRI:	International Resource Identifier
ISO:	International Standardization Organization
IVF:	In Vitro Fertilization
JSON:	JavaScript Object Notation
KE:	Knowledge Engineering
KP:	Knowledge Pattern
KRC:	Knowledge Rich Context
KWIC:	KeyWord In Context
LD:	Linked Data
LESS:	Laparoendoscopic single-site (surgery)

LESSCAR: Laparoendoscopic Single-Site Surgery Consortium for Assessment and Research

LLOD: Linguistic Linked Open Data

LOD: Linked Open Data

LUI: Lexical Unique Identifier

MeSH: Medical Subject Headings

NCI: National Cancer Institute

NHS: National Health Service

NLM: National Library of Medicine

NOTES: Natural Orifice Transluminal Endoscopic Surgery

NSAID: Nonsteroidal Anti-inflammatory Drug

OBO: Open Biomedical Ontologies

OGMS: Ontology for General Medical Science

OWL: Web Ontology Language

PHR: Personal Health Record

QoL: Quality of Life

RDF: Resource Description Framework

RRF: Rich Release Format

SKOS: Simple Knowledge Organization System

SPARQL: SPARQL Protocol and RDF Query Language

SUI: String Unique Identifier

SW: Semantic Web

TKB: Terminological Knowledge Base

TS: Terminological System

UML: Unified Modeling Language

UMLS: Unified Medical Language System

URI: Uniform Resource Identifier

URL: Uniform Resource Locator

W3C: World Wide Web Consortium

WERF: World Endometriosis Research Foundation

WES: World Endometriosis Society

WHA: World Health Assembly

WHO-FIC: World Health Organization - Family of International Classifications

WHO: World Health Organization

XML: eXtensible Markup Language

TABLE OF CONTENTS

RESUMO	iv
ABSTRACT	v
RÉSUMÉ	vi
LIST OF ABBREVIATIONS	vii
Introduction	1
A. Background and motivation.....	1
B. Objectives, scope, and research questions	2
C. Thesis structure.....	5
PART I: THE BRAVE NEW WORLD OF HEALTHCARE	8
Chapter 1 – Healthcare in the 21st century: challenges and opportunities for terminology work	9
1.1. Introduction	9
1.2. eHealth, mHealth, uHealth... where does it go from here?	11
1.3. Digital health literacy and the rise of the ePatient	16
1.4. Interoperability in healthcare: caught between hype, hope and reality	21
1.5. Healthcare at a crossroads: goodbye data silos, hello Linked Data?	25
PART II: THEORETICAL FRAMEWORK.....	30
Chapter 2 – Terminology: where the conceptual and the linguistic intertwine	31
2.1. Introduction	31
2.2. Terminology's double dimension: two sides of the same coin	33
2.3. Terminology and ontologies	35
2.4. The definition in Terminology	41
Chapter 3 – Representing, organizing and sharing knowledge within the biomedical domain: towards an interoperable future?	46
3.1. Introduction	46
3.1.1. The dynamics of knowledge representation in healthcare	48
3.2. Organizing biomedical knowledge: learning from the past and innovating towards the future	51
3.2.1. The early days of knowledge systematization and classification in Medicine: lessons from history	51
3.2.2. A disease is a disease is a disease: challenges for the 21st century biomedical terminological resources	57
3.2.3. Is What You See Really What You Get? Current biomedical terminological systems and <Endometriosis>	62
3.2.3.1. The International Classification of Diseases (ICD)	65
3.2.3.2. The Medical Subject Headings (MeSH).....	68
3.2.3.3. SNOMED CT	72
3.2.3.4. The Disease Ontology (DO).....	82
3.2.3.5. The Unified Medical Language System (UMLS)	86
3.3. Sharing biomedical knowledge: can standardization and diversity coexist?	93
3.4. Setting the scene for biomedical terminological systems in the new millennium.....	96

PART III: ENDOTERM – RATIONALE, METHODOLOGY, AND IMPLEMENTATION	101
Chapter 4 – Endometriosis: the most common disease one has never heard of?	102
4.1. Endometriosis: facts and figures	102
4.2. Endometriosis: a riddle wrapped in a mystery inside an enigma	106
4.3. The burdens of endometriosis: a delayed opportunity for millions of women?	113
4.4. Endometriosis and its many ‘faces’: between concept and term	117
4.5. The dilemmas of endometriosis: what to classify and how?	121
Chapter 5 – An overview of EndoTerm.....	130
Chapter 6 - EndoTerm’s conceptual backbone	134
6.1. Getting to know the subject field: the essential first step in terminology work.....	134
6.2. Identifying EndoTerm’s concept of departure or finding the needle in the haystack	137
6.3. Knowing a concept by the company it keeps: EndoTerm’s conceptual framework	148
Chapter 7 – Text analysis for onomasiological purposes: EndoTerm’s linguistic dimension	161
7.1. Introduction	161
7.2. The EndoCorpus: criteria for corpus design and collection	163
7.3. The EndoCorpus: the semi-automatic processing stage	167
7.4. The example of “hysterectomy” in the EndoCorpus: data analysis and systematization	173
7.5. Saying vs modelling? The added value of a complementary approach	181
Chapter 8 – LESS can indeed be more: linguistic and conceptual challenges in the age of interoperability	184
8.1. Introduction	184
8.2. LESS: a Brave New World for surgery?	185
8.3. The conceptual side of LESS	190
8.4. Terminological definitions in EndoTerm: two examples	200
8.5. SNOMED CT integration: a proposal	203
Concluding remarks.....	206
Bibliography	210
LIST OF FIGURES	230
LIST OF TABLES	232
ANNEXES.....	233
ANNEX 1	234
ANNEX 2	236
ANNEX 3	239
ANNEX 4	241
ANNEX 5	244

Introduction

A. Background and motivation

Over the last few decades, the biomedical domain has been undergoing major changes at various levels, leading to significant challenges and opportunities that several scientific areas have been trying to handle. Terminology, with its interdisciplinary potential, can make a valuable contribution to the biomedical domain, either in what concerns the improvement of specialized communication among the different stakeholders or as regards more effective knowledge representation, organization, and sharing.

From a knowledge organization perspective, Terminology can contribute by supporting the development of new healthcare-related resources that rely on a stable and reliable conceptual framework, such as classifications, thesauri, and other databases. The concrete impact can be felt, for instance, in the field of information retrieval, which can greatly benefit from the optimization strategies stemming from terminological work. Furthermore, the increasingly stronger collaboration between Terminology and ontologies, to be explored in this thesis, allows the resulting terminological products to be operationalized, i.e. manipulated and shared by a computer, which is in line with the ‘interoperability-Semantic Web-Linked Data’ framework that has been characterizing healthcare in recent years. Such a setting can then be materialized in Electronic Health Records (EHRs), Personal Health Records (PHRs), or in patient-oriented apps.

On the other hand, with the rising need for reliable and ‘patient-friendly’ health-related information (in the sense that patients can clearly understand it), it is believed that terminological work can also play a key role in the development and dissemination of products targeted at human users, such as specialized reference materials (e.g. online dictionaries, glossaries, etc.), training materials for expert patients or society in general, or patient-oriented ‘products’ provided by healthcare institutions themselves, namely brochures about a given disease, or even the information that is made available at an institutional website.

In this respect, this thesis will focus on the domain of endometriosis, a chronic gynecologic disease which arises when tissue similar to the endometrium (i.e. the lining of the uterus) is found outside the uterine cavity (Dunselman et al., 2014). It affects approximately 176 million women worldwide (Adamson, Kennedy, & Hummelshoj, 2010; Rogers et al., 2009), and has a significant economic and social impact in today's society, accounting for severe absenteeism and productivity loss among millions of women each year. Despite its worldwide prevalence and considerable impact, endometriosis is still a relatively unknown medical condition, even among healthcare professionals, and is underrepresented in most biomedical terminological resources.

Hence, this research work aims to make a contribution to bridging that gap by describing the conception, development, and implementation of EndoTerm, a knowledge-based terminological resource focused on endometriosis (cf. Carvalho, Costa, & Roche, 2016a, 2016b; Carvalho, Roche, & Costa, 2015) which seeks to contribute to a more effective way of organizing and sharing the current knowledge regarding this particular subject field.

B. Objectives, scope, and research questions

Thus, and as will be shown throughout this work, EndoTerm may constitute a valuable foundation for existing or future terminological and/or lexicographic resources concerning the subject field of endometriosis and which can be targeted either at human or at machine users, since EndoTerm content also allows operationalization, i.e. it can be manipulated and shared by a computer. In fact, EndoTerm's theoretical principles rely on the synergies between Terminology - with its double-dimensional nature - and ontologies. This is also illustrated by EndoTerm's core structure, composed of two autonomous, yet interrelated modules (conceptual and linguistic), supported by a mixed methodology (onomasiological and semasiological).

When applied to the biomedical domain and, in this case, to the EndoTerm project, these principles have contributed to substantiate the following assertions:

[a] the fact that, at a conceptual level, a conceptualization proposal can greatly benefit from the combination of i) a specific difference, Aristotelian-based approach; ii)

hierarchical and non-hierarchical conceptual relations; and iii) a customized categorial structure (in this case, the adaptation of the ISO 1828 (2012) categorial structure of surgical procedures);

[b] the fact that, at the linguistic level, the tasks related to corpus processing and corpus analysis can be optimized if supported by a previous (semi-formal) systematization of the field's conceptual framework, as well as by the use of Knowledge Patterns (KPs), derived from the notion of Knowledge-Rich Contexts (KRCs) (Meyer, 2001);

[c] the fact that the semantic information extracted from the corpus may, in turn, lead the terminologist to elicit conceptual information, thereby contributing to enrich the existing conceptual systems, in an iterative process;

[d] the fact that, on the one hand, formal concept definitions can provide a stable and logically consistent foundation, either to the drafting of new natural language definitions or to the enhancement of existing ones;

[e] the fact that both the conceptual and the linguistic information can be operationalized in both human- and machine-readable formats, thereby allowing the resulting terminological information to be interoperable.

Ultimately, it is believed that one of the key contributions of this research work is its focus on formal conceptual definitions and on the drafting of natural language definitions, as well as on how the former can help ground the latter, especially via a set of definitional templates. Such a template-based methodology may provide a more consistent approach to the drafting of natural language definitions in more specialized – and hence, more restricted – communicative settings, in addition to contributing towards a more effective way of devising such natural language definitions in multiple languages.

In short, the current research work aims to provide answers to the questions that follow:

[1] What is a concept within the biomedical domain and how is it represented?

[2] How are concepts related to each other within the existing biomedical terminological systems?

[3] Are hierarchical relations enough to fully define a concept and distinguish it from another concept, or should non-hierarchical relations be taken into consideration as well?

[4] How can that conceptual framework assist in the analysis of a specialized corpus?

[5] On the other hand, how can conceptual information be expressed – either explicitly or implicitly – in that specialized corpus?

[6] How can the aforementioned conceptual framework support the drafting of natural language definitions?

[7] How do the conceptual and the linguistic dimensions interrelate and how can that be systematized and shared among both human and machine users?

[8] How can a conceptualization be operationalized to achieve the aspired interoperability in this field?

To address the preceding research questions, a set of deliverables have been developed within the scope of this thesis, namely:

a) a **domain tree** resulting from the domain acquisition stage and presenting an overview of endometriosis as the subject field under analysis;

b) a **semi-formal conceptualization proposal** concerning a type of surgical procedure commonly undertaken in cases of severe endometriosis. This proposal derived from the analysis of a set of biomedical terminological resources, as well as from the conjugation of an Aristotelian, specific difference approach with an adaptation of the ISO 1828 (2012) categorial structure of surgical procedures;

c) a **monolingual (English) corpus of specialized texts from this domain** – the EndoCorpus –, specifically created within the scope of this thesis;

d) a **terminological resource** – containing both conceptual and linguistic information – focused on surgical procedures in endometriosis, whose conceptual module can be exported into a formal ontology, and whose linguistic module can be shared via an HTML web page. Underlying this resource is also a theoretical and methodological framework that supports content development and future expansion.

C. Thesis structure

As far as its core structure is concerned, the body of the thesis has been divided into three main parts. The first part, which entails **Chapter 1**, addresses some of the key concepts in healthcare at the onset of the 21st century – such as eHealth, mHealth, interoperability, the Semantic Web (SW), and Linked Data (LD) – and reflects upon the role that Terminology, especially within the scope of knowledge organization, can play in a setting characterized by numerous societal, economic, and technological challenges. In addition, this chapter focuses on the recent trend towards a patient-centered healthcare model, where patients take the driver's seat and play an increasingly active role in everything related to their health. Hence, the chapter also explores topics such as patient empowerment, health literacy, and the relatively recent notion of expert patient.

The second part of the thesis focuses on the theoretical framework underlying the EndoTerm project. **Chapter 2** introduces the theoretical framework that supports this research work. It starts by focusing on Terminology's double dimension and how that approach has impacted the current trend towards the operationalization of terminological products, namely via the promising connection between Terminology and ontologies. Examples of this recent approach, such as Termontography (Kerremans & Temmerman, 2004; Kerremans, Temmerman, & Tummers, 2004; Temmerman & Kerremans, 2003) and Ontoterminology (Roche, 2012, 2015; Roche, Calberg-Challot, Damas, & Rouard, 2009), will be described. The following section of the chapter refers to definitions and their role in Terminology. It builds up on the classical Aristotelian approach to explore the notion of intensional definition advocated, for instance, in the ISO standards - 1087-1 (2000) and 704 (2009). The final section refers to the specificities and collaboration potential involving formal concept definitions and natural language definitions, namely by expanding on the topic of definitional templates (cf. (Durán-Munoz, 2016; Faber, 2012, 2015) and how these can contribute to the drafting of more consistent natural language definitions.

Chapter 3 deals with the dynamics of knowledge representation, knowledge organization, and knowledge sharing within the biomedical domain. Firstly, it outlines the diversity that underlies knowledge representation in this particular field, and the more recent trend towards increasingly visual and multimedia-based forms of representation. Secondly, the chapter provides a thorough historical background regarding the development of terminological resources in this subject field. It highlights the considerable development of biomedical terminological resources, particularly in the last few decades: from plain lists, these resources have evolved into compositional systems and, more recently, into formal concept representation systems with a clear link to Artificial Intelligence- and Knowledge Engineering-based approaches (Duclos et al., 2014). Afterwards, an overview of some of the current terminological resources in healthcare is presented, with special emphasis on the International Classification of Diseases (ICD), the Medical Subject Headings (MeSH), the Disease Ontology (DO), SNOMED CT, and the UMLS. This outline will allow a more comprehensive look into how these resources specifically depict Endometriosis. The final section of the chapter resorts to a concrete example of a surgical procedure undertaken in severe cases of endometriosis to illustrate how knowledge stemming from the resources above can be shared, especially within the scope of the SW and LD.

The third and final part of the thesis is dedicated to the conception and implementation of EndoTerm: **Chapter 4** deals with the thorough presentation of endometriosis as the chosen subject field. It explores not only its clinical dimension (i.e. signs and symptoms, etiology, diagnosis, treatment) but also its economic and social burden. In addition, it outlines the existing classification systems referring to the disease. **Chapter 5** provides an overview of the EndoTerm project, focusing mainly on its core structure, use cases, and supporting methodology.

The two subsequent chapters characterize and explore EndoTerm's conceptual and linguistic dimensions, respectively. **Chapter 6** starts by describing how knowledge concerning the subject field in question has been acquired by the terminologist and presents the first attempt to systematize the gathered information via a domain tree. The core of the chapter, however, is dedicated to EndoTerm's conceptual backbone and to the different steps that led to the conception of a semi-formal conceptual

representation supported by the specific difference approach, a set of predefined non-hierarchical conceptual relations adapted from SNOMED CT and the UMLS, as well as a categorial structure. This semi-formal representation has been created via CMap Tools. **Chapter 7**, on the other hand, is dedicated to the collection, processing, and analysis of a specialized corpus devised especially for this purpose – the EndoCorpus. The chapter aims to show how the previous conceptual systematization of the domain supported the corpus analysis, which has also relied on a set of internal and external Knowledge Patterns (cf. Meyer, 2001). The corpus data has led to the creation of a set of lexical networks, also via CMap Tools, that have been compared to the existing micro-conceptual systems. The final section of the chapter shows how these lexical networks can help the terminologist elicit the necessary conceptual information so as to enhance the aforementioned conceptual representation. Finally, **Chapter 8** aims to illustrate, via a case study involving a specific type of <Surgical procedure>, how the conceptual and the linguistic dimensions can interrelate and contribute to mutual enrichment, with special emphasis being placed on the drafting – or enhancement – of natural language definitions. Furthermore, the chapter will show how EndoTerm content can be shared with both human- and machine-users, with the latter being in line with current trends in healthcare and biomedical terminological systems as regards interoperability.

Part I: The Brave New World of Healthcare

Chapter 1 – Healthcare in the 21st century: challenges and opportunities for terminology work



Figure 1 – Word cloud with some of the key notions of healthcare today¹

1.1. Introduction

In recent decades, healthcare, as other sectors, has been impacted by a series of critical changes resulting mainly from societal, economic, and technological factors. Population ageing, for instance, is one of the most significant challenges for humankind as a species, both nowadays and in the upcoming years: according to the latest projections (Statista, 2018; The World Bank, 2018; United Nations, 2015), the number of people aged 60 or over, currently at around 950 million, will increase dramatically to 2.1 billion in 2050. Conversely, the “oldest-old persons” (United Nations, 2015), i.e. those aged 80 and over, will experience an even more striking rise, from the present 140 million to almost 450 million in 2050.

Due to the stated population ageing and to a growing prevalence of chronic or noncommunicable diseases (NCDs), namely cardiovascular diseases, cancer, chronic respiratory diseases and diabetes², there has been a steady increase in public health expenditure worldwide, as well as constraints on healthcare budgets. A report issued by the European Union (EU) (Przywara, 2010) projected that in 2060, public health

¹ Created using the <https://www.wordclouds.com/> software.

² The World Health Organization [WHO] estimates that each year, NCDs are collectively responsible for approximately 70% of deaths worldwide (<http://www.who.int/ncds/en/>).

expenditure would account for approximately 10% of the Member States' Gross Domestic Product (GDP), as opposed to the 6.8% in 2010. The sustainability of public healthcare systems has, in fact, become paramount in the current political, social and economic debate (Liaropoulos & Goranitis, 2015).

In an attempt to provide solutions for the problems described above, as well as for other pressing issues, such as the ongoing health workforce shortage of about 10 million professionals globally – expected to become approximately 13 million in 2035 (Aluttis, Bishaw, & Frank, 2014; WHO, 2016b), technology has been playing a decisive role in the last few years by revolutionizing healthcare, not only regarding service provision itself, but also the interaction between the different stakeholders in this sector.

Technological innovation has, indeed, become the driving force of this healthcare revolution, undergoing an exponential growth that has been unprecedented thus far. Top advances such as nanotechnology, 3D printing, robotic surgery, genomics³, wearable technology and sensors⁴, as well as the use of virtual, augmented and/or mixed reality, constitute the basis of “Quantified Health” (Kraft, 2017), an extension of the Quantified Self movement (Swan, 2013; Wolf, 2010). According to this approach, patients (and citizens in general) use smart devices and applications to track health data – such as heart rate, glucose level or blood pressure –, to prevent and manage acute and chronic diseases. Besides adopting healthier lifestyles, these users also share this data with their physicians and healthcare institutions.

In other words, the era of **P4 Medicine**, as coined by (Flores, Glusman, Brogaard, Price, & Hood, 2013; Hood & Flores, 2012), appears to be in full development nowadays, as healthcare is becoming more **Predictive** (by predicting specific health threats more

³ In this respect, reference should be made to the latest trend towards affordable and fast genome sequencing, with several start-ups attempting to reach the promising goal of providing a full personal genome sequence under 1,000 US\$ before 2020. Very recently (March 2018), an American research team has actually broken the Guinness World Record by sequencing a human genome in less than a day (about 20 hours). (<https://bit.ly/2FKoTKB>)

⁴ This refers to the application of the Internet of Things (IoT) concept to healthcare. As an example, some of the latest wearable technologies, used mainly for remote health monitoring, include the Qardicore, a wireless ECG chest strap (<https://www.getqardio.com/>); the Accendowave, a headband which aims to minimise pain and discomfort (<http://accendowave.com/>); and PillDrill, a smart medication tracking system (<https://www.pilldrill.com/>).

accurately and effectively and by tackling appropriate responses, both locally and worldwide), **Personalized** (by allowing a better and more customized response to disease via personalized medication, for instance), **Preventive** (by preventing the emergence of potential outbreaks or other global incidents), and **Participatory** (by resorting to active and networked consumers).

At the core of this paradigm shift are some of the concepts illustrated by the word cloud, especially **eHealth**. Often used as an umbrella term to refer to the use of Information and Communication Technologies (ICT) in the prevention, diagnosis, treatment, monitoring and management of diseases, it has gradually been accompanied by **mHealth**, the application of the eHealth concept to the increasingly pervasive use of mobile devices. This chapter will also explore the notions of **patient empowerment**, **health literacy**, and **expert patient**, all of which referring, to some extent, to the patients' pivotal role regarding their health, a trend which has been paving the way into an increasingly patient-centered healthcare model. The following sections will focus on the challenges that underlie the dissemination of healthcare provision across multiple departments, institutions, and countries. Therefore, current 'hot topics' in healthcare, particularly **interoperability**, **Big Data**, **Linked Data** and the **Semantic Web** will also be addressed. The final section aims to discuss the subsequent impact of all these ongoing changes in healthcare worldwide, as well as the role that terminology work and this particular project can play in such a context. Disruptive phenomena like these do entail challenges, barriers and limitations, but they also bring massive opportunities and benefits, as will be shown in this first chapter.

1.2. eHealth, mHealth, uHealth... where does it go from here?

Regarded as one of the key elements underlying the technology-based revolution in healthcare, eHealth (or e-health) is a somewhat recent term, coined originally by the industry and marketing sectors, in line with other "e-" buzzwords. It has been gaining momentum in the health-related literature since 2000 and is now a rapidly expanding field at a global level.

Due to the hype within this particular sector, a plethora of related terms, such as Medicine 2.0, Health 2.0, Health 3.0, mHealth and Telehealth, have been created and

used interchangeably throughout the literature, being often perceived as synonyms of eHealth (Boogerd et al., 2015). However, and within the context of this work, the umbrella term **eHealth** will be used rather than Medicine 2.0, Health 2.0 or Health 3.0, since these notions are not consensual in the literature, and are, furthermore, too closely linked to terms describing the evolution of the Web⁵, which may affect their scope and understanding. Also, the term eHealth has been adopted as such by international guidelines on this matter and by institutions such as the WHO and the European Commission (EC). On the other hand, mHealth and Telehealth have been identified as examples of eHealth, as will be shown in the following paragraphs.

Providing a clear and precise definition for eHealth has also proven to be challenging. From 2001 up to now, several authors have attempted to do that, with little success (Boogerd et al., 2015; Eysenbach, 2001; Oh et al., 2005; Pagliari et al., 2005)⁶ and with proposals ranging from very generic (“the use of ICT for health” – WHO⁷) to highly specific (“an emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies” (Eysenbach, 2001, p. 1)). Among the different definitions, Eysenbach’s has been one of the most consensual and most widely cited, as it aims to go beyond the basic postulate of *eHealth = Internet + Medicine* by acknowledging eHealth also as “a state-of-mind, a way of thinking, an attitude, and a commitment for networked, global thinking, to improve healthcare locally, regionally and worldwide” (ibid.).

In light of the preceding context, eHealth has been identified as a priority by the WHO and the EC, with both institutions launching national and global eHealth implementation programs, supported by a set of key documents, most notably: i) the *World Health Assembly’s (WHA) eHealth resolution* (58.28, 2005)⁸; ii) a joint initiative

⁵ A diagram containing the various stages of the WWW can be consulted at <https://bit.ly/2K5Ejed>.

⁶ Boogerd et al. have, more recently, reinforced the need to devise a clear definition of eHealth, as well as a taxonomy of the field. For that reason, they submitted a call to the *Journal of Medical Internet Research* in 2015, asking readers to answer a four-question survey on their ideas concerning the definition of eHealth. This survey is still available at <http://tinyurl.com/eHealthdef>.

⁷ cf. <http://www.who.int/ehealth/en/>.

⁸ This resolution highlighted the importance of eHealth and urged the Member States to collaborate in its implementation at an international level. It is available at <https://bit.ly/2HhUdUR>.

entitled *National eHealth Strategy Toolkit* (2012)⁹, promoted by the WHO and the International Telecommunication Union (ITU); iii) and the EC's *eHealth Action Plan* (2012), which details the Commission's operational objectives to be carried out within the 2012-2020 timeframe¹⁰.

At European level, some eHealth pilot projects have already been tested and put into action, especially those on a cross-border scale, as is the case of *ePrescription* and *eDispensing*¹¹, both of which developed within the scope of the former European Patients Smart Open Services (epSOS) project¹². **Table 1** comprises further examples of eHealth tools and services:

Category	Examples
1. Individual electronic health information	• Electronic Health Records (EHRs) • Electronic Medical Records (EMRs) • Personal Health Records (PHRs)¹³
2. Healthcare communication and collaboration	• Care plan management • Appointment booking and management • ePrescribing • eDispensing • Electronic health event summaries
3. Healthcare service delivery tools	• Medications management • Clinical decision support • Alerts monitoring and management • Chronic disease management

⁹ This document is a step-by-step practical guide aimed at the national implementation of an eHealth strategy, destined to ministries of health and other relevant government departments and agencies. It can be consulted via the following link: <https://bit.ly/2qPAFNu>

¹⁰ The document is available for consultation at <https://bit.ly/2DPtYyL>

¹¹ **ePrescription** is defined as "the electronic prescribing of medicine with the use of software by a legally authorized health professional and the electronic transmission of said prescription data to a pharmacy, where the medicine can then be dispensed"; **eDispensing**, on the other hand, is "the electronic retrieval of a prescription and the dispensing of the medicine to the patient as indicated in the corresponding ePrescription" (<http://www.epsos.eu/epsos-services/eprescription.html>).

¹² Although this project was discontinued in 2014, it is still possible to access relevant data and documentation under <http://www.epsos.eu/home.html>.

¹³ **EHRs** can be defined as "records in electronic format that contain a patient's medical history, diagnoses and treatment, medications, allergies, immunizations, radiology images and laboratory results. They provide immediate and secure information to authorized users across different healthcare institutions"; **EMRs**, in turn, are "in-house digital versions of the traditional paper charts in clinicians' offices, clinics and hospitals". They contain notes and information collected by and for the clinicians and are used by providers for diagnostic and treatment within one specific healthcare organization; Finally, **PHRs** refer to "records designed to be set up, accessed, managed, shared and controlled by patients themselves. They can include information from a variety of sources: clinicians, home monitoring devices and the patients". Sources: <https://bit.ly/2K50Jwc>; <https://bit.ly/2K8FY2z>.

	● Telehealth¹⁴ ● mHealth
4. Healthcare information and knowledge	● Consumer health knowledge sources ● Healthcare provider knowledge sources ● Distance learning for healthcare professionals and respective electronic sources
5. Healthcare management and administration	● Adverse event monitoring ● Risk analysis ● Surveillance and at-risk identification ● Health policy development ● Health program design and optimization

Table 1 – Categories within eHealth and respective examples

(adapted from: (WHO & ITU, 2012; WHO, 2016a))

When compared to other subtypes of eHealth, **mHealth** (= mobile health), defined as the medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), smart watches and other body-worn devices (cf. (EC, 2014, p. 2; WHO, 2011b, p. 6)), has acquired a particular status nowadays, due to the impressive and rapid advances in mobile technologies and applications¹⁵. According to the report by WHO (2011, p. 1), mHealth “has the potential to transform the face of health service delivery across the globe”, as it allows a more widespread access and provision of healthcare, especially in remote areas and in developing countries, which lack the infrastructure to support the Internet or traditional health services. The substantial proliferation of health-related applications for mobile devices is, however, “blurring the distinction between the

¹⁴ Telehealth can be defined as the “delivery of healthcare services, where patients and providers are separated by distance” (WHO, 2016a, p. 56). It can be either synchronous or asynchronous and encompasses five subtypes: teleradiology; teledermatology; telepathology; telepsychiatry; and remote patient monitoring.

¹⁵ It is estimated that, in 2020, there will be 6.1 billion smartphone subscriptions worldwide (compared to 2.7 billion in 2014) and a striking amount of 17 Exabytes of monthly smartphone traffic (versus 2.1 Exabytes in 2014) (Cf. <https://bit.ly/2Hmq0Aj>; <https://bit.ly/2dk8wHh>). Besides that, projections by PricewaterhouseCoopers (<https://pwc.to/2tIPIAa>) and Statista (<https://bit.ly/2Ac9zIR>) set the global mHealth market at around \$60 billion in 2020, and in that same year, more than one-third of smartphone users (2.6 billion) will have downloaded an mHealth app at least once (<https://bit.ly/2qPhTFV>).

traditional provision of clinical care by physicians, and the self-administration of care and wellbeing” (EC, 2012, p. 9).

Both the EC and the WHO have acknowledged the importance of mobile health by providing a framework for its steady implementation, as it had happened with eHealth (EC, 2014; WHO, 2011b). In addition to mHealth solutions such as healthcare telephone helplines, toll-free emergency numbers or appointment reminders, the focus has recently been placed on more widespread use of mobile health in what concerns patient monitoring, clinical decision support systems, and patient and professional education (mLearning)¹⁶. Furthermore, mHealth has been supporting the advent of **uHealth** (ubiquitous healthcare), a current trend in the sector which relies on wireless and sensor-based monitoring technology to collect data 24/7, at any location (Eidam et al., 2017; Elgazzar et al., 2012; Kim et al., 2016; Seppälä, Nykänen, & Ruotsalainen, 2014).

The preceding context has contributed to the emergence of two co-occurring and relevant phenomena: on the one hand, the stakeholders involved in healthcare are now much more diverse and go far beyond the traditional ‘patient + provider + government + pharmaceutical industry (pharma)’ axis, as represented in **Figure 2**; on the other hand, current healthcare has become more and more patient-centered, with patients taking the driver’s seat and playing an increasingly active role in everything related to their health, as will be shown in the following section.

¹⁶ The recent creation of the multilingual “Zika app” by the WHO (<https://bit.ly/1VbCDl8>), aiming to provide essential information on the Zika virus and designed mainly for healthcare workers but also for the general public, illustrates how mobile technologies can support health promotion campaigns, as well as population surveillance and risk prevention/risk communication.

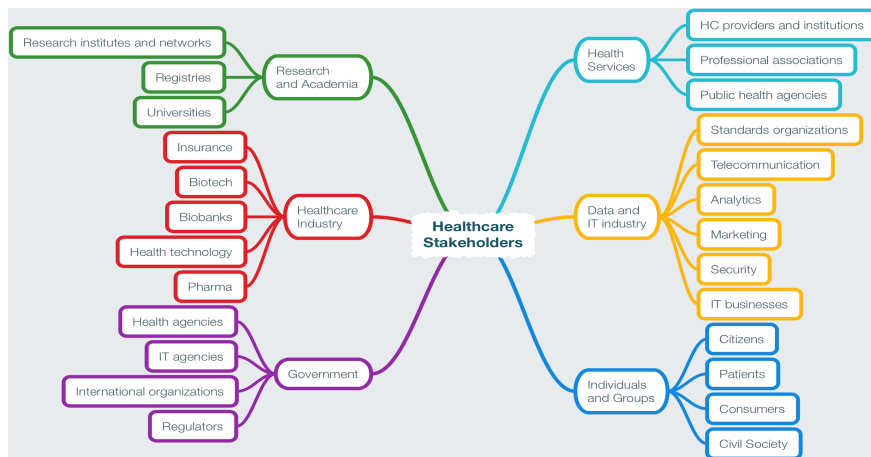


Figure 2 – Current stakeholders within the eHealth framework

(adapted from (Vayena, Dzenowagis, Brownstein, & Sheikh, 2018; WHO & ITU, 2012, p. 3)

1.3. Digital health literacy and the rise of the ePatient

The last few decades have been characterized by the so-called ePatient movement, with the ‘e’ pointing towards ‘empowerment’. Although this term may have different meanings and interpretations, within the scope of healthcare it generally refers to the process that allows an individual or a community to obtain the necessary knowledge and skills to make more informed choices regarding their health. The (WHO, 2009, pp. 191–192) identifies four components in the empowerment process, namely: i) patient participation (the ability to provide relevant background about one’s health history, to show self-interest and motivation for a beneficial outcome, and to be present whenever deemed necessary during care and/or treatment); ii) patient skills (such as self-efficacy, i.e. the belief that one is capable of producing an effect or reaching a certain goal, and health literacy, defined by the WHO as “the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand and use information in ways which promote and maintain good health”¹⁷); iii) patient knowledge (patient education can enhance information exchange and communication between patients and healthcare professionals, thus resulting in a true partnership, based on shared responsibility and shared decisions); and, finally, iv) the presence of a facilitating environment, i.e. one that fosters open communication.

¹⁷ Cf. <https://bit.ly/2oglxbc>

Patient participation has been occurring in several ways, most notably in what concerns chronic disease monitoring and management. Nowadays, it is common practice to have patients themselves collect and, in some cases, share health-related data with healthcare providers, thus avoiding unnecessary medical appointments or procedures. Yet the more involved patients become in their own healthcare, the more they demand to own and manage their own data, which poses a number of challenges to this sector, not only because it questions the traditional relationship between healthcare professionals (namely physicians) and patients, but also because there are several issues concerning data security and privacy that need to be dealt with.

In a more informal way, patients are participating via social media as well, creating vibrant and engaged online communities that connect people worldwide. Patients are using these platforms to learn more about their conditions, to provide or get feedback about a given healthcare institution or professional, to organize and/or participate in health promotion campaigns, to collect and/or submit health data and, furthermore, to raise awareness and put pressure on local, national and international institutions as far as legislation, research and investment are concerned¹⁸.

Among the aforementioned patient skills, health literacy is at the forefront of the most recent European and worldwide projects and other initiatives. The main goal is to provide patients with the necessary conditions to go beyond the basic notion of patient health literacy (which (Nutbeam, 2000) characterized as “functional”), fostering skills that can enable them not only to access/obtain relevant health information, but also to understand that information and, most importantly, to analyze it critically, thus supporting the patients’ more engaged role in their own healthcare issues (cf. (Sørensen et al., 2012) and Nutbeam’s top-level notion of “critical health literacy”). With the proliferation of electronic devices, the notion of health literacy eventually started comprising that digital component. According to (Norman & Skinner, 2006, p. 2), eHealth

¹⁸ One of the best-known examples of such a network is *PatientsLikeMe*, where people can submit their health data which, in turn, is organised and shared with the different stakeholders, thus allowing the development of more and better healthcare solutions (<https://www.patientslikeme.com/>). Other platforms include *RareConnect*, which aims to globally connect patients suffering from rare diseases (<https://www.rareconnect.org/en>), or more national-based networks, such as the *Patients en réseau* (<http://www.patientsenreseau.fr>), that defines itself as an association for the patients and their respective caregivers.

literacy implies that one must have the ability to “seek, find, understand and appraise health information from electronic sources, as well as use that knowledge to address or solve a health-related problem”.

Finally, patient knowledge, more specifically patient education, is one of the cornerstones of eHealth implementation programs, as it was previously mentioned (cf. point 4 in Table 1). Through such programs, patients (and often caregivers) are provided with formal and tailored education, concerning either general public health content or training on specific pathologies. This has taken one step further in recent years, with the advent of another type of ePatient, the so-called ‘expert patient’¹⁹. The notion refers to a patient with a chronic disease whose acquired knowledge and experience about that condition enable him/her to play an active role in its management. Nevertheless, that engagement is not limited to his/her situation, since an expert patient “peut aussi intervenir en tant que personne ressource pour les autres”²⁰, not only other patients, but also healthcare professionals and the remaining stakeholders.

The first initiative to provide formal education to patients, mainly with the purpose of supporting them in their own healthcare management, was developed at Stanford University, in the early 1990s. Nowadays, the Stanford Patient Education Research Center (SPERC) and its Chronic Disease Self-Management Program (CDSMP) are highly sought after. According to the Director of the SPERC (Lorig, 2014), about 100,000 people in 36 countries attended the CDSPM in 2013 alone. There are currently other similar self-management programs available at Stanford, namely for Diabetes, Chronic Pain, Cancer Survivors and HIV.

In the United Kingdom (UK), the National Health Service (NHS) also became aware, very early on, of the potential involving patient education, not to mention the possible cost reductions this could imply. As a result, the NHS/Department of Health launched the Expert Patient Programme (EPP) in 2002, which has now disseminated all across the UK. It is a free, self-management course aimed at people living with, or caring for someone with, one or more long-term health conditions (e.g. diabetes, high blood

¹⁹ Although equivalents in other languages are available in the specialised literature, namely in French (patient expert), Spanish (paciente experto) and German (Patientenexperte), nothing could be found in Portuguese, thus the term “paciente especialista” is proposed within the scope of this work.

²⁰ Cf. <http://projetsante.com/le-patient-expert/>.

pressure, depression, anxiety, asthma, arthritis and chronic obstructive pulmonary disease (COPD))²¹.

Another initiative at European level includes the European Patient's Academy (EUPATI), a multi-stakeholder consortium (involving the pharmaceutical industry, academia, not-for-profit and patient organizations) which focuses on patient education and training in the areas of medicine development and regulations, clinical trials, and health technologies assessment. It provides a free Patient Expert Training Course over a 14-month period and so far, EUPATI has trained almost 100 Patient Experts, as well as offered and maintained course materials on Medicine Development²².

The latest innovation in what concerns formal patient education is a concept originally launched in France: a university targeted mainly at patients, their relatives and caregivers. Founded in 2009 by Catherine Tourette-Turgis, the *Université des Patients*²³ is located at the Pierre and Marie Curie University (Paris 6) and provides expert patients with formal higher education training programs (in areas such as therapeutic education, chronic diseases, healthcare legislation, etc.), which can subsequently lead to a Higher Education qualification as, for instance, the *Diplôme Universitaire* or DU. The model above has already reached other European countries, namely Spain, with *La Universidad de Los Pacientes*²⁴, resulting from a partnership between the Autonomous University of Barcelona (UAB) and the Health and Environment Foundation of the UAB.

That being said, one could wonder where the notion of 'expert patient' would fit into all of this. What degree of expertise would such a person have if he/she had received formal training and possibly obtained a qualification? Would he/she be at the expert level? That could be problematic if there were no sublevels, as it could eventually place the expert patient at the same level as a physician with a given specialization. Could the expert patient, with the appropriate training, be at the same level regarding knowledge about a disease's etiology, symptoms, treatment options, etc. than a physical

²¹ A case study concerning the EPP can be consulted at <https://bit.ly/2aJHmrl>.

²² More information available at <https://www.eupati.eu/>.

²³ Cf. <http://www.universitedespatients.org/>.

²⁴ Cf. <http://www.universidadpacientes.org/>.

therapist, a nurse, or even a General Practitioner (GP)? Again, this could be difficult to ascertain.

Even if the answers to these questions are still far from being unequivocal, this emerging trend raises pertinent questions as regards the notion of 'expert' in Terminology and terminology work. In fact, it is believed that a more thorough discussion around this notion will be essential in the very near future, given that terminology work relies heavily - and rightly so - on the collaboration between terminologists and subject field experts. The notion of 'expert' has remained relatively stable since the early days of Terminology as a scientific area in its own right. According to (Picht & Draskau, 1985), for instance, an expert deals with specialized communication at the highest level of complexity, whereas for (Bowker & Pearson, 2002), one has to work and receive training in a given subject field to become a subject field expert. In addition, it is possible to distinguish different levels of specialization regarding the producers and the recipients of specialized knowledge, depending on their degree of expertise (e.g. non-experts, semi-experts and experts) (Bergenholtz & Kaufmann, 1997, pp. 100–101).

Taking the preceding notion of expert patient into account, the aforementioned levels of specialization may need to be revisited in Terminology: firstly, would a new category be necessary, to update the traditional non-expert/semi-expert/expert triple? As mentioned earlier, it is still not totally clear where the expert patient could be placed in what concerns the knowledge of a very specific disease; secondly, some initiatives have been developed where patients are actually authors - or co-authors - of scientific publications in peer-reviewed medical journals, such as the BMJ, with its "Patient Partnership Strategy"²⁵. Accordingly, one could wonder if these articles could be integrated as such into an 'expert-to-expert' corpus, for instance. What is more, when consulting a terminological resource, this patient will also have differentiated pre-acquired knowledge - and needs - when compared to other patients or the general public. Due to all of this, it is believed that this notion of the 'expert patient' will be the

²⁵ More information at: <https://www.bmj.com/campaign/patient-partnership>.

subject of analysis and discussion within Terminology in a very short-term, especially in terminological projects that focus on the healthcare domain.

Thus, within the scope of this work, and to further specify the end users of the resource being proposed in this research work, a distinction is made between ‘expert healthcare providers’ – referring to the physicians and other HC providers that have specialized in endometriosis via formal theoretical and/or practical training – and ‘expert patients’, i.e. patients that have received some type of training (i.e. via lectures or other theoretical sessions) concerning this particular pathology.

In short, this section has focused on the increasingly active role of patients in their own health and on the fact that digital health literacy is on its way to becoming one of the key skills of the 21st century. However, that active patient participation on so many different levels - also in what concerns the gathering of healthcare data - raises important issues concerning data protection and data privacy. With the recent Facebook scandal, these questions have become even more pressing: ultimately, who owns healthcare data, who has access to that data, and what can be done with it? That will be one of the main topics of the following section.

1.4. Interoperability in healthcare: caught between hype, hope and reality

The overall background provided thus far substantiates eHealth’s (and consequently mHealth’s) significant impact towards a more cost-effective healthcare access and delivery. However, it also points towards key areas that may either foster or prevent eHealth from reaching its full potential: i) Big Data; ii) legislation; and iii) interoperability.

Big Data (i.e. collecting and storing massive amounts of data electronically) is a fairly recent phenomenon in the healthcare sector, which has experienced a fast and substantial growth in the last decade (WHO, 2016a). By fostering the reuse and connection of existing datasets, it aims to respond to the ‘silo approach’ that has characterized healthcare database creation and maintenance (Bresnick, 2015; Zillner & Neururer, 2016). Despite the substantial benefits resulting from applying Big Data to healthcare, namely the ability to boost healthcare research and innovation, thus leading,

for instance, to individually targeted drug therapies, ethical issues with regard to data protection²⁶, privacy and security still constitute a major concern among local and international authorities. There has been, in fact, a pressing need to devise a clear legal framework on data privacy, security and confidentiality in this sector. Questions pertaining to data access and sharing rights, data quality and integrity, its ownership, as well as to liability, in case errors or damage occur, have become vital in eHealth policies (cf. (EC, 2012; WHO; ITU, 2012; WHO, 2016a).

Lastly, there is the issue of interoperability, which has become one of the key drivers of eHealth deployment. As previously mentioned, it is one of the ‘hot topics’ in healthcare, insofar as a successful implementation of interoperable solutions can, indeed, enhance the quality and outcomes in the sector, while decreasing costs. In short, it can be the best of all possible worlds. Yet, interoperability has also been one of the most challenging topics in this domain, due to the underlying complexity of delivering “the right information, at the right time, to the right place” (Benson & Grieve, 2016, p. 20). Adapting the title of a book by the physician Robert Wachter²⁷, one could say that there has been a considerable hype in healthcare about interoperability, as it represents the ultimate hope of solving all the information exchange and communication problems in this sector. However, reality has shown less optimistic, or at least less conclusive, results. Before focusing on what may be causing those hurdles, it is important to start by understanding what interoperability is.

In 1990, the Institute of Electrical and Electronics Engineers (IEEE) defined interoperability as “the ability of two or more systems or components to exchange information and to use the information that has been exchanged” (IEEE, 1990, p. 42). Other authors/organisations have then expanded this definition and adapted it to healthcare. One of most widely cited is that by the Healthcare Information and Management Systems Society (HIMSS), which refers to interoperability as “the ability of different information technology systems and software applications to communicate, exchange data, and use the information that has been exchanged”²⁸, in such a way that

²⁶ Personal data protection is an integral part of any citizen’s basic rights in Europe (cf. Art. 8 of the EU’s Charter of Fundamental Rights, 2012).

²⁷ The Digital Doctor: Hope, Hype, and Harm at the Dawn of Medicine’s Computer Age (2015).

²⁸ Cf. <http://www.himss.org/library/interoperability-standards/what-is>

the data can then be shared across different stakeholders, regardless of the application or application vendor.

Due to the key role of interoperability in eHealth, several initiatives have been carried out on a global scale in order to create and/or improve supporting legislation, tools and services. In 2013, the World Health Assembly approved a resolution endorsing eHealth standardisation and interoperability among all its Member States (WHA, 2013). Moreover, the European Commission's eHealth Network has been working towards an eHealth European Interoperability Framework (eEIF), whose goal is to effectively achieve a cross-border application. An updated version, described as a 'refined model' [ReEIF], has been published in 2015²⁹.

Overall, the concept of interoperability comprises different layers. However, it can be difficult to find agreement in the literature about this, both in what concerns the number of levels themselves and, mostly, the designations attributed to each of them.

Table 2 provides an overview of different approaches:

Authors	Layers of Interoperability (IO)
European Commission (2009)	• Technical • Syntactic • Semantic
(Elkin & Tuttle, 2012)	• Syntactic • Semantic • Pragmatic
HIMSS (2013)	• Foundational • Structural • Semantic
Benson & Grieve (2016)	• Technical • Semantic • Process (includes Clinical IO as a subtype of Process IO)

Table 2 – Types of interoperability layers (overview)

For the purpose of this work, the EC's terminology will be used, but also taking into consideration the features of each level as described by the HIMSS (2013) and by Benson & Grieve (2016). This is summarized in **Figure 3**:

²⁹ The model, the main outcome of the Antelope Project, is available at <https://bit.ly/2qUOs4q>.

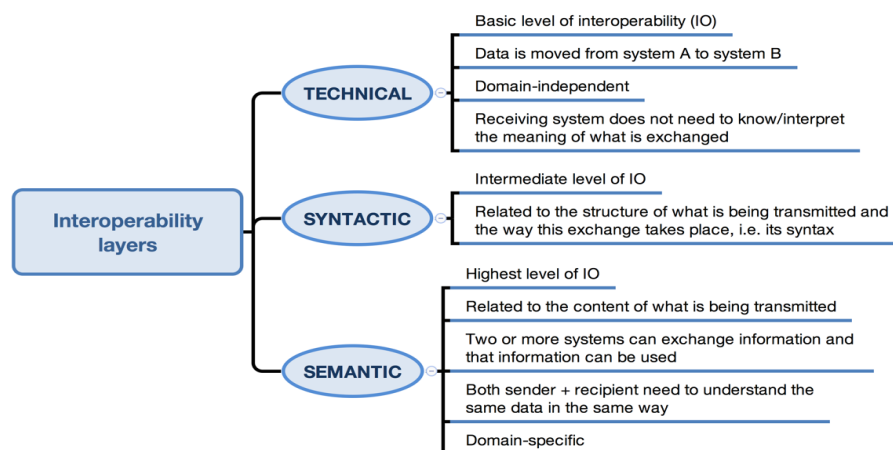


Figure 3 – Types of interoperability layers: terminology used within the scope of this work

Semantic interoperability is, in short, the highest possible level, contributing decisively to the fundamental goal within healthcare informatics, i.e. that clinical information systems are able to exchange information accurately, supported by a shared understanding, not only of the structure of what is being transmitted (**syntax**) but also of the meaning of that message (**semantics**) (Coiera, 2015). It is in this particular layer that Terminology and terminology work can play a key role, as terminologies, ontologies, classifications, thesauri and other terminological systems³⁰ provide the means to transmit content, thus allowing semantic interoperability to be achieved.

The question could be posed, however, of why this goal has been so difficult to reach, despite having acquired a priority status in key international organizations, as previously mentioned, and having been the main focus of multiple projects worldwide. The use case for semantic interoperability has, indeed, been hampered by several barriers. Some of the examples include the heterogeneity of the systems supporting HC information; the privacy and security concerns surrounding EHRs or even PHRs; the political issues regarding funding allocation and priority setting; the fact that different professional communities use their own jargon (terminology may change even from one medical specialty to another); and, last but not least, the inherent complexity of representing healthcare information and, subsequently, healthcare knowledge (Benson & Grieve, 2016; Sicilia & Balazote, 2013). The aforementioned ‘silo approach’ followed

³⁰ An overview of terminological systems in healthcare will be provided in **Chapter 3**.

by many companies and healthcare institutions can also be seen as a hurdle, as well as the lack of global standards to handle the current 'health data deluge'.

The need for a standard interchange language, i.e. a common format, that can be used across the various systems and support semantic interoperability is, therefore, of paramount importance. Yet, this need is far from being confined to healthcare, and it has become one of the cornerstones of the Linked Data (LD) approach. Given LD's potential in healthcare, the next section will provide a brief overview of LD and its grounding principles, followed by some examples of LD-based healthcare projects and resources.

1.5. Healthcare at a crossroads: goodbye data silos, hello Linked Data?

Linked Data is another buzzword in the healthcare sector: as it was mentioned before, the increasing need to connect clinical information from different sources across systems and institutions led to a search for solutions. Thus, LD emerged as an ambitious, yet feasible option to the existing data silos. Coined by Tim Berners-Lee (Berners-Lee, 2006), the term Linked Data refers to "a set of best practices for publishing and interlinking structured data on the Web" (Heath & Bizer, 2011, p. 7) resorting to international standards and recommendations issued by the World Wide Web Consortium (W3C).

It is one of the pillars of the **Semantic Web** (SW), as it provides the means to put the SW architecture into practice. In their seminal paper from 2001, Berners-Lee, Hendler & Lassila perceive the SW as an extension of the current one, in which machines can communicate more effectively because, in this context, "information is given well-defined meaning" (Berners-Lee, Hendler, & Lassila, 2001). There has been, therefore, a massive shift from a Web of documents, targeted mainly at humans and relying on Web standards such as the Uniform Resource Locator (URL) (for identification), the Hypertext Transfer Protocol (HTTP) (for access/retrieval) and the Hypertext Markup Language (HTML) (for content format), to a Web of Data, aimed essentially at machines and which builds directly on Web technologies, reusing HTTP, expanding the concept of URL into a

URI (Uniform Resource Identifier)³¹, and making use of a common data model – the Resource Description Framework (RDF) –, hence allowing data to be structured (cf. Berners-Lee, 2009; Bizer, Heath, & Berners-Lee, 2009; Wood, Zaidman, Ruth, & Hausenblas, 2014). More importantly, via this architecture, anything, from real-world objects to abstract concepts, can now be identified and linked. This forms the basis of the SW technology stack, depicted in **Figure 4**, and is also at the heart of the so-called Linked Data principles, specified by Berners-Lee in 2006:

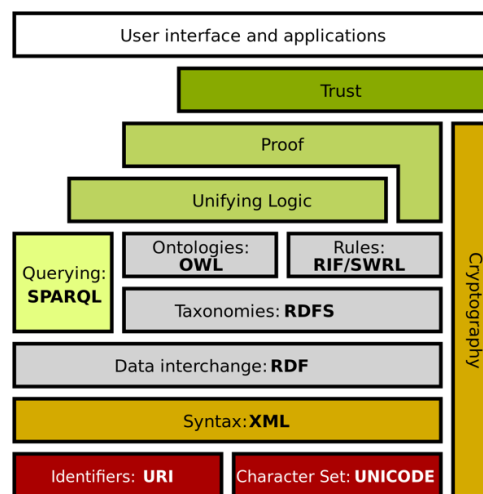


Figure 4 – The Semantic Web technology stack³²

[1] URIs should be used as names for things, i.e. for any entity, either concrete or abstract, so that anyone can link to them, refer to them or retrieve a representation of them; [2] HTTP URIs should be used, so that people can look up those names – this enables these URIs to be dereferenced via the HTTP protocol, enabling a description of the identified resource to be retrieved; [3] When a URI is looked up, useful information about it should be provided, using standards such as RDF, SPARQL (a query language for RDF data), among others. This principle advocates the use of RDF as the single data model for publishing structured data on the Web; [4] Links to other URIs should be included, so as to allow further connection between resources, i.e. using hyperlinks to

³¹ A URI is a generalised version of a URL, i.e. a URL is a type of URI. Gradually, URIs are being replaced by IRIs (International Resource Identifiers), i.e. global identifiers that allow non-ASCII characters to be used in the character string and, thus, the use of different languages in the identifiers themselves (cf. <https://www.w3.org/TR/ld-glossary/>; <https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140225/>). To further distinguish the three aforementioned notions (URL, URI, IRI), Gandon (2016) states that a URL identifies what exists on the Web; a URI makes it possible to identify, on the Web, what exists; whereas an IRI enables the identification, on the Web and in any language, of what exists (cf. <https://bit.ly/2HIUW7C>).

³² Source: <https://bit.ly/2gQHr5t>

connect not only Web documents, but any type of entity identified by a URI (Shadbolt, Hall, & Berners-Lee, 2006). In short, data only becomes Linked Data when “it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can, in turn, be linked to from external data sets” (Bizer et al., 2009, p. 2). One of the key drivers of LD that allows the link to external data sets consists of releasing the data under open licenses, thus leading to the notion of **Linked Open Data** (LOD).

LOD relies on a 5-star scoring system created by Berners-Lee (2006), according to which the data to be published needs to cumulatively meet a number of criteria in order to be granted 5 stars: [1 star] – the data is available on the Web, in whatever format, under an open license (e.g. to publish an image scan of a table on a website); [2 stars] – the data must be available as machine-readable structured data (e.g. to publish the table in MS Excel instead of the image scan); [3 stars] – all of the previous criteria plus the use of a non-proprietary format (e.g. to provide the table in CSV rather than MS Excel); [4 stars] – all of the above plus the use of open standards from W3C, such as RDF and SPARQL; [5 stars] – all of the above plus linking one’s data to other people’s data in order to provide additional context.

Thus far, several LOD data sets have been published and linked to each other in the so-called LOD cloud, which has been growing exponentially in recent years. Its version from February 2017 includes about 1,139 data sets and can be visualized in **Figure 5**. The number of data sets concerning the Life Sciences domain [in red] is particularly significant, as can be seen from the diagram.

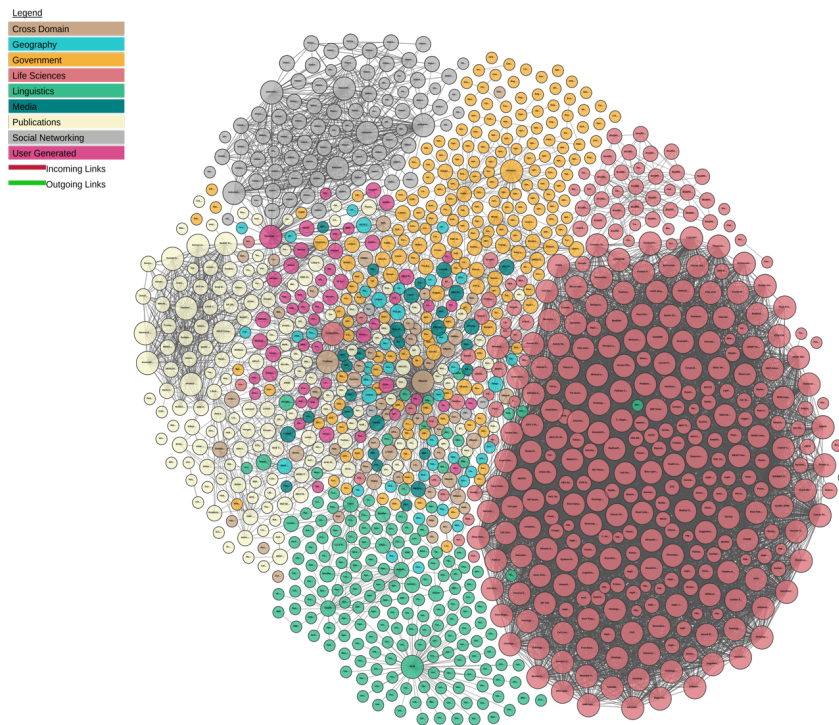


Figure 5 – The Linked Open Data cloud³³

Despite the importance of the remaining elements pertaining to the SW technology stack, including the RIF/SWRL rule languages, as well as ontologies, taxonomies and other vocabularies, the key role of RDF is indisputable as regards the implementation of SW applications. Accordingly, the following paragraphs will provide a more thorough description of this pivotal data format.

As one of the building blocks of the SW, RDF is essentially a framework for representing information on the Web³⁴. More specifically, it allows resources (e.g. documents, people, other physical objects, as well as abstract concepts) to be identified (via HTTP URIs), described (via properties and their respective values) and shared (cf. Linked Data Principles) in a consistent, machine-readable manner. It is a family of international standards produced by the W3C that has reached a certain level of maturity (it was firstly adopted in 1999) and is used in various domains, including healthcare.

There are currently several RDF syntaxes available, also called serialization formats: a) RDF/XML: the original RDF format in eXtensible Markup Language (XML); b)

³³ Latest consulted version: 2017-02-20 – available at <http://lod-cloud.net/>.

³⁴ Cf. RDF 1.1 Concepts and Abstract Syntax (<https://www.w3.org/TR/rdf11-concepts/>).

the Turtle family of RDF languages (e.g. Turtle, N-Triples, etc.): a simpler, human-readable format; c) RDFa: here, RDF is embedded in HTML attributes; d) JSON-LD: a more recent format based on the JavaScript Object Notation (JSON) language and adapted to LD, aimed primarily at Web developers. Underlying the description of any resource is RDF's core structure, which consists of the so-called triples: a) a subject (i.e. the URI identifying the described resource); b) a predicate, indicating the type of relation existing between the first and the last component (a URI also identifies the predicate); and c) an object, that can either be another resource (with the corresponding URI) or a literal, that is, an object with a literal value (e.g. a number, a date, a person's name, etc.). An object can also be a blank node, although these should be avoided whenever possible. Each RDF triple constitutes a statement, and, in turn, these statements can be represented as a graph, i.e. with the subject and object being the nodes, and the predicate connecting both.

The mapping potential provided by RDF should not be underestimated, particularly since the information to be mapped can also be linguistic. In a nutshell, within the LOD initiative one can also find the Linguistic Linked Open Data (LLOD) movement, which aims to apply the LD principles to lexical and other linguistic resources, mainly via RDF, as previously stated (Chiarcos, Nordhoff, & Hellmann, 2012). In this respect, the W3C Ontology-Lexicon Community Group (Ontolex) has recently proposed the Lexicon Model for Ontologies (LEMON), an RDF-based model intended to support the publication of lexical resources as Linked Data (McCrae et al., 2017; W3C Ontology Lexicon Community Group, 2016).

Ultimately, this first chapter aimed to explore some of the key concepts in healthcare at the onset of the 21st century, as well as some of the hurdles regarding the dissemination of clinical data across multiple providers, departments, institutions, and countries. Disruptive phenomena like the ones alluded to here do entail challenges, barriers, and limitations, but they also provide massive opportunities and benefits. It is believed that Terminology, especially in its knowledge organization facet, can play an influential role by helping to enhance not only communication between the different stakeholders, but also the way clinical data is represented, stored, and shared.

Part II: Theoretical framework

Chapter 2 – Terminology: where the conceptual and the linguistic intertwine

‘When I use a word,’ Humpty Dumpty said in rather a scornful tone, ‘it means just what I choose it to mean – neither more nor less.’

‘The question is,’ said Alice, ‘whether you *can* make words mean so many different things.’

‘The question is,’ said Humpty Dumpty, ‘which is to be master – that’s all.’

Lewis Carroll, *Through the Looking Glass* [Chapter 6]

2.1. Introduction

After nearly two decades into the 21st century, Terminology as a scientific discipline seems to be at a crossroads: to either maintain its autonomy while fully acknowledging its multidisciplinary scope, or to eventually dilute into areas such as specialized lexicography and translation, or more technologically-oriented – and increasingly popular – approaches that resort to text mining, deep learning, or Big Data to extract, represent, organize, and share specialized knowledge.

When 50 years have passed since Wüster’s publication of the seminal “Machine Tool” dictionary (Wüster, 1968), Terminology is still regarded as a scientific area in the confluence of many others (e.g. linguistics, logic, information science, ontology, lexicography, and computer science, to name just a few) (Cabr , 1999; ISO, 2009; J. Sager, 1990; W ster, 1974), and one where its boundaries towards those related disciplines did not seem to get less blurry over time. Despite the challenges and the debate regarding Terminology’s core identity, it is believed that its interdisciplinary nature constitutes, in fact, one of its greatest assets, providing the study of specialized knowledge and specialized language with a “plurality of theoretical approaches” (Costa, 2006) that, especially since the onset of the new millennium, have been aiming at a more consistent interrelation between what is linguistic and what is conceptual. In short, Terminology as an autonomous scientific discipline should be perceived as much more than merely the sum of the different ‘parts’ that comprise it (Cabr , 1999, p. 25), and the challenge for the upcoming years may lie on defining the exact nature of the aforementioned boundaries, as well as the driving forces that support them (Rogers & Wright, 2006).

Within the biomedical domain, this interdisciplinarity allows fruitful collaborations to be established in multidisciplinary teams of terminologists, subject field experts – bearing in mind that this notion of expert may be evolving, as was shown in Chapter 1 with the ‘expert patients’ –, medical coders and/or librarians, together with Information Technologies (IT) and KE professionals. In fact, the advent of the Semantic Web (SW) and, more recently, of the Linked Data (LD) initiative, has paved the way for new perspectives and opportunities for Terminology, namely in what concerns the operationalization of terminological products (i.e. the development of a computational representation of their concept system), namely via ontologies (in the sense of KE).

Furthermore, such collaborative work also demands, from the terminologist, the ability to adapt to and learn about related areas. In this particular project, it was necessary to acquire a deep understanding of key concepts in the present healthcare setting (e.g. interoperability, LD, or the SW, for instance), as well as of current biomedical terminological resources.

Hence, the remainder of this chapter is dedicated to outlining the theoretical framework of this thesis, which rests upon three main postulates:

- [1] the added value concerning the interdisciplinary nature of Terminology and its status as an autonomous scientific area, a topic which was explored in the preceding paragraphs;
- [2] Terminology’s double dimensional approach, which focuses on the synergies between what is conceptual and what pertains to the linguistic, while also addressing the impact of such a perspective within terminological work;
- [3] the increasingly fruitful collaboration between Terminology and ontologies, the latter regarded in the sense of KE, and how that can contribute to the aforementioned operationalization of terminological products.

2.2. Terminology's double dimension: two sides of the same coin

As mentioned earlier, at the heart of the work carried out in this research project is the assumption that Terminology has a double dimension, linguistic and conceptual, which has been more thoroughly described by (Costa, 2013; Roche, 2012, 2015; Santos & Costa, 2015). According to this approach, terminology work needs to consider not only the discourses produced by experts but also the formal (or semi-formal) representations of the shared knowledge regarding their respective domains. In short, both linguistic and conceptual information become indispensable and need to be gathered, analyzed, and systematized by the terminologist, supported by a theoretical and methodological framework that allows the potential within each dimension and the synergies resulting from their interaction to be maximized.

In order for that to happen, it is necessary to perceive the notions of **concept** and **term** as autonomous entities that are, nonetheless, interrelated in terminology work (cf. (Silva, 2014)): the former allows us to conceptualize the world, whereas the latter allows us to talk about that conceptualization and share it with other human beings. In short, terminology is regarded as both “a science of objects and a science of terms” (Roche, 2015, p. 136). These two core notions have been defined quite differently by the various theoretical approaches in Terminology (Bourigault & Slodzian, 1999; Cabré Castellví, 2003; Cabré, 1999; Faber Benítez, 2009; Faber, 2012, 2015; Felber, 1984; Gaudin, 2005, 2007; Picht & Draskau, 1985; Temmerman, 2000; Wüster, 1991), yet for this thesis, the working definitions for concept and term are the ones put forward by the ISO 1087-1 (2000) standard, namely: a “unit of knowledge created by a unique combination of characteristics” and a “verbal designation of a general concept in a specific subject field”, respectively (pp. 2; 6).

In short, term and concept play a key role within an expert community and, thus, in terminology work, but the fundamental issue is, in fact, the **relation** between them. For Costa (2013), the specificity of Terminology as an autonomous scientific subject lies precisely in studying how both the conceptual and the linguistic interrelate and become complementary, as two sides of the same coin. Indeed, when one chooses to analyze the concept without taking the term into account, the scientific area underlying the analysis is no longer Terminology, but Knowledge Engineering. In turn, to focus on the

term while disregarding the concept is, again, not Terminology, but Specialized Lexicography. Therefore, acknowledging this double dimension in Terminology has clear theoretical – but also methodological – implications, as will be shown in Part III.

On the one hand, carrying out a systematic concept analysis at the very onset of a terminology project has had a long tradition in terminology work, with the advantages of such an approach being highlighted by several authors (Felber, 1984; Felber & Budin, 1989; Picht & Draskau, 1985; J. Sager, 1990; Suonuuti, 2001; Wüster, 1991). For (Meyer, Eck, & Skuce, 1997), this prior concept analysis can make a valuable contribution as regards:

[1] the preparation of lists of candidate terms, as the terminologist may be able to detect potentially “unsuitable terms” in a more thorough way, as well as identify areas in the concept systems where terms may still be missing. However, and as it will be shown later, the opposite situation may also occur, i.e. terms may point towards conceptual information that is not present in the prior conceptualization, thereby contributing to enhance it³⁵;

[2] the issues of terminological synonymy (Costa, 2017), as well as of interlingual equivalence. By previously identifying the characteristics of a given concept, the terminologist will be able to more accurately verify whether two or more terms designate that concept and gather their respective equivalents in different languages;

[3] the interaction with subject field experts, since doing one’s “conceptual homework” (Meyer *et al.*, 1997, p. 108) will undoubtedly facilitate the indispensable communication with the experts throughout a project;

[4] the drafting of natural language definitions. As will be explored in one of the final sections of this chapter, conceptual representations provide a valuable starting point regarding the creation of definitions and allow the drafting process to expand on the more ‘traditional’ intensional definition (generic-specific).

³⁵ It is believed that concepts do not exist in texts *per se*, but terms (as well as lexical-semantic relations) may supplement the terminologist’s linguistic knowledge and assist in the elicitation of some conceptual traits, thereby assisting the terminologist during corpus analysis, provided that he/she has previously acquired the necessary domain knowledge to interpret those ‘hints’.

On the other hand, specialized texts are, indeed, the natural habitat of terms. In most subject fields and communities of practice, texts are still the main vehicle to represent, structure, and disseminate knowledge (cf. (Costa & Silva, 2008; Costa, 2013)), in spite of the ambiguity and inconsistencies that characterize natural language. Specialized texts also play a fundamental role in fields where knowledge stabilization and knowledge dissemination are paramount, as is the case of endometriosis. All in all, they constitute an invaluable resource to the terminologist, especially in their first contact with a given domain, and the experts can – and should – advise the terminologist as to which texts are deemed representative and/or mandatory in a given area. That was the case in this particular project, as will be shown in Chapters 5 and 6.

Nowadays, however, and taking the previous considerations into account, it is becoming increasingly important that, depending on the objectives and target audience, terminological products should be represented in a computer-processable way, thus paving the way to the interoperability referred to in Chapter 1.

2.3. Terminology and ontologies

As mentioned earlier, the collaborative work involving Terminology and ontologies has been regarded as quite promising (Roche, 2015) and has led to the development, in recent years, of numerous resources focusing on several domains, one of them being Medicine. But what is an ontology? According to the Oxford English Dictionary, it is “[1] the branch of metaphysics dealing with the nature of being; and [2] the set of concepts and categories in a subject area or domain that shows their properties and the relations between them³⁶”.

From this example, it is possible to see that the notion of ontology can have two distinct meanings: when written in upper case (Ontology), it refers indeed to a branch of philosophy that focuses on the study of being, of what is, or of *Being qua Being*, as stated by Aristotle in his Book IV – Metaphysics (Aristotle, 1989). Named “first philosophy” by Aristotle himself, Ontology (or *ontologia*, in Latin) was firstly coined as a term in the 17th century, with the main objective of understanding the nature of things,

³⁶ <https://en.oxforddictionaries.com/definition/ontology> (16.03.2018).

i.e. their essence³⁷. Based on that study of being, Aristotle proposed a categorization of the world that surrounded him, and that proposal was then visually represented via the so-called Porphyry Tree (cf. next section), developed by the philosopher Porphyry in the introduction (*Isagoge*) to Aristotle's treaty on Categories (Smith, 2018). This Tree is considered one of the most influential hierarchical classifications to this day and is often seen as the first "prototype" for current ontologies. In the 17th and 18th centuries, however, philosophy became mostly interested in the nature of knowledge itself (cf. the works by Descartes and Kant) rather than in the nature of things, and Ontology was eventually overshadowed by epistemology.

Since the end of the 20th century, particularly the 1980s, ontologies (this time in lower case and in the plural (cf. Guarino & Giaretta, 1995)) have taken on a new role and importance in the fields of Computer Science, Artificial Intelligence (AI), and KE. According to this perspective, ontologies are computational artifacts used to "formally model the structure of a system" (Guarino, Oberle, & Staab, 2009), namely via a conceptualization (the representation of the relevant types of entities in a given domain, together with the relations between them). Perhaps the most widely quoted definition of ontologies in this respect is the one by (Gruber, 1993; Gruber, 1993b), stating that "an ontology is an explicit specification of a conceptualization". A few years later, (Borst, 1997) expanded upon Gruber's proposal, introducing the notions of "formal specification" and of "shared conceptualization". One year later, (Studer, Benjamins, & Fensel, 1998) merged the preceding definitions, recommending that ontologies would be defined as "a formal, explicit specification of a shared conceptualization".

This latter proposition is the one being used within the scope of this thesis, as it introduces three critical features: firstly, the fact that this specification should be **explicit**, i.e. the type of concepts and the respective constraints on their use must be explicitly defined; secondly, that it should be **formal**, i.e. understandable by a machine; and thirdly, that it should be **shared**, i.e. it should capture knowledge that is consensual among a given community.

³⁷ This will be pursued further in this chapter, due to its relevance in the creation of definitions.

Regarding the different types of ontologies, several classification proposals have been put forward throughout the years, as the one by (Guarino, 1998), which divides the ontologies, according to their structure, into: [1] Top-Level or Upper ontologies, that represent very general concepts; [2] Domain ontologies, that depict the fundamental concepts of a given domain; [3] Task ontologies, which focus on the key concepts of a general activity or task; and [4] Application ontologies, which are specialized ontologies designed for a specific use or application focus.

Due to their operationalization potential, ontologies have become a ‘hot topic’ in several areas over the last few decades, as referred to earlier, and Medicine has been one of them. In fact, the search for language-independent³⁸ biomedical TS started in the mid-1990s, as a key objective of the groundbreaking GALEN project, which will be further explored in the following chapter. Also during the 1990s, Biology ‘discovered’ the potential of ontologies as regards knowledge representation, organization, and sharing, which led to the creation, in 1998, of the Gene Ontology – now GO Consortium –, whose mission has been to develop “an up-to-date, comprehensive, computational model of biological systems, from the molecular level to larger pathways, cellular and organism-level systems” (Gene Ontology, 2018). At the beginning of the new millennium, another major project began, namely the conception of the Basic Formal Ontology (BFO), an upper-level ontology designed to support “information retrieval, analysis and integration in scientific and other domains” (BFO, 2017). With the advent of the 21st century, and as will be seen in the next chapter, several biomedical resources have managed to integrate ontology-related structures and features, as well as convert their content into SW- and LD-compliant formats.

In recent decades, ontologies have also become a relevant topic in Terminology research and have been used as the foundation of several terminological resources. In this respect, reference should be made to the pioneering work by Meyer and colleagues in the 1990s, the first who aimed to incorporate aspects of KE into terminology work (Meyer et al., 1992; 1997) and, thus, explore the potential of this interrelation between what is linguistic and what is conceptual. Although the notion of ‘ontology’ was not used

³⁸ For the purposes of this work, language-independent is understood here as referring to a conceptualization that is not determined by or restricted to a given natural language.

by these authors, their COGNITERM project, for instance, was described as a Terminological Knowledge Base (TKB), making “explicit what a native speaker who is also a domain expert knows about concepts denoted by specialized lexical items” (Meyer et al., 1992, p. 956). This resource comprised a set of features that were innovative at the time, some of which can still be found in most of the current terminological resources, including EndoTerm: the combination of hierarchical and non-hierarchical conceptual relations, in order to make the content of the TKB more explicit; a graphic visualization tool, allowing the user to navigate the resource in a more interactive way; an embedded browser which enabled the user to search for content either via concepts or via characteristics; and, finally, the so-called Characteristic Comparison Matrix (CCM), a type of matrix where the characteristics of the several concepts were registered and compared, and which supported the drafting of intensional definitions.

This search for a more systematic integration of Knowledge Representation and KE tools and approaches into terminological resources regained momentum at the onset of the new millennium. An example of that evolution within an existing resource refers to the EcoLexicon, a multilingual TKB dedicated to the environmental domain and developed according to the theoretical premises of Frame-based Terminology (Faber, 2009; 2012; 2015). Since the existing relational database was found to be insufficient, EcoLexicon is now supported by a domain ontology in OWL (Web Ontology Language), which allows for the effective representation of domain entities and for inferences to be drawn from them (Faber, León-Araúz, & Reimerink, 2014). In the near future, it is expected that in addition to the corpus annotation via a part-of-speech (POS) tagger, EcoLexicon content can be included in the LOD initiative (Faber, León-Araúz, & Reimerink, 2016).

The need for a more substantiated presence of ontologies in Terminology and terminology work has also led to the advent of new theoretical and methodological perspectives, whose designations seek to illustrate that correlation. One of these approaches is Termontography, developed by the research group of the Centrum voor Vaktaal en Communicatie (CVC) in Brussels within the scope of the European project FF Poirot in the field of financial forensics (Temmerman & Kerremans, 2005). This multidisciplinary perspective combines Ontology Engineering, Terminography, and

Corpus Linguistics, namely by resorting to theories and methods for multilingual terminological analysis stemming from the Sociocognitive terminology theory proposed by (Temmerman, 2000), together with theories underlying ontology engineering (Temmerman & Kerremans, 2003).

From a methodological standpoint, Termonography relies on both top-down and bottom-up approaches: firstly, a “framework of categories and intercategory relationships” (Kerremans & Temmerman, 2004, pp. 4–5) is devised, in collaboration with subject field experts; secondly, this categorization is used as a template for the “manual and semi-automatic extraction of knowledge from a corpus” (ibid.); next, both the corpus data and the categorical framework are compared, and the latter may be expanded as a result of the corpus analysis. Hence, a bottom-up analysis is carried out. Finally, the results are inserted into a terminological database (Kerremans et al., 2004) and include culture or language-specific information.

Despite the validity of this approach, its cognitive-based principles, which depict the concept as a “unit of understanding” (Temmerman, 2000), do not fit the purposes of this research work. Furthermore, we do not share their view that knowledge can be directly extracted from a textual corpus via manual or semi-automatic corpus processing. In addition, and from a methodological point of view, it is not exactly clear, from the several consulted publications (e.g. (Kerremans, De Baer, & Temmerman, 2010; Kerremans et al., 2004; Kerremans & Temmerman, 2004; Temmerman & Kerremans, 2003, 2005)): [1] how the categorical framework actually supports the corpus collection and corpus analysis stages; [2] which associative relations are used in the framework and whether or not they are aligned with other existing resources so as to enable interoperability; [3] if the resulting “ontology” undergoes some validation in order to allow the verification of its logical properties; and [4] whether or not the resulting terminological database is available in a format that complies with the SW and LD specifications, thereby enabling the necessary interoperability.

Another approach that aims to integrate ontologies in terminology work is Ontoterminology, proposed by (Roche et al., 2009). Defined as a “terminology whose conceptual system is a formal ontology” (ibid., p. 325), this perspective regards the conceptualization of a given subject field as the starting point of any terminological

project. Rather than making them incompatible, Ontoterminology seeks to reconcile Terminology's linguistic and conceptual dimensions while maintaining their core identities and fundamental differences. This is visible in Roche's proposal (2012), shown in the figure below, to extend the classical semiotic triangle by (Ogden & Richards, 1923):

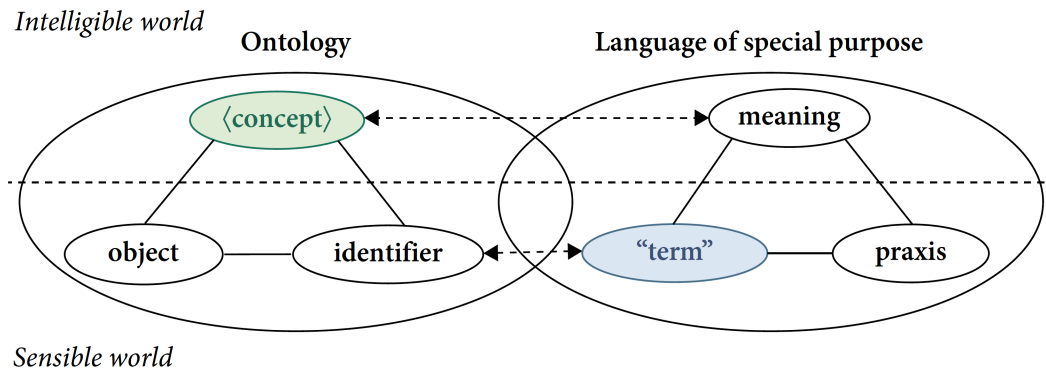


Figure 6 – The double semiotic triangle

From this diagram, it becomes clear that even though both dimensions are present in ontoterminological projects, they rely on two distinct semiotic systems that should not be confused – as pointed out by (Roche, 2007), “Saying is not modelling”. By separating signified (meaning) and signifier (“term”) – pertaining to Linguistics and natural language – from the <Concept> and its respective identifier – which are part of a formal system, Ontoterminology acknowledges a distinction between the definition of the term, written in natural language, and the definition of the concept, written in a formal language. This approach also highlights the key role of subject field experts in an ontoterminological project, stating these should be directly responsible, together with the ontoterminologist, for the conception of the core ontology. Finally, Roche (2007, 2012) advocates the specific difference, Aristotelian-based approach (cf. following section) as the basis for the proposed conceptualizations.

Ontoterminology's focus on consistent and constructive (i.e. allowing the conceptualization to be manipulated computationally) formal concept definitions (Roche, 2012, 2015) can be relevant within the scope of this thesis, as it substantiates the drafting of natural language definitions that will be illustrated in Part III. Moreover, it allows the core conceptualization to be operationalized, which is in line with the

objectives of this work. However, and as it will be shown in Chapter 6, the development of the EndoTerm core conceptual system did not resort solely to the specific difference, Aristotelian-based approach, the one advocated by Roche (2007, 2012), but to the combination of the former with a set of pre-defined categories, conceptual relations, and restrictions. To meet the specific needs of this project, it was believed that the somehow rigid, mono-axial structure of the specific difference approach might make a possible integration into existing biomedical terminological resources more challenging. Furthermore, EndoTerm's conceptualization was obtained neither directly from experts nor directly from texts. As described in Chapter 5, EndoTerm relied on a mixed methodology (both onomasiological and semasiological), where a conceptual analysis of the domain and the outlining of the resulting categories preceded – and supported – the actual corpus processing and analysis stages.

In short, the brief outline provided in this section illustrates an increasingly relevant turn towards the integration of ontologies into terminological products, thereby aiming at their operationalization. Despite the diverging theoretical and methodological approaches that seem to coexist, it is unlikely that this trend may weaken in the coming years. The ability to combine the logical consistency and the interoperability potential of ontologies with the preservation of linguistic diversity is, indeed, quite appealing. And one of the terminological 'products' that benefits the most from this 'alliance' involving the conceptual and the linguistic is precisely the definition, which will be addressed in the following section.

2.4. The definition in Terminology

For centuries, there has been a wide debate concerning the nature of definitions, which has involved various knowledge branches, such as philosophy, logic, linguistics, lexicography, or even mathematics. Within the scope of this research work, the current analysis will solely explore the way the notion of definition has been perceived in Terminology and, indeed, the topic has been widely debated in the community for quite some time (cf. de Bessé, 1997; Löckinger et al., 2015; Rey, 1995; Sager, 2000; Seppälä, 2007; Temmerman, 2000). However, philosophy is still a mandatory starting point in this regard, namely via the contribution from Aristotle. In his works on Logic, the Ancient

Greek philosopher outlined the postulates for a definition or “what it is to be for something”. In other words, what makes something unique (its nature or essence) must be accounted for in its definition (Smith, 2018). As a result, something is defined based on the classical formula *Definiendum* = *df* *differentia* + *genus*, i.e. the closest generic supertype plus a specific difference. In the *Introduction*, or *Isagoge*, to Aristotle’s *Categories*, Porphyry presents a visual proposal of a hierarchical classification depicting Aristotle’s desiderata. As illustrated by **Figure 7**, the process continues until the most specific species is reached. Furthermore, the inheritance principle should be observed, i.e. the species inherits the differentiae from the respective genera. As an example, a Human is a rational Animal and, because of that, also a sensitive Living, an animate Body and, ultimately, a material Substance.

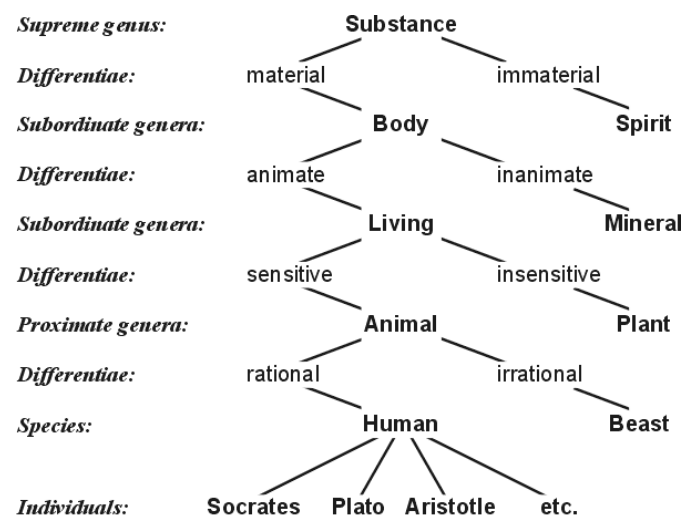


Figure 7 – Tree of Porphyry as drawn by Peter of Spain (1239) (Sowa, 2007)

This Aristotelian approach to definition, which allows the *definiendum* to be distinguished from other related notions while also providing logical consistency (e.g. a Body cannot be simultaneously animate and inanimate), has had a major impact in Terminology, especially in what concerns the set of Terminology-related ISO standards (ISO 1087-1, 2000; ISO 704, 2009).

According to ISO 1087-1 (2000, p. 6), a definition is a “representation of a concept by a descriptive statement which serves to differentiate it from related concepts.” Thus, and within this set of standards, the definition is one form of concept representation, drafted in natural language, whereas a term, as a “verbal designation of a general concept” (cf. section 2.2.), provides another way of representing a concept. This standard also distinguishes between intensional and extensional definitions, stating that the former “describes the intension of a concept (i.e. the set of characteristics which makes up that concept) by stating the immediate superordinate concept and the delimiting characteristics”, while the latter resorts to the enumeration of “all of a concept’s subordinate concepts under one criterion of subdivision” (ibid., p. 6-7). A delimiting characteristic, in turn, is a type of essential characteristic³⁹ that aims to “distinguish a given concept from other, related concepts” (ibid., p. 3).

Hence, the core structure of the intensional definition matches its Aristotelian counterpart – superordinate concept (genus) + delimiting characteristic (specific difference). The main role of an intensional definition is, therefore, to allow a given concept to be identified and differentiated from other concepts (cf. ISO 704, 2009, p. 22). Moreover, it should be built in such a way as to reflect the place of the concept within the concept system. On the other hand, the concept system can “facilitate the writing of definitions” (ISO 704, 2009, p. 19). Overall, intensional definitions have become the “standard way of illustrating concepts” (Löckinger et al., 2015, p. 66), so it is no wonder that several guidelines have been provided in the literature to support the drafting of this type of definitions (e.g. Cabré, 1999; ISO, 2009; Löckinger et al., 2015; Suonuuti, 2001). Ideally, and just to name a few examples, intensional definitions should be precise, concise, objective (i.e. unbiased), reliable as far as the source is concerned, linguistically accurate, non-circular, and should also refer to the immediate superordinate concept (or another one that is more generic), as well as to the domain in question.

³⁹ An essential characteristic is one that is “indispensable to understanding a concept” (ISO 1087-1, 2000, p. 3).

As evident as the preceding rules may be in theory, the actual process of drafting intensional definitions in natural language is not always exempt from problems: firstly, it can be very time-consuming, especially when expert participation and validation is involved. However, those hurdles can be minimized by adopting a clearly defined methodology and a set of mediation strategies (Costa, Silva, & Almeida, 2011, 2013; Costa et al., 2012; Silva, 2014). Secondly, and if supported exclusively on lexico-semantic relations, the definitions may very likely be inconsistent, since they will be language-dependent and, in addition, hampered by the problems that characterize natural language, such as polysemy, ellipses, etc. Thirdly, despite the support provided by a conceptual system, intensional definitions anchored in the aforementioned Aristotelian tradition may be too rigid and unable to fully depict the multidimensionality that characterizes some concepts and their respective subject fields (Bowker, 1997; Kageura, 1997). For (Seppälä, 2015), the definitions within some subject fields should be supplemented with non-hierarchical relations (e.g. agent, function, composition, etc.). In fact, and notwithstanding their secondary role in the current Terminology-related ISO standards⁴⁰, non-hierarchical conceptual relations are regarded as “equally important and more revealing about the nature of the concepts” (Sager, 1990, p. 34)⁴¹ and can, according to (García de Quesada, 2001), make a valuable contribution in accounting for the complexity of a given subject field.

Thus, some approaches have been put forward as a way to provide a more flexible, yet equally consistent, alternative to the Aristotelian-based intensional definition. One of the most prominent refers to the so-called definitional templates, derived from Frame-based approaches to terminology work (Faber Benítez, 2009; Faber, 2012, 2015), as well as to lexicography (Maks, 2006; Swanepoel, 2011). Following this theoretical perspective, the template-like structure comprises schematic representations of the most prototypical relations involving a set of concepts that belong to the same category. As a result, one obtains a stable framework that guides

⁴⁰ Both the ISO 1087-1 (2000) and the ISO 704 (2009) provide a rather vague definition of “associative relations”, stating that they exist “when a thematic connection can be established between concepts by virtue of experience” (ISO 704, 2009, p. 17) and listing a little under 20 of such relations (*ibid.*, pp. 17-18).

⁴¹ To this day, and to our knowledge, the most comprehensive inventories of non-hierarchical conceptual relations within the scope of Terminology have been proposed by (Sager, 1990) and (Nuopponen, 2011, 2014).

the drafting of natural language definitions, which are regarded as “mini-knowledge representations” (Faber, 2002, p. 345). In line with Martin’s proposal of frame-based definitions (Martin, 1998), these templates entail a slot-filler core structure, in which the slots correspond to domain-specific conceptual relations, whereas the fillers (values) are to be replaced by the concepts linked to the *definiendum* via the referred conceptual relations. Within the scope of Frame-based Terminology, such definitional templates have been used in various domains, namely healthcare (oncology) (García de Quesada, 2001), the environment (EcoLexicon) (Faber, 2012; Faber et al., 2014, 2016), or even tourism (Durán-Munoz, 2016).

As described in the previous section, the incorporation of ontologies in terminology work in the last few decades has further developed the results of that work. Natural language definitions were no exception. By providing a language-independent, logically consistent, and computer-processable foundation, ontologies can contribute to optimize the drafting process without threatening the added value brought about by linguistic diversity. In short, formal concept definitions can directly support the design of natural language definitions. The following chapters will show how that contribution is currently making its way into the biomedical domain and how, in what regards definitions, EndoTerm aims to complement the advantages of having a formal conceptual system containing both hierarchical and non-hierarchical relations with those of adopting a systematized and pattern-based definitional structure.

Chapter 3 – Representing, organizing and sharing knowledge within the biomedical domain: towards an interoperable future?

“We are emerging into a new information world, with profound impact upon our systems of healthcare. Underpinning this future must be a robust system of clinically describing what we see, do, learn, and discover. These descriptions, in turn, must be transmissible between systems, enterprises, knowledge resources, and user communities.” (Chute, 1998, p. 72)

3.1. Introduction

Published twenty years ago, the preceding quote is still relevant today, as it refers to some of the ongoing issues in healthcare and healthcare-related systems that have been explored in the previous chapter, namely the exponential growth of technology-based tools and procedures within healthcare, which, together with the emergence of computerized patient records, have been posing considerable challenges to the various stakeholders while aiming to respond to their multiple needs. Moreover, the increasing demand for shareable clinical information across different providers, institutions and countries, has placed special emphasis on the concept of interoperability analyzed in Chapter 1.

In fact, the key to an efficient and effective healthcare delivery appears to lie more and more on what could be described as the 3 Cs: the access to **consistent**, **comparable**, and **concept-oriented** patient data representations, which will, in turn, enable their reuse for billing and reimbursement, for resource allocation, or as basis for decision support systems, among other examples. However, to achieve that interoperability goal, it is crucial that clinical information – and also clinical knowledge – can be represented and organized in a systematic and, as much as possible, standardized way.

For decades, efforts have been undertaken to make this representation explicit, either through free text and semi-structured forms or by resorting to more formalized resources, such as coding schemes or classifications. Each of these examples has its specificities, advantages and limitations. Nonetheless, they share the challenging

objective of attempting to bring (some) order and structure to something as complex and multifaceted as healthcare terminology (cf. ISO, 2007).

In fact, the difficulties inherent to the creation, maintenance, and sharing of healthcare-related terminological resources derive from a set of linguistic, conceptual, and operational factors which will be further explored in this chapter. Firstly, the ambiguity that often characterizes natural language, even within specialized contexts, should be highlighted. The presence of synonymy, polysemy, ellipsis, vagueness, etc. conveys a certain degree of uncertainty to the interpretation of patient/clinical records, which can, in itself, be demanding for human users, let alone for computer-based systems (Coiera, 2015; Duclos et al., 2014).

Secondly, the notion of concept within the healthcare domain is not always linear and absolute (Chute, 2005; Thagard, 1997). Clinical concepts are deeply rooted in social, economic, cultural, technological, scientific, and even religious postulates, thereby leading to their evolution over time and to the existence of different conceptions of disease within different cultures (e.g. Western Medicine vs. Traditional Chinese Medicine (TCM) vs. Ayurvedic or Traditional Indian Medicine (TIM) - (Maciocia, 2015; Nordenfelt, 2007; Patwardhan, Warude, Pushpangadan, & Bhatt, 2005)).

Finally, the fact that such terminological resources often coexist in different and increasingly large systems, and that they are currently expected to tackle the needs of both human users and machines, whilst addressing a wide range of purposes (e.g. coding patient data, supporting information retrieval, indexing biomedical literature, etc.), can entail additional challenges from an operational perspective, which may be difficult to overcome (cf. Rector, 1999).

It is believed, however, that Terminology, given its double-dimensional nature (cf. previous chapter), can play a decisive role in dealing with these hurdles, as well as with the silo-approach referred to in Chapter 1. In a context characterized by the “inevitable uncertainty” of healthcare and healthcare terminology, the key to overcome these challenges can lie in the ability to represent “that uncertainty with precision” (Cimino, 1998). Therefore, and acting as a “conceptual glue” (Chute, Cohn, & Campbell, 1998, p. 509), especially within the scope of knowledge organization, Terminology can provide the necessary framework to devise consistent conceptual representations of a

given domain while maintaining and systematizing linguistic diversity. In short, it can contribute to create resources that can: [1] limit the ambiguities and imprecisions of natural language; [2] address the needs of both human and machine end-users; and [3] enable the reuse of clinical information for multiple goals. An ambitious endeavour indeed, as will be seen throughout this chapter. In fact, a multitude of medical vocabularies and terminological systems have been created in the last few decades, most of which have been progressively discarded and/or forgotten (Chute, 1998). Nevertheless, recent advances within the field of Biomedical Informatics⁴² (BMI) have led to the emergence and development of more advanced resources, with an increasingly solid knowledge-based approach and a clear aim towards semantic interoperability.

3.1.1. The dynamics of knowledge representation in healthcare

As regards the biomedical domain, as in others, knowledge can be **represented** in numerous formats, and **Figure 8** illustrates several examples specifically related to the field of endometriosis: [1] a scientific article presenting a consensus statement about the disease; [2] the institutional website from a healthcare provider containing information about the disease and targeted at a broader audience; [3] a monolingual glossary of endometriosis, available in the global online platform endometriosis.org⁴³; [4] a diagram containing the representation of endometriosis according to SNOMED CT (to be explored later in this chapter); [5] an image depicting the different locations of endometriosis lesions; and [6] a screenshot of a video with 3D animations describing a hysterectomy.

⁴² Over the last decade, especially since 2012, with the publication of a White Paper on that matter by the Board of the American Medical Informatics Association (AMIA), the designation Biomedical Informatics (BMI) has gradually replaced “Medical Informatics” as the one most widely accepted for the “interdisciplinary field that studies and pursues the effective uses of biomedical data, information, and knowledge for scientific inquiry, problem solving, and decision making (...)” (Kulikowski et al., 2012; Shortliffe & Cimino, 2013).

⁴³ <http://endometriosis.org/>

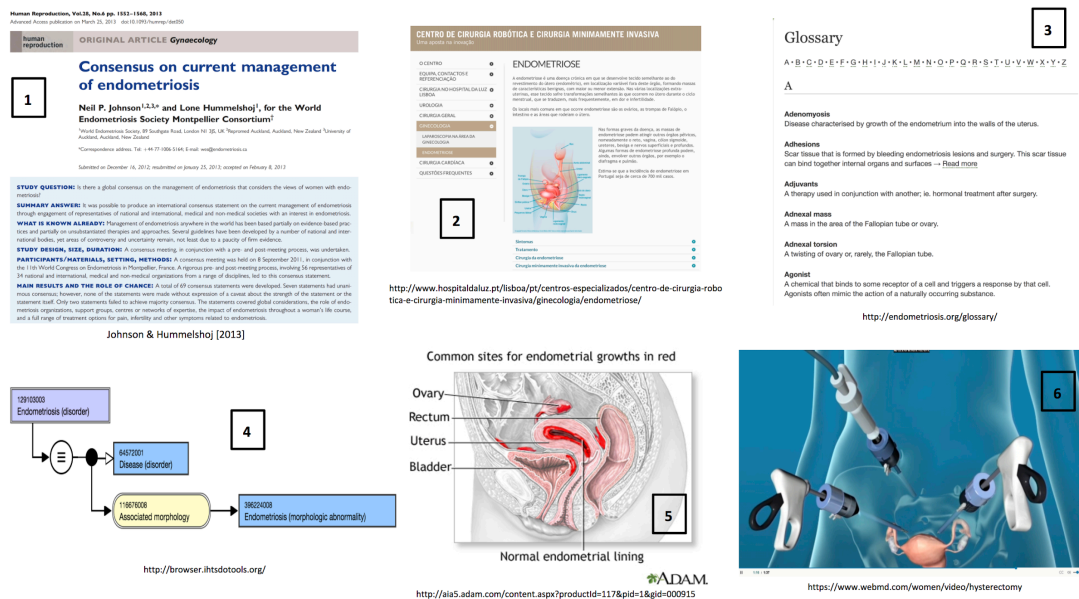


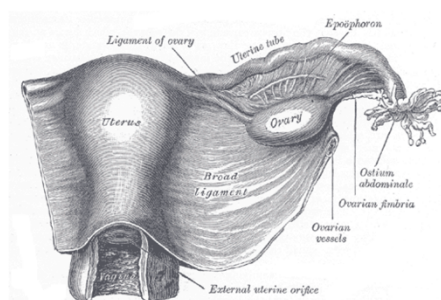
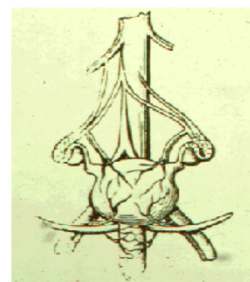
Figure 8 – Examples of knowledge representation within the scope of endometriosis

Although the preceding resource sample is rather small, it is possible, nonetheless, to observe the diversity that underlies knowledge representation in this particular field, as in others within the biomedical domain. This diversity also illustrates the previously stated exponential technological progress, which has given rise to new techniques, procedures, instruments, findings, in short, to new knowledge, whose representation is increasingly visual and multimedia-based, thus differing from more ‘traditional’ forms, such as plain text or illustrations. On the one hand, this visual approach to knowledge representation is not new in Medicine, and its relevance has been unequivocal. For centuries, medical illustrations have contributed significantly to the representation and dissemination of medical knowledge and their quality and degree of precision have gone hand in hand with scientific and technological progress (some of the work by Leonardo da Vinci, Andreas Vesalius or Henry Gray has been depicted in **Figure 9**, especially the one related to the female reproductive system).



Da Vinci's *Fetus in the Womb* (1510-1513)

Vesalius's *De Humani Corporis Fabrica* (1543)



Gray's Anatomy (1918)

Figure 9 – Examples of medical illustrations

On the other hand, new multimedia-based resources are becoming more and more important within the medical community, which has led more recent terminological projects within the healthcare domain to rely on multimodal corpora (cf. (Prieto-Velasco & López Rodríguez, 2009; Prieto Velasco, 2016; Prieto Velasco & Tercedor Sánchez, 2014)). One of the most promising examples of such multimedia-based resources is the **medical video article**, a peer-reviewed and indexed scientific article fully presented in video format. As stated by (Falcone & Setúbal, 2013), this type of video article constitutes a new paradigm in clinical research, since these videos “are not adjunct material but the focal point of the research” and represent “a unique way to present scholarly material, (...) a visual presentation of research”. Hence, it is believed that if issues concerning access restrictions and copyright can be overcome, these medical video articles can soon become indispensable sources for multimodal corpora in terminological projects within the biomedical domain.

As will be shown throughout this chapter, the diversity that characterizes medical knowledge representation is also present in the way that knowledge is organized and shared. Accordingly, the following section of this chapter will focus on

knowledge organization, particularly on some of the main terminological resources in healthcare today (e.g. the International Classification of Diseases (ICD), the Medical Subject Headings (MeSH), the Disease Ontology (DO), among others), as well as on their core structure, with the purpose of exploring how these resources depict Endometriosis⁴⁴. The final section of the chapter will be dedicated to how knowledge stemming from these resources can be shared, especially within the scope of the Semantic Web and Linked Data.

However, and as it is important to understand how the systematization of healthcare terminology evolved, the next subsection will present a historical outline regarding the development of terminological resources in this subject field.

3.2. Organizing biomedical knowledge: learning from the past and innovating towards the future

3.2.1. The early days of knowledge systematization and classification in Medicine: lessons from history

The need to organize and provide some structure to whatever surrounds us is inherent to human nature. Medicine and its adjacent fields are no exception. From very early on, attempts have been made to classify, or group together, phenomena such as diseases and their respective etiology, location, and symptoms, as well as procedures, both surgical and non-surgical.

The first proposal to categorize diseases, at least the first from which there are clear, written records, was put forward by the Ancient Greek physician Hippocrates, often regarded as the father of Western medicine. Besides introducing designations such as symptom, diagnosis, therapy, trauma, and sepsis, he was a pioneer in encouraging objective observation of the patient and critical deductive reasoning, as opposed to the current view at that time, which perceived disease as a supernatural phenomenon (Kleisiaris, Sfakianakis, & Papathanasiou, 2014). Hippocrates's approach, claiming that disease resulted from an imbalance of one or more of the four humors

⁴⁴ This will, in turn, provide a fundamental basis for the domain-based conceptualization proposal put forward in Chapter 6.

(blood, phlegm, yellow bile, and black bile), constituted the basis of a classification-like outline, included in Vol. V – *Affections, Diseases I & II* (Hippocrates of Cos, 1988) - of what is known as the Hippocratic Corpus⁴⁵. In **Figure 10** below, adapted from the preceding work, it is possible to observe the existence of different axes underlying the disease arrangement, which refer either to their anatomical location (Intestinal Diseases), to their severity (Acute Diseases) or to their probable etiology (Summer Fevers and Pains).

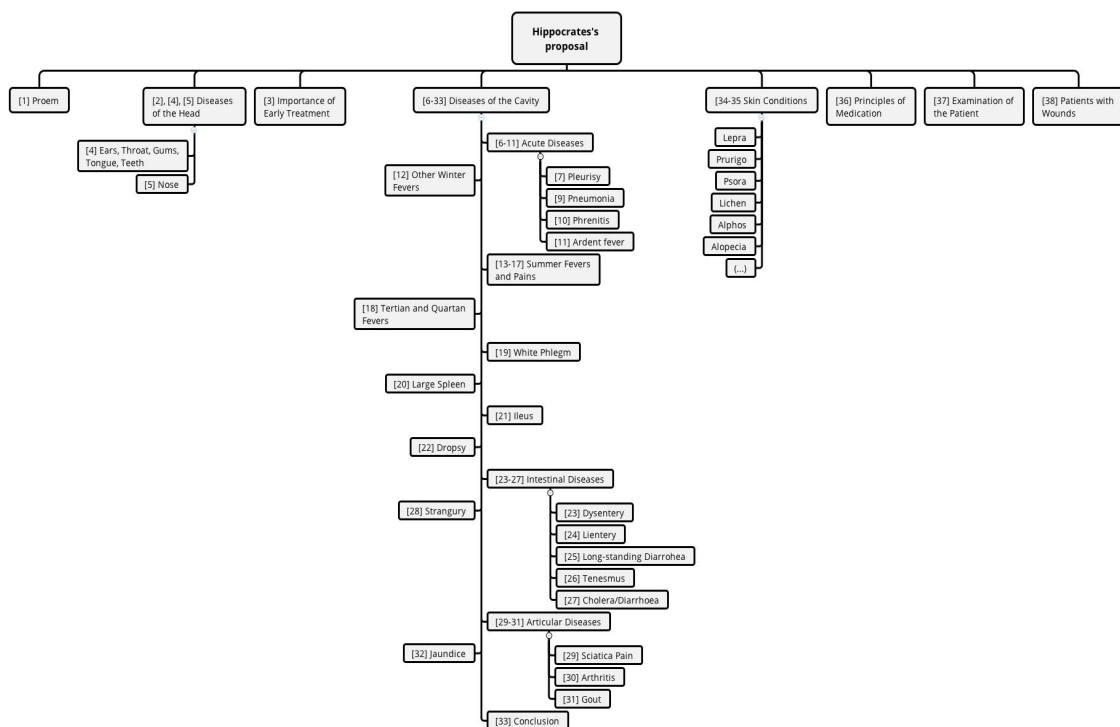


Figure 10 – Hippocrates’s proposal concerning the classification of diseases

Another contribution to this topic was provided by Plato: in *Timaeus*⁴⁶, written around 360 BC, the monologue by Timaeus of Locri, the main character, aimed to thoroughly describe the creation of the universe and included a proposal to classify diseases of both body and soul by dividing them into three classes: i) those due to the imbalance or maldistribution of the elements (air, water, earth, and fire); ii) those due

⁴⁵ A collection of about 60 early Ancient Greek medical texts, presumably written between the VI and IV centuries BC, and commonly associated with Hippocrates and his teachings.

⁴⁶ Available at <http://classics.mit.edu/Plato/timaeus.html>.

to a disordered sequence of tissue formation; and iii) those due to an abnormal accumulation of air (wind), phlegm or bile (Zeyl & Sattler, 2017).

A few centuries later, Aulus Cornelius Celsus, in Book 3 (*Specific Diseases*) of his seminal treatise *De Medicina*, referred to the premise, advocated by the Ancient Greek, of dividing diseases into acute and chronic, as unsatisfactory, considering that “it is clear that there are more than two classes” (Celsus, 1971). According to this Roman encyclopedist (and likely practicing physician), diseases could either be [1] acute, [2] chronic, [3] both acute and chronic at times (e.g. fevers), and [4] neither clearly acute - as they were not fatal - nor chronic, since they could be readily cured if treated. He thereby focused on body structure as the main axis of his disease classification proposal, distinguishing between “those [diseases] which appear to have their seat in the body as a whole, and (...) those which originate in particular parts” (Celsus, 1971).

One of the most influential contributions to this matter was also developed by the Greek physician, writer and philosopher Galen of Pergamum. He envisioned disease as something contrary to nature, further exploring the Hippocratic assertion of it being a state of imbalance⁴⁷. In his ground-breaking works - *De morborum differentiis* and *De symptomatum differentiis* -, he attempted to define and classify both diseases and symptoms in a systematic way, despite the lack of clear-cut boundaries of both concepts at that time (and even now, one could claim) (Johnston, 2006). The types of ‘imbalance’ brought about by disease were thus organized in three *genera*, depending on which anatomical structures were affected:

[1] *Dyscrasias*, which were applicable to *homoiomeres*, or uniform body parts (e.g. muscle, bone, cartilage, arteries), and brought about by an imbalance in one or more of the four primary elements (heat, cold, moisture and dryness), either due to excessive constriction or to excessive dilatation;

[2] Abnormalities of morphology and composition, applicable to organs (e.g. heart, liver, lungs). These could then be further subdivided into: [2.1.] Disorders of formation (abnormalities of overall form, both congenital and acquired; changes in the number and configuration of channels and cavities; and changes in the roughness and

⁴⁷ This notion of disease as an imbalance is still widely disseminated, also in Chinese Traditional Medicine and Ayurveda, or Traditional Indian Medicine.

smoothness of their surfaces); [2.2.] Disorders of the number of parts (either by deficiency or excess); [2.3.] Disorders of the size of parts (by being bigger or smaller than they should be); [2.4.] Disorders of relative position (e.g. dislocations and hernias).

[3] Dissolution of continuity, which could be applied to both *homoiomeres* and organic structures (e.g. a limb that is simultaneously inflamed/swollen and dislocated) (Johnston, 2006).

Notwithstanding the reservations surrounding this proposal, even at that time, Galen's attempt to provide an explicit framework (anatomical structure) for the categorization process was innovative and had a key impact in medical theory and practice in Europe from the Middle Ages until the mid-17th century.

It was precisely in the 17th century that another breakthrough took place, namely that human diseases should be thoroughly registered and classified, similarly to what botanists were doing with plants. This branch of medical science that deals with the classification of diseases would later be named nosology. In this context, one of the most significant inputs was provided by the English physician Thomas Sydenham, often regarded as one of the precursors of clinical medicine and epidemiology. In his work on acute and chronic diseases, Sydenham stressed the importance of grouping diseases under the same genus and then specifying each one of these pathologies according to a set of criteria, in particular by enumerating "both the peculiar and constant phenomena or symptoms, and the accidental ones separately" (Sydenham, Greenhill, & Latham, 1850).

The development of a clear and coherent foundation for the classification of diseases that could enable data comparison was also at the heart of John Graunt's work. Generally considered to be the main founder of demography, Graunt proposed a pioneering analysis, primarily for statistical purposes, which was based on the London Bills of Mortality, the weekly statistics of deaths in early modern London, created as an attempt to anticipate the onset and spread of the bubonic plague. In the work *Natural and Political Observations Made upon the Bills of Mortality* (Graunt, 1977 edition), Graunt was able to estimate the city's population, to infer ages at death and to outline the proportion of deaths in different age groups, based on his classification proposal of causes of death (cf. Chute, 1998; Coiera, 2015).

The 18th century witnessed a deepening connection regarding the methodical arrangement and classification of both diseases and plants since some of the most prolific authors and scientists of that time were involved in these two fields. François Boissier de Sauvages de Lacroix, for instance, a French physician and botanist, is credited with the first attempt to classify diseases systematically, in line with Sydenham's work. In his 5-volume *Nosologie Méthodique, ou Distribution des Maladies en Classes, en Genres et en Espèces, suivant l'Esprit de Sydenham et la Méthode des Botanistes* (Sauvages, 1772), he presented a remarkably comprehensive list of individual diseases organized into 10 classes, 295 genera and 2,400 species. Linnaeus's *Genera morborum* (Linnaeus, 1763) and William Cullen's *Nosology: or a Systematic Arrangement of Diseases, by Classes, Orders, Genera, and Species* (W. Cullen, 1800) were two further examples of prominent disease classification proposals at that time, with the latter becoming the most widely used for statistical purposes both in the 18th and 19th centuries. As Sydenham had done before, Cullen also highlighted the distinction between what is "accidental" and what is "essential" to establish clear-cut categories when devising a disease classification.

In the first half of the 19th century, the so-called Cullen classification was chosen as the basis for the work to be developed by the newly founded General Register Office (GRO) of England and Wales (created in 1836), namely in keeping a registration of both births and deaths in the aforementioned territories to help with the transfer of land property rights from one generation to the next (Lilienfeld, 2007). Yet a few years later, this classification was deemed unsatisfactory by William Farr, the GRO's first medical statistician, for whom the advantages and importance of a uniform statistical nomenclature were "obvious", since, and according to Farr, in the Cullen classification and in others used at that time "each disease has, in many instances, been denoted by three or four terms, and each term has been applied to as many different diseases: vague, inconvenient names have been employed, or complications have been registered instead of primary diseases" (WHO, 2016c, p. 192). Thus, Farr concluded that "the nomenclature is of as much importance in this department of inquiry as weights and measures in the physical sciences and should be settled without delay" (ibid.). Farr's subsequent work, which consisted of classifying diseases as epidemic, constitutional

(general), developmental, resulting directly from violence, as well as local diseases arranged according to anatomical site, was presented to and adopted by the International Statistical Congress in the 1850s. His model, especially the idea of differentiating general diseases from those located in a particular organ or anatomical site, survived as the core principle of what later became the International List of Causes of Death.

Also known as the Bertillon classification, due to the valuable contribution by Jacques Bertillon, the Chief of Statistical Services of the City of Paris who chaired a committee responsible for the preparation of a uniform classification of causes of death, the International List of Causes of Death began to have widespread international use and was the forerunner to the current International Classification of Diseases (ICD), which will be presented in more detail in one of the following sections (Coiera, 2015; WHO, 2016c).

As can be seen from the previous paragraphs, the systematic organization and consequent classification of diseases and respective symptoms have evolved and played an important role in attempting to solve practical issues as they emerged. It also becomes clear that they served mainly epidemiological, statistical, administrative and fiscal purposes. With the advent of technology-based healthcare, however, needs have arisen that go far beyond the purposes above. One of the most immediate responses at the end of the 20th century was to extend the content and structure of those traditional resources to meet those demands. However, it quickly became clear that to adequately represent the complexity inherent to healthcare and related terminology, as mentioned in the introductory section of this chapter, simple coding systems and classifications were no longer suitable. A more profound change was, therefore, necessary in this field and indeed, the last two decades have witnessed substantial and noteworthy progress. The following section will, therefore, be devoted to this evolution and to the 21st-century biomedical terminological resources resulting thereof.

3.2.2. A disease is a disease is a disease: challenges for the 21st century biomedical terminological resources

As mentioned above, considerable changes have recently occurred within the scope of biomedical terminological resources, often grouped under the umbrella notion of Terminological Systems (TS) by the ISO Technical Committee 251 (“Health Informatics”). Defined as a “set of designations within the domain of health care with, when appropriate, any associated rules, relationships and definitions” (ISO, 2012), this notion, as it is, does not fit the objectives of this thesis and the project it aims to present, since it does not address the conceptual dimension of Terminology and the resulting terminological resources within this subject field. Hence, it is believed that the definition proposed by (Duclos et al., 2014, p. 22) would best suit the purposes of this research work: a terminological system aims to organize “the relationships between terms and concepts in a domain with, when appropriate, any associated rules, relationships, definitions and codes”.

Indeed, in recent years, the core structure of the TS above and several others has become increasingly concept-oriented, thus paving the way to a knowledge-based approach (Cimino, 2000, 2001; Elkin & Tuttle, 2012). More specifically, many resources have been designed or redesigned to incorporate ontology-related elements, such as formal concept definitions and computer-processable formalisms, thereby making the task of defining clear-cut boundaries between these resources more and more difficult.

Concept orientation has, therefore, become one of the key principles in current biomedical terminology and was, in fact, one of the twelve desiderata that, according to Cimino, should support terminological systems in this subject field in the 21st century. In his seminal paper, (Cimino, 1998) proposed a set of 12 principles that should underpin the development of “standard, reusable, multipurpose, and shareable controlled medical vocabularies for the 21st century”. As previously mentioned, one of Cimino’s fundamental ideas was that the concept, defined as a “clinical meaning”, should become the core unit within the terminological resource. In that sense, each concept should have a single, coherent meaning, i.e. being non-vague, non-ambiguous, and non-redundant, and would be defined formally - via a collection of relationships to other concepts in the TS (both hierarchical and non-hierarchical) -, in a way that would allow computational manipulation. In order to support this system’s implementation, each concept should

be assigned a unique non-semantic identifier - in other words, an identifier that would not be tied to hierarchical information (i.e. the concept's position within the resource's hierarchy) or to any other implicit meaning.

Given the dynamic nature of these 21st century resources, Cimino also postulated that the necessary content expansion should be anchored in a solid methodology that would facilitate a compositional, rather than an enumerative, approach⁴⁸. Furthermore, multipurpose resources would need to enable concept arrangements that would suit all the different requirements, thus allowing multiple hierarchies to coexist in a resource - i.e. polyhierarchy. In addition, and due to the multitude of users and respective demands, the TS should be able to depict differences in granularity through which the concepts would be expressed (e.g. a medical student, a general practitioner (GP), or a gynecologist/obstetrician (GYN/OBN) would very likely require different levels of granularity when consulting a resource on Endometriosis). This diversity should also be illustrated in the numerous view(s) each type of user could have access to within the resource. Regardless of the view, the concept information should always be consistent with the predefined conceptual model and respective definition. Aside from these structural principles, Cimino outlined a set of other rules concerning the maintenance of these systems (update frequency, version control, etc.), as well as mechanisms for coping with redundancy and nonsense representations (e.g. use of constraints or sanctions).

Cimino's proposal has, indeed, paved the way for a major shift in biomedical TS. As a consequence, other authors or organizations have followed or expanded his *Desiderata* – and Cimino himself reinforced his proposal eight years later (Cimino, 2006). Later in 1998 and in the following year, (Chute et al., 1998; Rector, 1999) referred to Cimino's work as a stepping-stone to their contributions in devising a framework for representing, updating and sharing biomedical information in an increasingly computer-

⁴⁸ A compositional approach, also called post-coordination, implies that a resource's basic building blocks must be previously identified and subsequently combined within that resource according to a set of rules and constraints, in order to avoid nonsense representations. An enumerative or pre-coordinated approach, on the other hand, implies that any complex component must be pre-assembled within the resource, i.e. all the elements to be used should be listed in advance (Coiera, 2015, pp. 375–376; 406–308; Duclos et al., 2014, pp. 20–21). The implementation of the aforementioned rules and constraints, which takes place within the scope of the so-called categorial structure, will be further explored in this subsection.

based healthcare system. **Table 3** summarizes the similarities regarding these three perspectives:

Cimino [1998]	Chute et al. [1998]	Rector [1999]
Content, Content, and Content	Completeness + Comprehensiveness Atomic concepts + compositional approach	Scale and diversity of use [2.1]
Concept orientation	Nonredundant and integrated terminologies	Separation of language + concepts [3.1] Formal concept representation [3.3]
Concept permanence	Unique identifiers	Separation of language + concepts [3.1] Formal concept representation [3.3]
Non-semantic concept identifier	Context-free identifiers	Multiple views and multiple hierarchies [3.3.3]
Polyhierarchy	Multiple hierarchies	Multiple views and multiple hierarchies [3.3.3]
Formal definitions	Explicit definitions Logical structure	Formal concept representation [3.3]
Reject "Not Elsewhere Classified"	-----	The legacy of existing classifications and coding schemes [4.2]
Multiple granularities	-----	Differences in granularity [3.3.2]
Multiple consistent views	Consistency of view in multiple hierarchies	Multiple views and multiple hierarchies [3.3.3]
Beyond medical concepts: representing context	-----	Meeting the needs of EHRs [4.3]
Evolve gracefully	Version control	Change management [4.4]
Recognize redundancy	Synonymy	Criteria for classifying concepts + criteria for determining when 2 concepts are equivalent [3.3.1]
	Centralized coordination	
	Wide distribution and affordable pricing	
	Funding from both public and private sectors	

Table 3 – Overview of Cimino (1998), Chute et al. (1998) and Rector's (1999) approaches

Through the information above, it can be concluded that one of the most fundamental aspects uniting these approaches is the distinction between what is conceptual and what pertains to natural language. The theoretical and methodological implications of this assumption have underpinned the much-needed operationalization of biomedical TS in recent years. In this respect, reference should be made to the GALEN project, one of the leading and pioneering initiatives in that area.

An acronym for General Architecture for Languages, Encyclopedias and Nomenclatures in Medicine, the GALEN consortium was funded by the European Commission and integrated academic, clinical, and industrial partners from seven countries (e.g. the University of Manchester, Hewlett-Packard, the Hôpital Cantonal Universitaire de Genève, etc.)⁴⁹. Its goal was to support what was believed to be an indispensable paradigm shift from traditional, enumerative-type-classifications that mixed concepts and terms, lacked formal structure and relied on static systems, into a compositional approach based on formal concept definitions that could support language-independent concept systems, computational manipulation and, hence, more dynamic terminological resources and services (Rector et al., 1997; Rector, Nowlan, & Glowinski, 1993; Rector et al., 1994; Rector, Nowlan, & Consortium, 1994; Rogers et al., 2001).

At the heart of this project was the GALEN Terminology Server, which aimed to deliver terminological services within the biomedical domain so as to “mediate between different systems and applications” (Rector, Nowlan, & Glowinski, 1993) such as EHRs, clinical user interfaces, decision support systems, natural language processing or knowledge management systems. Its foundation was anchored in the Concept Reference (CORE) Model, which, as the name indicates, provided a conceptual framework that, according to current patterns, could be perceived as an “ontology” (Rector & Rogers, 2006), with its clear focus on the distinction between concept and term, as well as on formal concept definitions.

⁴⁹ The project is currently deactivated, but a knowledge repository can be found at <http://www.opengalen.org/>.

To enable this latter point, the CORE Model resorted to a Description Logic-related formalism known as GRAIL (GALEN Representation And Integration Language). Among its distinguishing features, GRAIL allowed a homogeneous classification of categories and entities⁵⁰, along with the differentiation between defining and essential criteria (cf. previous chapter for the essential and delimiting characteristics). Besides the CORE Model, the Terminology Server included a Multilingual Model, which aimed to supply lexicons and grammatical information in multiple natural languages, thereby allowing phrases to be expressed and understood (Rector et al., 1997). In addition to laying the foundations for the compositional approach, the GALEN project also provided a valuable contribution to outlining a set of rules and constraints that should govern the way that composition took place. GALEN's "sanctioning" principle allowed the system's developer to know which compositional representations were coherent (Rector, Nowlan, & Glowinski, 1993).

This approach has also been adopted by the main standards in the field of Health Informatics - the ISO 17115:2007, the ISO 1828: 2012, and the CEN 12264: 2005, under the notion of categorial structure, defined by the ISO 17115:2007 standard as a "minimal set of domain constraints for representing concept systems in a subject field" (p. 4). In short, the categorial structure comprises a set of basic sanctions or limitations required to represent a given concept system within a subject field. These constraints will, in turn, determine which categories (i.e. concepts in DLs) and their respective relations will be authorized, to prevent nonsense representations (e.g. a liver fracture). Ultimately, the categorial structure can play a key role in TS, not only by supporting reasoners with automatic inference and subsumption of the subordinate concepts but also by acting as a sort of "template" for more consistent formal concept definitions. In fact, this notion will be particularly useful in the development of EndoTerm's conceptual framework, as well as in assisting the definition drafting process (cf. Chapters 6, 7 and 8).

So far, the core principles underlying the large majority of current TS have been outlined: concept orientation, language-independent concept systems, as well as the compositional approach, supported by the categorial structure. Yet in order for these resources to become truly machine-processable, a formalism is required to adequately

⁵⁰ Known in Description Logics as concept and individual, respectively (Baader, 2003).

represent these foundations. In addition to GRAIL, there are other formalisms to be considered, whose use depends on the resource's requirements and fundamental structure. Within the biomedical subject field, examples of TS can be found in the last few decades which have resorted to conceptual graphs as devised by (Sowa, 1992, 2000) (e.g. the CANON Group – (Evans, Cimino, Hersh, Huff, & Bell, 1994)), to frame-based structures as proposed by (Minsky, 1977) (e.g. the Medical Entities Dictionary – (Cimino, 2000, 2001)), and, more recently, to OWL (Kibbe et al., 2015; Schriml et al., 2012), as well as to RDF (Boehr & Bushman, 2018; Salvadores, Alexander, Musen, & Noy, 2013).

These core principles regarding today's most prominent terminological systems, as well as their connection to endometriosis, will be seen in more depth in the following subsection.

3.2.3. Is What You See Really What You Get? Current biomedical terminological systems and <Endometriosis>

Taking the preceding aspects into account, it was believed that the analysis of how the concept of <Endometriosis> is depicted in a set of biomedical terminological systems could provide a valuable insight into this subject field and, in turn, play a critical role in the development of the conceptualization proposal put forward in Chapter 6. It would, however, be unwise to compare those results and draw overall conclusions from them, as each one of these resources serves a particular context, purpose, and set of users, which then determine their epistemological principles, core structure, concept organization, as well as the language(s) of expression. For Coiera (2015, p. 397), one should avoid extrapolating results and conclusions when comparing the various resources, since, and “unsurprisingly, the same clinical concept may look very different when coded using different classification systems”.

Several TS typologies have been proposed throughout the years, as summarized in the table below:

TERMINOLOGICAL SYSTEMS	(de Keizer et al., 2000)	(CEN, 2005)	(ISO, 2007)	(Duclos et al., 2014)
Classification	✓	✓	✓	✓
Coding system	X	X	✓	✓
Coding scheme	X	✓	✓	X
Nomenclature	✓	✓	X	✓
Ontology	✓	X	X	✓
Taxonomy	X	X	X	✓
Terminology	✓	✓	✓	✓
Thesaurus	✓	✓	X	✓
Vocabulary	✓	✓	X	✓

Table 4 – Typologies of biomedical terminological systems

It should be noted, though, that the boundaries among these different types of resources have become more and more blurred, in such a way that the term “ontology” is often used indistinctly to refer to all of them. (Grabar, Hamon, & Bodenreider, 2012, pp. 376–377) refer, in fact, to “the lack of precise distinction among [...] resources” that can be found in the biomedical literature. Consider, for instance, BioPortal⁵¹, an open repository of biomedical ontologies, that grants resources such as SNOMED CT, the Medical Subject Headings, the International Classification of Diseases-10th edition, and the Medical Dictionary for Regulatory Activities, among others, the exact same status, i.e. of “ontologies” (ibid.), despite their fundamental differences, as will be illustrated below.

Therefore, and among the types of TS to be presented in the following subsections of this chapter, five were chosen as the most representative: [1] a classification; [2] a thesaurus; [3] a clinical terminology; [4] an ontology; and [5] a clinical information repository. According to the ISO 17115:2007 (p. 6), a **classification** is “an exhaustive set of mutually exclusive categories to aggregate data at a pre-prescribed level of specialization for a specific purpose”. In healthcare informatics, there are several

⁵¹ Cf. <https://bioportal.bioontology.org/>

examples of classifications which aim to code diseases and provide statistical support to international comparisons and clinical decision-making. Some of these statistical classifications refer solely to diseases (e.g. the International Classification of Diseases – ICD, which will be further explored below), while others focus on procedures, both surgical and non-surgical, such as the upcoming International Classification of Health Interventions – ICHI⁵². A **thesaurus**, on the other hand, is a “controlled and structured vocabulary in which concepts are represented by terms, organized so that relationships between concepts are made explicit, and preferred terms are accompanied in lead-in entries for synonyms or quasi-synonyms” (ISO, 2011, p. 12). The example to be presented further ahead, MeSH (Medical Subject Headings), is used for indexing biomedical literature. A **clinical terminology** is defined as a “terminology required directly or indirectly to describe health conditions and healthcare activities” (ISO 17115:2007, p. 6), and is used mostly to support the representation of detailed clinical information in computerized EHRs, as is the case of SNOMED CT. Within the scope of Knowledge Engineering (KE), and as seen in the previous chapter, an **ontology** is regarded as a “formal, explicit specification of a shared conceptualization” (Studer et al., 1998). The example to be explored here is that of the Disease Ontology, which strives to provide a conceptual framework for human disease. Finally, the Unified Medical Language System (UMLS) will be addressed as a **repository** which aggregates clinical information from different resources and merges that information in a systematic and computer-processable way.

As will be shown in the following pages, one of the distinctive features among the stated resources refers to their various notions of concept, term, and relationship often represented through several – and completely unrelated – graphical notations. Despite the wide variability in this respect, it was decided that the results within each resource should be presented according to its own particular graphical notation features. However, and within the particular scope of EndoTerm, a more systematized approach concerning these graphical representations will be followed, in line with what has been proposed by (Roche, 2015). Thus, and as regards EndoTerm, concepts are

⁵² The ICHI and other procedure terminological systems played a critical role in the development of EndoTerm’s conceptualization, as will be shown in Chapter 6.

capitalized and written between single chevrons (hence the example of <Endometriosis> in the title of the current section), whereas terms are presented in lower case and between double quotation marks (e.g. “endometriosis”). Conceptual relations are represented through lower case, bold, and underscores between each element (e.g. **is_a**), while lexical semantic relations differ from the former by not resorting to the underscore (e.g. **is a**, in case this appears in a specialized text).

The order in which the following terminological resources will be described further below reflects an increasingly formalized concept-based representation that, in turn, can facilitate cross-resource mapping and the ambitioned interoperability. The key aspect that must be taken into consideration, however, is that all these resources serve different purposes or “intended objectives” (cf. Cimino, 1998; Duclos et al., 2014), which are reflected in their core structure, epistemological principles, and even graphical notations, as mentioned above.

3.2.3.1. The International Classification of Diseases (ICD)

Often regarded as the “standard” statistical classification of diseases, the ICD allows the reporting, comparison and monitoring of the incidence and prevalence of diseases and related conditions across institutions, countries, and different time periods⁵³. Endorsed and maintained by the World Health Organization (WHO) since 1948, this classification evolved from Jacques Bertillon’s *International List of Causes of Death* (cf. Section 3.2.1.) to include not only mortality but also morbidity data. It is used in more than 100 countries nowadays⁵⁴ as a Reference Classification, i.e. one that encompasses “the main parameters of the health system” (e.g. death, disease, functioning, disability, and health interventions) and that, due to its broad acceptance, is often “approved and recommended as a guideline” (WHO, 2016c, pp. 5–6)⁵⁵. Given that the ICD often does not serve all the needs of its various users, particularly in some medical specialties, there are also Derived Classifications that provide additional detail

⁵³ Cf. <http://www.who.int/classifications/icd/en/>

⁵⁴ It has also been translated into 43 languages, and Portuguese (in this case, Brazilian Portuguese), is one of them.

⁵⁵ At the moment, the WHO has, in addition to the ICD and to the upcoming International Classification of Health Interventions (ICHI), another Reference Classification – the International Classification of Functioning, Disability and Health (ICF). The three comprise the WHO’s International Family of Classifications.

to the core resource, such as the ICD for Oncology (ICD-O), the Application of the ICD to Neurology (ICD-10-NA), or to Dentistry and Stomatology (ICD-10-DA), just to name a few. At the heart of this terminological system are, on the one hand, a variable-axis structure supported by William Farr's categorization proposal (cf. 3.2.1.) and, on the other hand, a coding system which facilitates data storage, retrieval and analysis. The current version of the classification (the 10th), which has been in use since 1994, is supported by an alphanumeric code and provides a disease arrangement into mutually exclusive categories, which are, in turn, divided into 22 chapters, listed below:

Chapter	Title
I	Certain infectious and parasitic diseases
II	Neoplasms
III	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
IV	Endocrine, nutritional and metabolic diseases
V	Mental and behavioural disorders
VI	Diseases of the nervous system
VII	Diseases of the eye and adnexa
VIII	Diseases of the ear and mastoid process
IX	Diseases of the circulatory system
X	Diseases of the respiratory system
XI	Diseases of the digestive system
XII	Diseases of the skin and subcutaneous tissue
XIII	Diseases of the musculoskeletal system and connective tissue
XIV	Diseases of the genitourinary system
XV	Pregnancy, childbirth and the puerperium
XVI	Certain conditions originating in the perinatal period
XVII	Congenital malformations, deformations and chromosomal abnormalities
XVIII	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
XIX	Injury, poisoning and certain other consequences of external causes
XX	External causes of morbidity and mortality
XXI	Factors influencing health status and contact with health services
XXII	Codes for special purposes

Table 5 – Core chapters of the ICD-10

These chapters are then further subdivided into homogeneous blocks of three-character categories that aim to reflect the different axes of classification. **Figure 11** shows the results obtained from the ICD-10 browser for <Endometriosis>⁵⁶, which will be used to observe and analyze the resource's core structure.

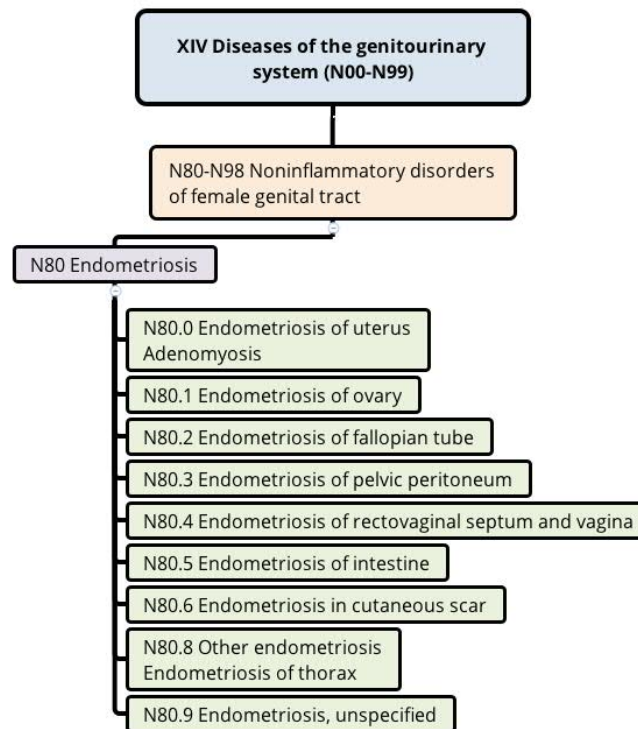


Figure 11 – ICD-10 results for <Endometriosis>

As can be seen above, the ICD-10 follows a hierarchical structure, progressing from generic to specific in the sense that N80 Endometriosis belongs to the category of N80-N98 Noninflammatory disorders of female genital tract, which, in turn, integrates the broader category of XIV Diseases of the Genitourinary System (N00-N99). This part of the classification seems to depict both etiology- and topology-based axes.

Concerning the aforementioned alphanumeric code, it is possible to identify the first character as a letter, associated with a particular chapter, and the remaining characters as numbers. In this case, the 3-character category refers to a single condition

⁵⁶ The ICD-10 browser is available at: <http://apps.who.int/classifications/icd10/browse/2016/en>

(N80. Endometriosis), with the fourth character reflecting a topological axis of the classification, i.e. specifying the different anatomical locations of the disease (e.g. N80.1 Endometriosis of ovary). This latter character, following a decimal point, represents an additional level of detail in the classification that can be included if necessary. It should be noted that the .8 and .9 subcategories – N80.8 Other endometriosis and N80.9 Endometriosis, unspecified – represent “other conditions belonging to the 3-character category” and “the same meaning as the 3-character category title, without adding any further information”, respectively (WHO, 2016c, pp. 17–18). Unlike Cimino's proposal (Cimino, 1998), the ICD's alphanumeric code contains ‘semantic’ information about the concepts it refers to, i.e. it points towards the concept's position within the hierarchy, making it relatively easy for a human user – but certainly more challenging for a machine – to identify the concept in the classification, as well as its respective supertypes and subtypes (e.g. N80.1 Endometriosis of ovary is a subtype of the supertype N80. Endometriosis).

Therefore, to maintain its key role among the current healthcare terminological resources and be integrated into the interoperability revolution that has been taking this field by storm, the ICD has been undergoing a substantial revision which will be released in 2018. Among other changes, ICD-11 will allow polyhierarchy – unlike the current version –, post-coordination, as well as consistent mapping to and into SNOMED CT⁵⁷.

3.2.3.2. The Medical Subject Headings (MeSH)

MeSH is the U.S. National Library of Medicine's (NLM) controlled vocabulary thesaurus and is used by the NLM for indexing and cataloguing biomedical articles for the MEDLINE®/PubMed® database in a uniform and consistent way, while also playing an important role in supporting information retrieval. MeSH was created in 1960, with the adoption of a single subject headings authority list for both books and periodical

⁵⁷ All the information regarding the ongoing ICD-11 Revision, including links to the new Coding Tool and the browser's beta version, can be found at <http://www.who.int/classifications/icd/revision/en/>.

articles and reviewed three years later to include a set of main categories and subcategories, many of which still exist today⁵⁸.

Since the early 2000s, MeSH has undergone significant changes to meet the needs of a rapidly and constantly evolving field. One of the most noteworthy is related to its transition from a term-based into a concept-oriented structure, with the concept taking on a central role. Currently, MeSH has a three-level structure supported by [1] descriptors, [2] concepts, and [3] terms.

Descriptors are considered the unit of indexing and retrieval at MeSH. They are defined as “classes of concepts” and are organized into a numbered tree structure that allows users to browse from general to more specific topics. They may provide information concerning the subject of an indexed item [Main Headings], its genre [Publication Characteristics/Types], or the geographic (or physical) location of the conducted study, for instance [Geographics]. Hence, MeSH’s core tree structure⁵⁹ comprises 16 categories (cf. **Table 6**), and each is further subdivided into subcategories, in which descriptors are arranged hierarchically, in up to thirteen different levels. The aforementioned tree structures can be seen as a compromise rather than an exhaustive or universally accepted classification: their primary goal is to serve the needs and requirements of the various disciplines and users. Each descriptor is followed by an alphanumeric string whose purpose is to indicate the descriptor’s location within each tree. An example concerning <Endometriosis> will be provided below.

Categories
Anatomy [A]
Organisms [B]
Diseases [C]
Chemicals and Drugs [D]
Analytical, Diagnostic and Therapeutic Techniques, and Equipment [E]
Psychiatry and Psychology [F]
Phenomena and Processes [G]

⁵⁸ More information can be found under <https://www.nlm.nih.gov/mesh/introduction.html>.

⁵⁹ <https://meshb-prev.nlm.nih.gov/treeView>

Disciplines and Occupations [H]
Anthropology, Education, Sociology, and Social Phenomena [I]
Technology, Industry, and Agriculture [J]
Humanities [K]
Information Science [L]
Named Groups [M]
Health Care [N]
Publication Characteristics [V]
Geographicals [Z]

Table 6 – Main hierarchy in MeSH

The category “Publication Characteristics [V]”, or Publication Types, will be further explored in Chapter 6 since it was used as the foundation of the text typology that supported the proposed conceptualization.

In MeSH, concepts are “common meanings shared by synonymous terms”⁶⁰. Each concept can be designated by one or more terms, and with the current restructuring of this resource, each concept is now explicitly represented by a Unique Identifier. The final element of this 3-tiered structure is the term, which designates a concept. The set of terms that designate the same concept are considered synonymous, and each term is also labelled with a Unique ID (Lipscomb, 2000).

As can be seen from the following examples, which are based on <Endometriosis>, MeSH comprises both relationships between descriptors and those within descriptors. The former, substantiated as Broader Than / Narrower Than relationships, constitute the basis of the MeSH hierarchies (**Fig. 12**), whereas the latter correspond to the 3-level structure mentioned earlier (Descriptor > Concept > Term)

⁶⁰ https://www.nlm.nih.gov/mesh/concept_structure.html

(Table 7)⁶¹. In contrast to what occurs in the ICD, there are no subtypes of <Endometriosis> in the MeSH tree structure.

It is also important to highlight that even though descriptors may appear in more than one place in the MeSH trees, the same does not occur with concepts, i.e. MeSH allows polyhierarchy in the arrangement of the descriptors but NOT of the concepts (Nelson, Johnston, & Humphreys, 2001).

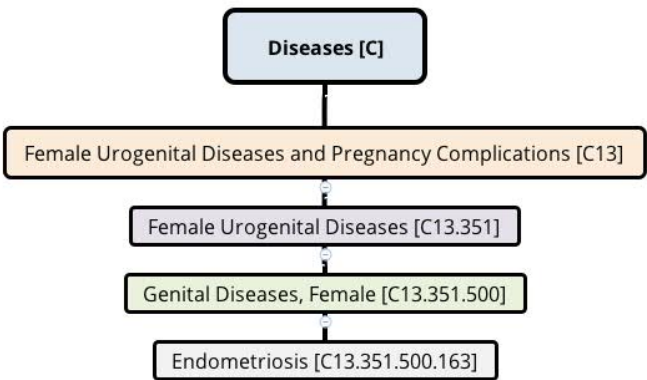


Figure 12 – MeSH tree structure for <Endometriosis>

MeSH Heading	Endometriosis
Tree Number	C13.351.500.163
Unique ID	D004715
Preferred Concept	Endometriosis
Concept UI	M0007413
Scope Note	"A condition in which functional endometrial tissue is present outside the uterus. It is often confined to the pelvis involving the ovary, the ligaments, cul-de-sac, and the uterovesical peritoneum"
Preferred Term	Endometriosis
Term UI	T014432
Date	01/01/1999

⁶¹ Both the visual representation in **Figure 12** and the table were based on information available at the MeSH Browser: <https://meshb.nlm.nih.gov/search>

Related Concept	Endometrioma
Concept UI	M0406918
Scope Note	"An enlarged area of Endometriosis that resembles a tumor. It is usually found in the ovary. When it is filled with old blood, it is known as a chocolate cyst"
Preferred Term	Endometrioma
Term UI	T472933
Date	12/10/2001

Table 7 – MeSH results for <Endometriosis>: metadata

The example above shows that the name of the descriptor is the Preferred Term of the Preferred Concept. There is just one Preferred Concept for each MeSH record and, in turn, each concept has one Preferred Term (and may have coexisting terms designating the same concept – synonyms – which is not the case here). Only concepts have scope notes. It can also be seen from the example that the Descriptor ID referred to before [D004715, in this case] corresponds neither to the tree number [C13.351.500.163] – used simply for location purposes – nor to the Concept Unique Identifiers [M0007413 and M0406918] or to the Term Unique Identifiers [T014432 and T472933], respectively.

Finally, it also becomes clear that MeSH’s hierarchical trees are not and should not be used as a concept representation language, since they serve mainly indexing, cataloguing, and retrieval purposes and are structured as such, containing “other needs” that are not confined to subsumption (Nelson et al., 2001).

3.2.3.3. SNOMED CT

SNOMED CT is defined as a comprehensive and multilingual clinical healthcare terminology that, in the short term, aims to become the most widely used biomedical terminological resource at an international level. Due to its Description Logic basis, SNOMED CT supports the representation of clinical content in Electronic Health Records in a consistent, comparable, reliable and computer-readable way (IHTSDO, 2018d).

As can be seen from the image below (**Figure 13**)⁶², SNOMED CT's inception occurred in 2002, and it evolved from two existing resources: the Systematized Nomenclature of Medicine (SNOMED) and the Clinical Terms - Version 3. It is currently owned and administered by SNOMED International, a trading name of the International Health Terminology Standards Development Organisation (IHTSDO). Very recently, namely on April 1, 2018, SNOMED CT has fully replaced the Read Codes within the National Health Service's Primary Care System and started being adopted by all UK General Practitioners⁶³.

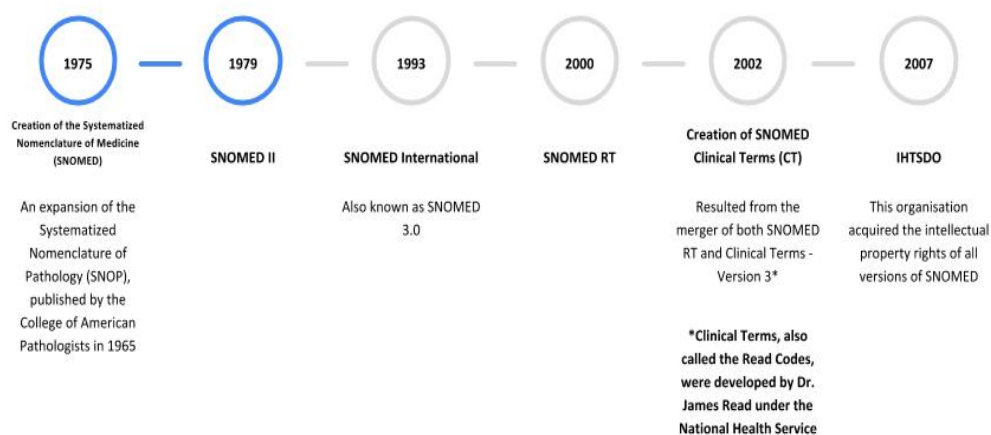


Figure 13 – Historical overview of SNOMED CT

SNOMED CT is also a concept-oriented terminological resource, following practically all of Cimino's Desiderata (1998). In fact, in addition to these principles and recommendations, new content in SNOMED CT must follow a URU approach, i.e. it must be understandable to people who may not be subject field experts, reproducible in several platforms and contexts, and useful to the various healthcare stakeholders (IHTSDO, 2018d, p. 44). This is also a multilingual resource, with its International Edition, for instance, being available in both English and Spanish, and including additional Local Extensions (in national languages) for Denmark, Sweden, the Netherlands, Canada, Australia, and Uruguay. As far as maintenance is concerned, SNOMED CT is regularly

⁶² Based on: <https://www.snomed.org/snomed-ct/what-is-snomed-ct/history-of-snomed-ct>

⁶³ More information about this transition is available at <https://digital.nhs.uk/SNOMED-CT-implementation-in-primary-care>

updated, with a biannual release of the International Edition (January 31 and July 31) in their proprietary Release Format 2 (RF2).

At its core, SNOMED CT's logical model comprises three main components (cf. **Figure 14** below, all of which are graphically represented in initial upper case and between vertical bars, as defined by the resource's Compositional Grammar, a lightweight syntax that is both human-readable and computer-processable, thus enabling interoperability (IHTSDO, 2018a).

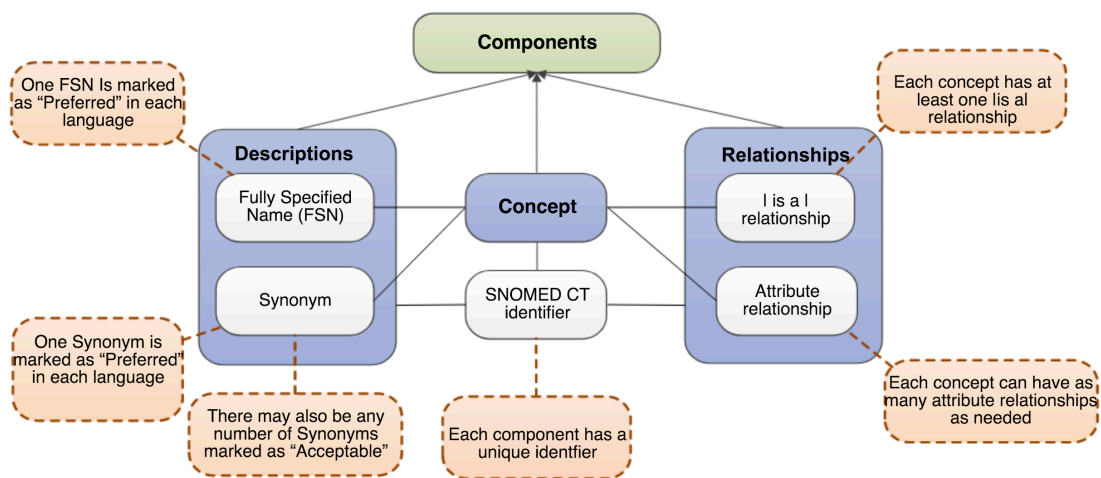


Figure 14 – SNOMED CT's Logical Model

1) Concepts represent clinical meanings in SNOMED CT. They are organized into hierarchies, ranging from general to specific, and each one of them is assigned a numeric and machine-readable concept identifier, which is non-semantic in the sense of Cimino's proposal (1998) and constitutes an unambiguous and unique reference to a particular concept.

2) Descriptions link human-readable terms to the concepts above. Therefore, one concept can have more than one associated description: in each language, one of these descriptions will be regarded as the concept's Fully Specified Name (FSN) and used for disambiguation between different concepts; and the remaining ones will be synonyms, in the sense that they describe the same clinical concept. Once more, in each language, one of these synonyms will be marked as preferred and the others as acceptable. Similarly to the concepts, each description has a unique numeric ID.

3) Finally, the relationships connect each concept to other related concepts. They also have a unique numeric relationship ID and provide the framework that underpins the concepts' logic definitions in a computer-processable way, which will be explored below. Overall, there are two main types of relationships in SNOMED CT: the |Is a| or subtype relationships and the attribute relationships. The former constitute the backbone of SNOMED CT's core hierarchies and are the most widely used within this resource. Primarily, they express that the source concept is a subtype of the destination concept, i.e. the source concept is subsumed by the target concept. To be considered as such, each concept in the system must be allocated to at least one |Is a| relationship (IHTSDO, 2018d, pp. 12–14).

As regards the attribute relationships, they contribute to further define a source concept (e.g. |Endometriosis of uterus|) by connecting it to a given value of a defining characteristic. The characteristic (attribute) is specified by the relationship itself (e.g. |Finding site|) and the value is provided by the destination concept (e.g. |Uterine structure|). It should be noted, though, that the applicability of each of these attribute relationships is constrained to a predefined domain (source concept) and range (destination concept or value), to ensure consistency and coherence (cf. the notion of categorial structure introduced in 3.2.2.). These specifications will, in turn, contribute to support logical definitions and infer additional relationships, hence enhancing data entry, decision support systems, and information retrieval (Benson & Grieve, 2016; IHTSDO, 2018c)⁶⁴.

The preceding core components can also be supplemented by Reference Sets (RefSets) and Extensions, which provide additional flexible features, enable customization and contribute to enhance the terminology by addressing different requirements, such as language preferences, value sets of concepts, maps to or from other code systems, etc. (IHTSDO, 2018d, pp. 46–48).

⁶⁴ The full set of attribute relationships available in SNOMED CT, as well as their respective inverse counterparts, are available at <https://bit.ly/2ke6Z90>, and played a key role in devising the conceptualization underlying the EndoTerm project.

SNOMED CT's Top Hierarchy includes 19 Top-Level Concepts, which represent entities as diverse as clinical findings, events, procedures, substances, organisms or pharmaceutical/biologic products, for example. **Figure 15**⁶⁵ depicts this core hierarchy:



Figure 15 – SNOMED CT's Top-Level Hierarchy

The table below (**Table 8**) builds up on the preceding hierarchy by providing a scope-note-like description for each of these Top-Level Concepts, as well as concept examples for each category.

Top Level Concept	Brief description	Examples
Body structure (Body structure)	represents normal and abnormal anatomical structures.	Mitral valve structure
Clinical finding (finding)	represents the result of a clinical observation, assessment or judgment and includes normal and abnormal clinical states. It includes concepts used to represent diagnoses.	Asthma Headache
Environment or geographical location (environment/location)	represents types of environments as well as named locations such as countries, states and regions	Intensive care unit
Event (event)	represents occurrences excluding procedures and interventions.	Flood Earthquake
Observable entity (observable entity)	represents a question or assessment which can produce an answer or result.	Systolic blood pressure

⁶⁵ Source: The SNOMED CT Browser - <http://browser.ihtsdotools.org/>

Organism (organism)	represents organisms of significance in human and animal medicine.	Streptococcus pyogenes
Pharmaceutical/biologic product (product)	represents drug products.	Amoxicilin 250mg capsule
Physical force (physical force)	represents physical forces that can play a role as mechanisms of injury	Friction
Physical object (physical object)	represents natural and man-made physical objects	Implant device
Procedure (procedure)	represents activities performed in the provision of health care. This includes not only invasive procedures but also administration of medicines, imaging, education, therapies and administrative procedures.	Appendectomy Physiotherapy
Qualifier value (qualifier value)	represents the values for some SNOMED CT attributes, where those values are not subtypes of other top-level concepts	Left Severe
Record artifact (record artifact)	represents content created for the purpose of providing other people with information about record events or states of affairs.	Record entry Family history section
Situation with explicit context (situation)	represents concepts in which the clinical context is specified as part of the definition of the concept itself. These include presence or absence of a condition, whether a clinical finding is current, in the past, or relates to someone other than the subject of the record.	Endoscopy arranged Family history of glaucoma
SNOMED CT Model Component (metadata)	contains technical metadata supporting the SNOMED CT release.	
Social context (social concept)	represents social conditions and circumstances significant to health care.	Occupation
Special concept (special concept)	represents concepts that do not play a part in the formal logic of the concept model of the terminology, but which may be useful for specific use cases.	Alternative medicine poisoning
Specimen (specimen)	represents entities that are obtained (usually from the patient) for examination or analysis.	Urine specimen
Staging and scales (staging scale)	represents assessment scales and tumor staging systems	FIGO staging system
Substance (substance)	represents general substances, the chemical constituents of pharmaceutical/biologic product, body substances, dietary substances and diagnostic substances.	Methane Insulin

Table 8 – Overview of SNOMED CT's Top-Level Concepts

In the following paragraphs, SNOMED CT's results concerning <Endometriosis> will be used to illustrate two core features of this terminological resource: the fact that it is grounded in formal concept definitions, and that it allows pre- and post-coordination (cf. Section 3.2.2.).

As previously mentioned, both the |Is a| and the attribute relationships in SNOMED CT provide the foundations of a concept's formal definition from a logical perspective, in the sense that each concept's defining characteristics are explicitly put forward according to a set of rules and constraints (cf. the notion of categorial structure

explored in Section 3.2.2.). As an example, **Figure 16** shows the definition of <Endometriosis> by resorting to a diagrammatic notation used within the SNOMED CT environment. This diagram-based approach aims to provide consistent concept representations, thereby reducing ambiguity and potential misunderstandings (IHTSDO, 2018b).

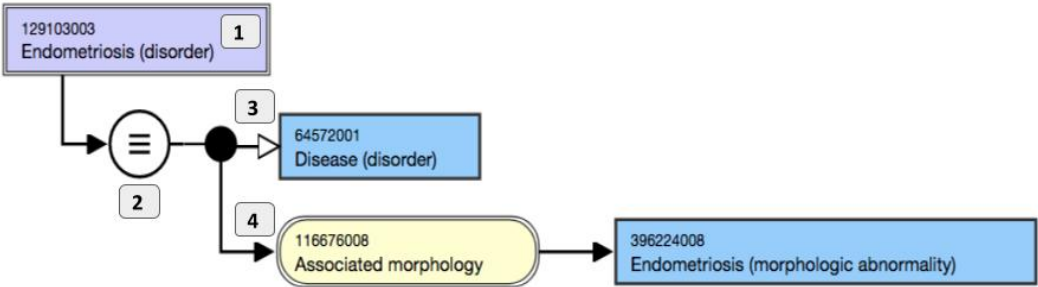


Figure 16 – <Endometriosis> in SNOMED CT

Number [1] refers to the concept being defined and, in this case, the rectangle depicts not only the SNOMED CT unique ID [129103003] but also the concept’s Fully Specified Name. The latter indicates that <Endometriosis> belongs to the supertype concept <Disorder> that, in turn, is subsumed by a SNOMED CT Top Level Concept, as will be seen shortly. The fact that Endometriosis is represented in purple means that it is perceived by the SNOMED CT Content Development Team as a Fully Defined concept, i.e. its defining characteristics are sufficient to distinguish its meaning from other similar concepts. This ‘Fully Defined’ status is also conveyed by the symbol represented in [2]. The first (and main) relationship underlying this concept definition is a subtype relationship [3], which states that |Endometriosis (disorder)| |Is a| |Disease (disorder)|. This second concept, as well as |Endometriosis (morphological abnormality)|, are displayed in blue, which indicates that they are Primitive concepts, in the sense that their defining characteristics are not sufficient to uniquely distinguish them from other similar concepts (IHTSDO, 2018c). The 4th element of the diagram is an **attribute** relationship – |Endometriosis (disorder)| |has Associated Morphology| |Endometriosis (morphologic abnormality)| –, which plays a key role in further specifying the definition and thus contributing to the Fully Defined status of the concept under analysis. The following table (**Table 9**) contains other concept details, namely the

terms that designate that concept and their acceptability status, i.e. the FSN and two synonyms for US English, the Preferred and the Accepted, respectively.

SCTID	129103003
	Fully Defined Concept
FSN	Endometriosis (disorder)
SYN (Pref.)	Endometriosis (clinical)
SYN (Accepted)	Endometriosis

Table 9 – SNOMED-CT’s information on the descriptions associated to <Endometriosis>

As regards the remaining subtype relationships involving |Endometriosis (disorder)|, SNOMED CT identifies 15 children altogether, which, in turn, are subdivided into a total of 42, mainly according to the disease **location** (e.g. |Endometriosis of intestine (disorder)|), although a few examples are grounded in disease **etiology** (e.g. |Endometriosis of the spermatic cord following estrogen therapy (disorder)|) and disease **severity** (e.g. |Deep endometriosis|). When analyzing the profusion of subtype concepts (available in **Annex 1**), two relevant points should be considered: firstly, the confirmation that SNOMED CT supports polyhierarchy (cf. |Endometriosis of rectovaginal septum and vagina (disorder)| |Is a| |Endometriosis of pelvis (disorder)| **and** |Endometriosis of vagina (disorder)|), and secondly, the fact that the creation of Fully Specified Names in US English does not occur randomly, but follows a set of patterns (example for a topology-based axis of analysis: “disease” + “of” + “location”) summarized in the Editorial Guide (IHTSDO, 2018c).

It should be noted that the preceding diagrammatic notation is particularly useful for human users, but not for machines, which is why SNOMED CT resorts to an additional way to represent concepts and their respective meanings via the aforementioned Compositional Grammar. The following figure represents the same clinical content as the previous diagram but resorting to SNOMED CT’s compositional grammar instead (**Figure 17**):

```

129103003 |Endometriosis (disorder)|
== 64572001 |Disease (disorder)| :
    116676008 |Associated morphology| = 396224008 |Endometriosis (morphologic abnormality)|

```

Figure 17 – Definition of <Endometriosis> using SNOMED CT's compositional grammar

As can be seen from the example, the first concept right after the one being defined represents the |Is a| relationship and is followed by a refinement (colon), which subsequently introduces the attribute relationship. The attribute is separated of the value (destination concept) by an equals sign. The inclusion of the unique concept IDs ensures machine-readability if needed, yet this representation is still readable and understandable by a human user.

As stated above, one of SNOMED CT's key features - and which contributes to its solid, expressive power and rising popularity - is the fact that it allows pre- and post-coordination. Hence, two different ways of representing the same concept can coexist: [1] the pre-coordinated expression, representing the meaning of a given concept, with a unique concept ID; and [2] the post-coordinated expression, which represents the meaning of that concept through the combination of two or more existing concepts, together with their respective identifiers and defining relationships. Both pre- and post-coordination can be displayed in SNOMED CT through the previously referred diagramming notations or via the compositional grammar. However, this compositional approach must take the predefined domain-range constraints into account, to avoid nonsense representations. All of this can be seen in **Figure 18** below, which depicts a surgical procedure that can be used within the context of Endometriosis: a laparoscopic radical hysterectomy, i.e. a surgical procedure resorting to a laparoscope and which aims at the total excision of the uterus, as well as the cervix, and part of the vagina. In this procedure, the so-called uterine *adnexae* structures are also removed⁶⁶.

⁶⁶ <https://www.cancer.gov/publications/dictionaries/cancer-terms?cdrid=322879>

Precoordinated expression

```
236891001 |Laparoscopic radical hysterectomy (procedure)|
```

Postcoordinated expression

```
== 116142003 |Radical hysterectomy (procedure)| :  
  { 405813007 |Procedure site - Direct| = 181452004 |Entire uterus (body structure)|,  
    260686004 |Method| = 129304002 |Excision - action (qualifier value)|,  
    425391005 |Using access device| = 86174004 |Laparoscope, device (physical object)| }  
  { 260686004 |Method| = 129304002 |Excision - action (qualifier value)|,  
    425391005 |Using access device| = 86174004 |Laparoscope, device (physical object)|,  
    405813007 |Procedure site - Direct| = 23043003 |Uterine adnexae structure (body structure)| }
```

Figure 18 – Definition of <Laparoscopic radical hysterectomy> via pre- and post-coordinated expressions using SNOMED CT's compositional grammar

When compared to the prior example that depicted SNOMED CT's compositional grammar, this post-coordinated expression goes one step further by including grouping of attribute relationships within a refinement, hence the curly braces. For example, the first group indicates that a given method (|Excision - action (qualifier value)|) applies directly to a specific procedure site (|Entire uterus (body structure)|), using a specific access device (|Laparoscope, device (physical object)|). The various attribute-value pairs are separated by commas.

To conclude this subsection, it is also important to emphasize that SNOMED CT's most recent efforts towards becoming the most widely used resource within the context of EHR content creation and sharing have led to an increasingly solid cooperation between SNOMED International and other organizations. Since 2013, the LOINC⁶⁷ codes have been mapped to SNOMED CT, and two years later, an additional mapping process was launched, this time from SNOMED CT to ICD-10⁶⁸. More recently, studies concerning

⁶⁷ LOINC stands for Logical Observation Identifiers Names and Codes and is defined as a common language (set of identifiers, names, and codes) used to identify health measurements, observations, and documents. Ultimately, LOINC's goal is to be able to devise different codes "for each test, measurement, or observation that has a clinically different meaning" (<https://loinc.org/>). Additional information on the agreement between SNOMED International and the Regenstrief Institute (responsible for LOINC) is available at <https://www.snomed.org/snomed-ct/mapping-to-other-terminologies/loinc>.

⁶⁸ Cf. https://www.nlm.nih.gov/research/umls/mapping_projects/snomedct_to_icd10cm.html;

a deeper integration between SNOMED CT and ICD-11 have also been published (J.-M. Rodrigues et al., 2013, 2014, 2015).

3.2.3.4. The Disease Ontology (DO)

Due to a growing need for a standardized representation of human disease that would allow, among other things, disease concepts to be mapped across various resources, the Disease Ontology was created in 2004 and is still currently developed by the Center for Genetic Medicine of Northwestern University and the Institute for Genome Sciences, from the School of Medicine of the University of Maryland.

This is an open source, community-driven ontology which aims to provide a disease-oriented framework that supports the creation of clear and semantically consistent definitions, thus enabling interoperability among other biomedical resources (Schriml et al., 2012). It is one of the few resources that is explicitly organized around the concept of ‘human disease’, especially disease etiology, although disease topology/location can also be used as a criterion when etiology is unknown. In order to substantiate the pivotal role of ‘disease’, the DO has adopted the Ontology of General Medical Science’s (OGMS) ontological definition for it: “A disposition (i) to undergo pathological processes that (ii) exists in an organism because of one or more disorders in that organism” (Scheuermann, Ceusters, & Smith, 2009, p. 118). Despite what could be easily perceived as a clear-cut concept orientation, the core elements within the resource are always called “terms” rather than concepts (Kibbe et al., 2015; Schriml et al., 2012; Schriml & Mitraka, 2015). Therefore, the use of “term” in this context reflects the theoretical approach followed by those responsible for the resource.

Starting with the DO’s core structure, depicted in **Figure 19**, it can be seen that it comprises the major types of disease arranged in eight categories that are mainly etiology-based, although the second (disease of anatomical entity) is topology-oriented, as it will focus on the affected body systems:

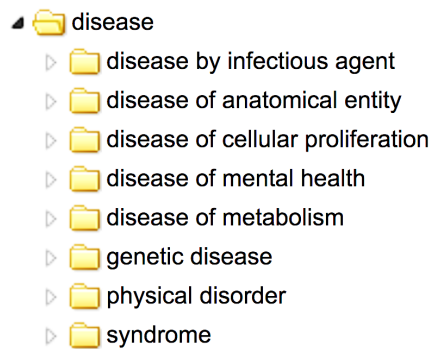


Figure 19 – The Disease Ontology’s core hierarchy

This resource can be consulted and explored via a web browser divided into three main panels (cf. **Figure 20** below):

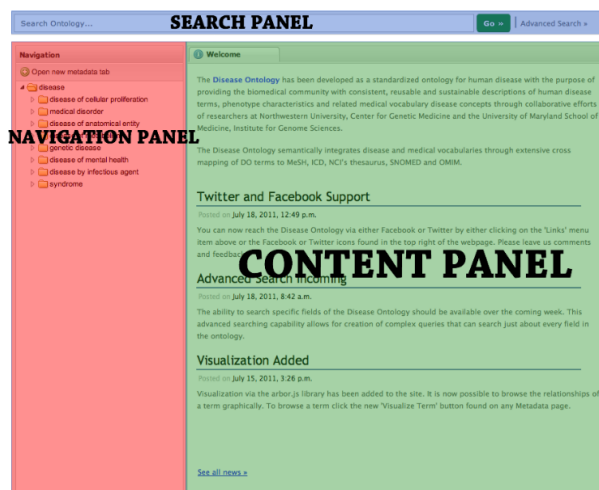


Figure 20 – The Disease Ontology’s web browser⁶⁹

The Search Panel provides all the necessary tools to conduct basic or advanced queries on the Disease Ontology; the Navigation Panel includes an interactive hierarchy-based ‘tree’, built upon subsumption or is_a relationships, and which enables the user to navigate and explore the different sub-hierarchies; finally, the Content Panel displays search results, metadata, as well as graphic visualizations of both searched disease(s) and respective relationships on a single page with multiple tabs (Schriml et al., 2012). Both the concepts and their respective relationships are written in lower case, with no additional graphical features besides an underscore separating the two elements in the core is_a.

⁶⁹ Available at <http://disease-ontology.org/>.

Following the same approach as in previous terminological systems, the results for <Endometriosis> will be used to illustrate the DO's main features. In the following image (**Figure 21**), one can observe, on the left, the hierarchy leading to our key concept, which is mainly topology-based: endometriosis is_a female reproductive system disease, which in turn, is_a reproductive system disease, is_a disease of anatomical entity, and, finally, is_a disease.

The Metadata Panel is then displayed on the right and comprises a Disease Ontology numeric and unique ID (DOID), the so-called 'Name' of the concept (endometriosis), a definition in natural language which relies partially on the preceding hierarchical information (in this case, it includes the parent concept, which is also listed in the metadata), and the several cross-references (XRefs), i.e. links to synonymous disease concepts in MeSH, ICD-9 and -10-CM, SNOMED CT, the National Cancer Institute (NCI), and the Experimental Factor Ontology (EFO), based on the Unified Medical Language System's (UMLS) Concept Unique Identifiers (CUI), which will be thoroughly described in the following subsection (Kibbe et al., 2015; Schriml et al., 2012).

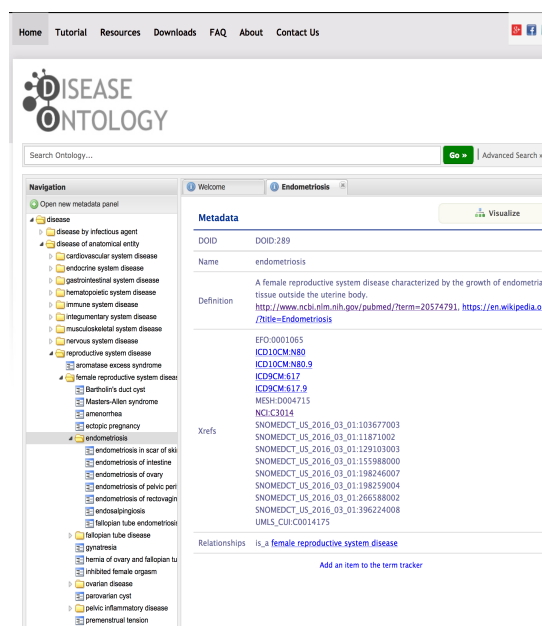


Figure 21 – The Disease Ontology's metadata for <Endometriosis>

When selecting the 'Visualize' option (top right-hand corner), a new tab opens in the Metadata Panel, displaying an interactive canvas which includes: a) the target node - in red; b) the parent node(s) - in green; and c) the leaf nodes - in grey. In nodes with five or more children, the number of children is represented by a golden circle

which can then be selected and expanded, hence providing the user with a visual representation, as depicted in **Figure 22**:



Figure 22 – The Disease Ontology’s visualization canvas for <Endometriosis>

As can be seen from the image above, the subtypes of Endometriosis indicate mainly a topology-oriented conceptualization, which is in line with the results obtained so far. Furthermore, it is important to point out that in the last few years, the DO has aimed to further explore the distinctions between <Disease>, <Disorder>, and <Phenotype> proposed by the OGMS, as well as the complex relationships between them. Consequently, a set of supplementary relations have been added to the core ontology (cf. **Table 10**). As mentioned earlier when describing SNOMED CT’s attribute relationships, these most recent relationships included in the DO were also taken into account when devising the conceptualization supporting the EndoTerm project.

Disease Ontology		
complicated_by	inheres_in	realized_by
composed_of	is_a	realized_by_suppression_with
derives_from	located_in	results_in
has_material_basis_in	occurs_with	results_in_formation_of
has_symptom	part_of	transmitted_by

Table 10 – The Disease Ontology’s relationships

Between 2012 and 2015, the DO has undergone several enhancements: in addition to the regular content updates, there has been an increase in the number of textual definitions, a stronger presence of genetic and rare diseases, better and more consistent mapping procedures (especially involving the UMLS, SNOMED CT, and the upcoming ICD-11), as well as an extended usage of the DO as a disease domain ontology in BioPortal (Kibbe et al., 2015). As one of the main developments in the short term, Kibbe et al. (*ibid.*) highlight the “imperative need to provide textual definitions for all DO terms”, something which will be further explored in the following chapter.

3.2.3.5. The Unified Medical Language System (UMLS)

Regarded by (Coiera, 2015, p. 394) as something like “the Rosetta Stone” of international terminologies, the UMLS was developed and is still maintained and distributed by the U.S. National Library of Medicine (NLM), with the purpose of helping computer programs “understand” the meaning behind user inquiries, thus being able to support and enhance the retrieval of machine-readable biomedical information (Humphreys & Lindberg, 1993; Lindberg, Humphreys, & McCray, 1993).

In fact, the UMLS tries to address two significant hurdles concerning the effective management, retrieval and sharing of biomedical information: on the one hand, the wide variety of ways in which the same concept can be expressed and, on the other hand, the lack of a standard format for distributing terminologies (Bodenreider, 2004). In short, the UMLS aims to provide some sort of integrated access to various biomedical TS, most of which vary in nature, scope, size, and purpose.

This aggregating resource was first created in 1986 and involved a multidisciplinary team at the NML, which included experts in medical librarianship, medical informatics, computer science, medicine, linguistics, among others. The first few years (1986-1988) were spent assessing needs and defining the necessary components for the resource. In the second stage of the project (1988-1991), the initial versions of these components were developed and distributed, with subsequent updates in 1991 and 1992. Until the mid-1990s, other applications were developed, namely in the fields of information retrieval, as referred to before, indexing, and natural

language processing (Humphreys & Lindberg, 1993; Lindberg et al., 1993). Nowadays, the UMLS is distributed free of charge (license agreement required) by the NLM via the UMLS Terminology Services (UTS)⁷⁰. At the core of the UMLS are the so-called Knowledge Sources - the Metathesaurus, the Semantic Network, and the SPECIALIST Lexicon - which will be further described below.

The UMLS Metathesaurus is defined as a “large, multi-purpose and multilingual vocabulary database” containing information about biomedical and other health-related concepts, their designations, as well as the relationships between them (NLM, 2009, pp. 11–12; NLM, 2016). In other words, it is a large repository that collects information from more than 100 different resources – the so-called “source vocabularies” -, such as thesauri, classifications, code sets, controlled vocabularies, clinical terminologies, ontologies.

The main principle underlying the Metathesaurus is its concept orientation, with concept being defined as a “meaning” (Bodenreider, 2004; NLM, 2009). This principle, in turn, allows the information that is present in the several “source vocabularies” (e.g. different concept names, respective relationships, etc.) to be connected around a given concept, thereby preserving that information according to the source vocabulary’s structure and specifications. This includes graphical notations, which means that, in the UMLS, coexist disparate graphical features, from concepts written in upper case, lower case, and even all-caps, to relationships with or without underscore, etc.

To facilitate implementation, each concept is assigned a unique and permanent identifier (CUI), which has no intrinsic meaning (i.e. non-semantic) and is permanent (in other words, it is never deleted or reused, regardless of changes that may occur over time)⁷¹. Concepts are then linked to other concepts via a set of relationships, both hierarchical and associative. Moreover, all the gathered information is stored - and shared - in a common format: either the Rich Release Format (RRF) (standard format) or the Original Release Format (ORF).

⁷⁰ <https://uts.nlm.nih.gov/home.html>

⁷¹ Again, in line with Cimino’s Desiderata (1998, 2006).

Despite this ability to congregate a wide range of biomedical information, the ultimate purpose of the Metathesaurus is not to become a single, standard-like resource covering the whole subject field of Medicine. Its goal, as referred to in its reference manual, is rather to preserve “the many views of the world present in its source vocabularies because these views may be useful for different tasks” (NML, 2009, p. 12).

As has been done before, the following figures contain examples regarding the concept of <Endometriosis> which will allow further clarification of the Metathesaurus and its respective framework.

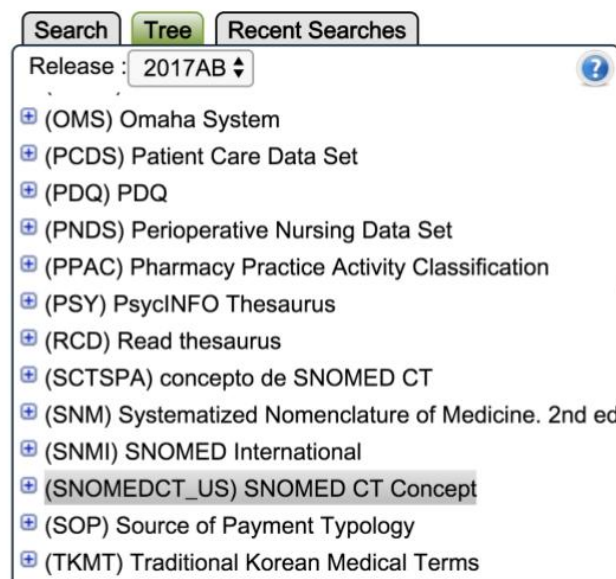


Figure 23 – The UMLS Metathesaurus tree (partial view)

Figure 23 shows part of the Metathesaurus tree, comprising the several source vocabularies. One of them, which will be used in our example below, refers to SNOMED CT (US Edition). The user can, therefore, either explore the tool via the tree, to see how a given concept or set of concepts are depicted in a given source vocabulary; or use the search bar to look for a specific concept. In **Figure 24**, the results for <Endometriosis> are displayed, allowing the user to know about the attributed CUI, the definitions in natural language that can be found in some of the source vocabularies (identified by the

respective abbreviated form), as well as the information that this concept belongs to the Semantic Type⁷² Disease or Syndrome [T047].

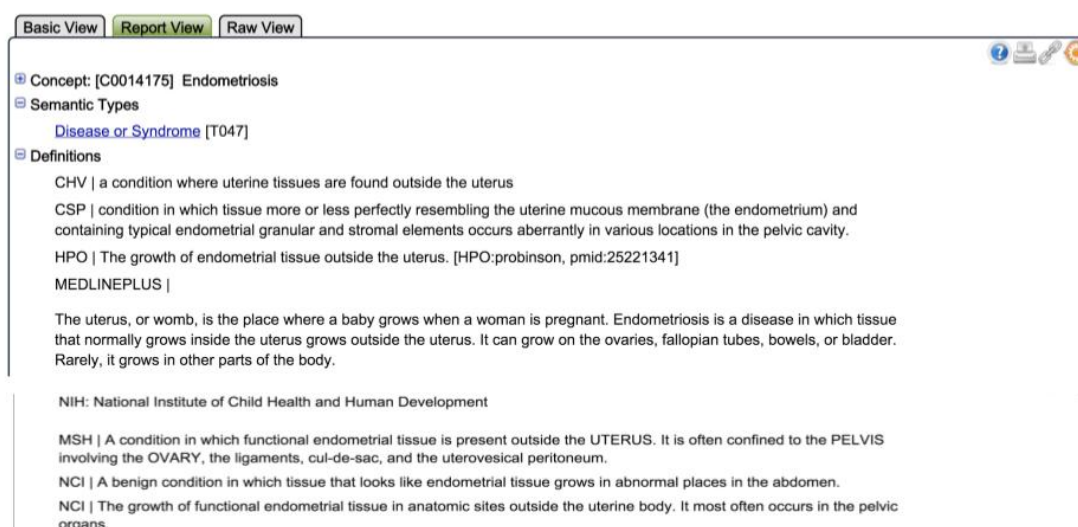


Figure 24 – The UMLS Metathesaurus: results for <Endometriosis>

As far as the core structure of the Metathesaurus is concerned, it should also be noted that there are four main types of subunits: the Concepts, the Terms, the Strings, and the Atoms. **Table 11**, compiled from information from the “Raw View” of the Metathesaurus Browser in RRF format⁷³, makes it possible to undertake a more in-depth analysis of each one of these subunits. The central unit - the concept - is identified by its respective CUI. Each concept can then be expressed in a certain way in a given natural language, i.e. via different terms. Also, the Metathesaurus structure can accommodate translations of its source vocabularies into languages other than English. A Lexical Unique Identifier (LUI) (light green lines) is therefore assigned to every term. As can be seen below, the first LUI refers to the term “endometriosis” in English, the second to the same term, but in Spanish, whereas the latter refers to the English term “endometriosis (disorder)” (due to the disparities among the several source vocabularies, EndoTerm’s approach to graphical notation is being used here). Within each LUI, there are the

⁷² As will be seen below, the Semantic Network is one of the remaining UMLS’s Knowledge Sources. It includes Semantic Types, which aim to provide “consistent categorisation of all the concepts represented in the Metathesaurus” (NLM, 2009, pp. 89-90). All concepts in the Metathesaurus are attached to at least one Semantic Type.

⁷³ https://uts.nlm.nih.gov/metathesaurus.html#C0014175;0;1;CUI;2017AB;EXACT_MATCH;CUI;*;

strings, identified by a String Unique Identifier (SUI) (light orange lines). These illustrate any changes concerning upper-lower case, character set, or punctuation. Finally, the concrete realization of each string in a given source vocabulary represents an atom, perceived as the basic building block upon which the Metathesaurus is built, and identified via an Atom Unique Identifier (AUI) (light grey lines).

[C0014175] Endometriosis	
[L0014175] Endometriosis [ENG]	
[S0037680] Endometriosis	
	[A0054291] Endometriosis [ICD10]
	[A7570294] Endometriosis [NCI]
	[A21238961] Endometriosis [OMIM]
[S0418288] endometriosis	
	[A18609769] endometriosis [CHV]
	[A478503] endometriosis [CSP]
[S0369090] ENDOMETRIOSIS	
	[A4008850] ENDOMETRIOSIS [WHO]
[L0342592] ENDOMETRIOSIS [SPA]	
[S0450897] ENDOMETRIOSIS	
	[A0514761] ENDOMETRIOSIS [WHOSPA]
[L3008244] Endometriosis (disorder) [ENG]	
[S3276228] Endometriosis (disorder)	
	[A3414225] Endometriosis (disorder) [SNOMEDCT_US]

Table 11 – AUI, SUI, LUI and CUI: the building blocks of the Metathesaurus⁷⁴

Within the Metathesaurus, it is also possible to have access to the concept relations from the source vocabulary. **Figure 25** depicts partial results for the concept |Endometriosis (clinical)|, from the SNOMED CT US Edition. Here, besides providing the source vocabulary's Unique Identifier [129103003], the results allow one to identify a list of supertype and subtype concepts, respectively: the former states that |Endometriosis (clinical)| |Is a| child of |Disease|, whereas the latter refers to the fact that |Endometriosis (clinical)| |Is a| parent of |Endometriosis of intestine| in this source vocabulary.

⁷⁴ The abbreviated forms for each one of the represented source vocabularies correspond to the following: **ICD10** = the International Classification of Diseases - 10th edition; **NCI** = National Cancer Institute Thesaurus; **OMIM** = Online Mendelian Inheritance in Man; **CHV** = Consumer Health Vocabulary; **CSP** = CRISP (Computer Retrieval of Information on Scientific Projects) Thesaurus; **WHO** = World Health Organization; **WHOSPA** = World Health Organization (Spanish Version); **SNOMEDCT_US** = SNOMED CT (US Edition).

Endometriosis (clinical) [A3414224/SNOMEDCT_US/PT/129103003]					
Attributes (6) Name Value RSAB					
Relations (44) REL RELA RSAB [SType1 - SType2] STypeld String CUI					
[: 1 - 10 : ➤]					
CHD	isa	SNOMEDCT_US [SCUI - SCUI]	64572001	Disease	C0012634
CHD	isa	SNOMEDCT_US [SCUI - SCUI]	362969004	Disorder of endocrine system	C0014130
CHD	isa	SNOMEDCT_US [SCUI - SCUI]	244938009	Disorder of female genital organs	C0017411
CHD	isa	SNOMEDCT_US [SCUI - SCUI]	373673007	Disorder characterized by pain	C1300028
PAR	inverse_isa	SNOMEDCT_US [SCUI - SCUI]	198259004	Endometriosis NOS	C0014175
PAR	inverse_isa	SNOMEDCT_US [SCUI - SCUI]	5562006	Endometriosis of intestine	C0156347
PAR	inverse_isa	SNOMEDCT_US [SCUI - SCUI]	26681001	Endometriosis of pelvis	C0241880
PAR	inverse_isa	SNOMEDCT_US [SCUI - SCUI]	89001007	Implanted endometriosis	C0269103
PAR	inverse_isa	SNOMEDCT_US [SCUI - SCUI]	46073007	Deciduomatosis	C0269104
PAR	inverse_isa	SNOMEDCT_US [SCUI - SCUI]	55850004	Endosalpingiosis	C0269106

Figure 25 – The UMLS Metathesaurus: partial results for <Endometriosis> within SNOMED CT

The second Knowledge Source under analysis is the UMLS **Semantic Network**, which constitutes a conceptual framework for all the concepts within the UMLS and that belong to the biomedical domain. (McCray, 2003) considers this resource to be an upper level ontology, in the sense that it aims to provide all UMLS concepts with a consistent and semantically coherent representation, supported by explicit concept relationships.

Hence, the Network comprises 134 Semantic Types and 54 relationships. The first are hierarchically arranged and linked by *is_a* relationships and, as stated above, each concept is assigned at least one Semantic Type within the hierarchy. With the exponential increase of the UMLS, it became necessary to try to reduce the level of conceptual complexity within the Semantic Types. (McCray, Burgun, & Bodenreider, 2001) proposed to aggregate the 134 types into a set of 15 broader categories, which are still in use today. The relationships, on the other hand, can be either hierarchical or associative and will be referred to once again in Chapter 6, as they made a valuable contribution to the conceptualization developed within the EndoTerm project. **Figures 26 and 27** illustrate a portion of the Semantic Network and a systematization of the Semantic Relations, respectively (McCray, 2003, p. 82; NML, 2009, p. 93).

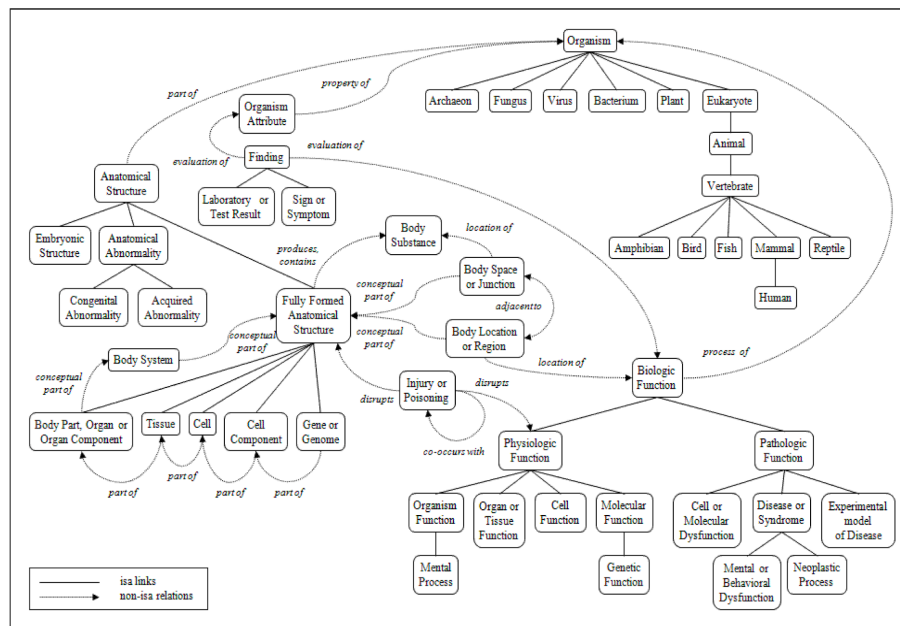


Figure 26 – A portion of the UMLS Semantic Network (with relations) (NLM, 2009)

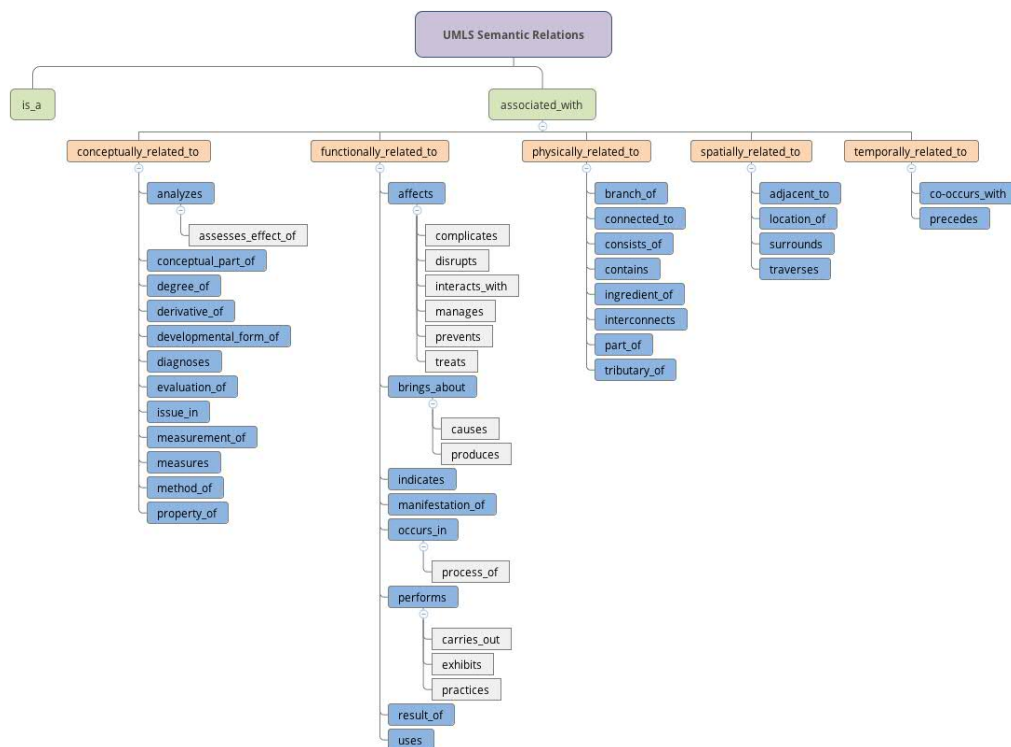


Figure 27 – Systematization of the UMLS Semantic Relations

Finally, the third Knowledge Source of the UMLS is the SPECIALIST Lexicon which, as the name indicates, provides all the necessary information for the underlying Natural Language Processing (NLP) features of the system, namely regarding syntax,

morphology and spelling (only available for English). It started being developed in the mid-1990s, while the encouraging advances in natural language processing applications occurred (Bodenreider, 2004). Some of the sources for the lexicon entries include Dorland's Illustrated Medical Dictionary, MEDLINE citations records or the UMLS Test Collection of MEDLINE abstracts (NML, 2009, pp. 101–127).

In the last few years, the UMLS has been working towards continuous improvement, which implies, for once, a decrease in mapping inconsistencies. Furthermore, and following deeper integration efforts with other TS, the UMLS Terminology Server currently provides direct access to SNOMED CT, both to its International Edition and to its specific UTS Browser.

3.3. Sharing biomedical knowledge: can standardization and diversity coexist?

Following what has been described in the previous sections of this chapter, as well as in Chapter 1, it becomes clear that knowledge sharing may, in fact, be the decisive factor in the current healthcare setting. And despite the indisputable value of human-oriented tools and resources, the focus has shifted towards machines and the aforementioned semantic interoperability. For that reason, and to overcome – or, at least, minimize – the hurdles caused by the coexistence of so many biomedical terminological systems, the quest for a standardized and reliable resource or format that facilitates machine-based communication in this domain has become paramount.

As far as a standardized resource is concerned, SNOMED CT seems to be taking the lead as the biomedical TS with a potentially global coverage in the near future, particularly due to its comprehensiveness and to its ability to provide a consistent and computer-processable representation of clinical content in EHRs, as shown in subsection 3.2.3.3 (Millar, 2016). Its concept model also facilitates interoperability, and the relatively recent initiatives involving cross-mapping with the ICD-10, LOINC, and the upcoming ICD-11 seem to be in full swing (Mea et al., 2014; J.-M. Rodrigues et al., 2013, 2014, 2015).

As for the use of a standardized format for clinical data sharing, the debate is far from complete. The reality is that current biomedical resources still hold a profusion of data formats. In biomedical ontologies, for example, it is possible to find at least three main syntaxes: the OBO format (the basis of the Gene Ontology and of the Disease Ontology⁷⁵); the Rich Release Format (RRF), used mainly by the NML; and OWL (which has become increasingly popular among several biomedical ontologies, such as the Foundational Model of Anatomy⁷⁶). Therefore, some recently developed projects aim at the standardization of the different formats in healthcare and advocate, in particular, the role of RDF as a common representation of biomedical information.

One of them is the BioPortal, where the roughly 600 biomedical ontologies gathered so far have been transformed, i.e. mapped, into RDF triples (Salvadores et al., 2013), that can be queried via SPARQL⁷⁷. Another initiative that is worth mentioning is the Yosemite Project, created in 2014 as a follow-up to the so-called Yosemite Manifesto⁷⁸, and whose mission is to “achieve semantic interoperability of all structure healthcare information” by relying on the use of RDF as a “universal information representation”. Moreover, the MeSH thesaurus is now available in a Linked Data/RDF beta version via the N-Triples format (Boehr & Bushman, 2018)⁷⁹. Whether the ultimate goal of semantic interoperability encompassing the whole healthcare domain is feasible or not, remains to be seen in the upcoming years. RDF’s added value as a common tool for representing healthcare information, however, seems to be indisputable.

⁷⁵ In fact, the DO has recently migrated their curation efforts into Protégé, in order to resort to OWL and reasoning to further strengthen the updating processes, as well as system and resource integration (cf. Schriml & Mittra, 2015). The DO .owl version is available here:

<https://raw.githubusercontent.com/DiseaseOntology/HumanDiseaseOntology/master/src/ontology/doid.owl>

⁷⁶ <https://bioportal.bioontology.org/ontologies/FMA>

⁷⁷ cf. the beta version at <http://sparql.bioontology.org/>

⁷⁸ The Yosemite Manifesto was a position statement with more than 100 signatures from leading experts in healthcare or healthcare-related fields which advocated RDF as the best candidate for a universal healthcare exchange language. In addition, it encouraged harmonization among standards, as well as government incentives that would allow a full implementation program. Additional information concerning both the Yosemite Manifesto and the Yosemite Project can be found at <http://yosemitemanifesto.org/> and at <http://yosemiteproject.org/>, respectively.

⁷⁹ The MeSH RDF data can be found at <https://id.nlm.nih.gov/mesh/>

Taking this framework into account, one might wonder if there is still room in biomedical TS for natural language definitions of concepts. It would appear so. One of the short-term objectives of the Disease Ontology, for instance, is to expand the number of textual definitions until reaching full coverage (Kibbe et al., 2015). The 11th version of the ICD (ICD-11), to be released later this year, should include “a short concise textual definition” for each entity, a feature that is not available in the current version (WHO, 2011a, p. 17). That will also be the case with the ICHI, now awaiting its official release, and where natural language definitions will be used to “describe the intervention” and “assist the user in selecting the most appropriate [intervention] code” (WHO, 2018).

Moreover, and even though the current version of SNOMED CT lacks natural language definitions, it is also likely that this issue will be addressed soon. On the one hand, 63% of SNOMED CT users stated, in a 2010 survey, that textual definitions would be extremely relevant (Elhanan, Perl, & Geller, 2011). On the other hand, SNOMED CT’s expected widespread use at an international level (e.g. it has fully replaced the Read Codes in the UK National Health Service’s Primary Care System in April 2018⁸⁰) will presumably gather various stakeholders with different areas of expertise and subsequently raise particular needs, one of them probably being natural language definitions.

Notwithstanding this growing interest in textual definitions, no unequivocal guidance has been explicitly provided by the aforementioned biomedical TS or their supporting guidelines on how to draft such definitions. In ICD-11, for instance, contributors who wish to propose a definition are advised to “describe the entity clearly and concisely” (WHO, 2011a, p. 19), as well as to resort to existing definitions as much as possible. However, no further, more specific, drafting recommendations are outlined. The overall picture is not very different in the remaining resources. In fact, one of the few – and pertinent – references to the governing principles of such definitions is to be found at the Draft ICHI Guidelines, which state that the definitions should “reflect the (...) axis categories from which the code is constructed” (WHO, 2018), thereby pointing

⁸⁰ <https://digital.nhs.uk/services/terminology-and-classifications/snomed-ct/snomed-ct-implementation-in-primary-care> (20.12.2017).

towards the conceptual core structure of the classification as a useful starting point in the development of natural language definitions.

This seems to be in line with what was explored in the previous chapter, especially in the sections concerning the link between Terminology and ontologies, and in matters regarding the drafting of natural language definitions. Once again, no additional information is provided in this respect by the supporting materials of such biomedical resources. Bearing in mind all of the above, it is believed that the current research work can make a contribution to systematizing the natural language definition drafting process, in this case within the subject field of endometriosis.

3.4. Setting the scene for biomedical terminological systems in the new millennium

In short, this chapter has highlighted the considerable development of biomedical terminological resources, particularly in the last few decades: from plain lists, these resources have evolved into compositional systems and, more recently, into formal concept representation systems with a clear link to Artificial Intelligence- and Knowledge Engineering-based approaches (Duclos et al., 2014, pp. 38–39). Driven, to a large extent, by the valuable contributions of Cimino (1998, 2006) and the GALEN Project, these resources are becoming increasingly computerized, multilingual, and multipurpose, and in fact, previously competing resources have now become complementary, due to more consistent and effective mapping. To quote the openGALEN project, “the possibility of multilingual systems based on underlying shared concept systems appears realistic in medicine⁸¹.” The advantages stemming from this progress are significant, and the latest advances into RDF/Linked Data have undoubtedly been paving the way to a more interoperable healthcare environment.

A simple example from the healthcare domain – more specifically, from Gynecology, the medical specialty that deals with endometriosis – will be provided below, to illustrate the transition from statements to data graphs, then to RDF graphs and, finally, to one of the serialization formats referred to in Chapter 1.

⁸¹ <http://www.opengalen.org/background/background12.html>

Example: A given patient (identified in the system as 123), called Jane Doe, underwent a radical hysterectomy – which is a type of hysterectomy⁸² – on August 15, 2017, at the Florida Hospital East Orlando (USA), whose homepage is www.floridahospital.com/east-orlando. The same information could be represented in triple statements as follows (Table 12):

SUBJECT	PREDICATE	OBJECT
patient123	has name	Jane Doe
patient123	has surgical procedure	radical hysterectomy
radical hysterectomy	is a	hysterectomy
radical hysterectomy	has date	15-08-2017
radical hysterectomy	has location	Florida Hospital East Orlando
Florida Hospital East Orlando	has homepage	www.floridahospital.com/east-orlando

Table 12 – Example of RDF statements

Alternatively, the six previous statements could also be depicted as a data graph⁸³ (Figure 28). In this case, no RDF serialization format has been used yet.

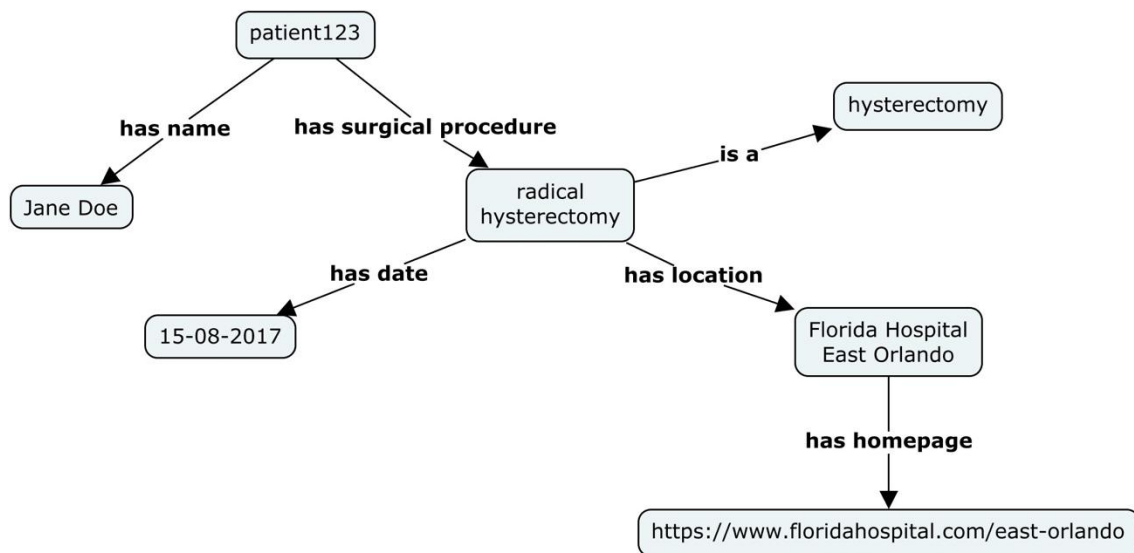


Figure 28 – Example of a data graph

⁸² As will be further explored in the third part of this work, a hysterectomy is a surgical procedure whereby a woman's uterus is removed. It is often regarded as a last resort measure for women with endometriosis. In a radical hysterectomy, the uterus, cervix and upper vagina are removed, together with the parametrium (i.e. round, broad, cardinal, and uterosacral ligaments) (<https://www.uptodate.com/contents/radical-hysterectomy>).

⁸³ The graph was created via CMap Tools, a free concept-mapping software developed by the Florida Institute for Human & Machine Cognition (IHMC) and available at <http://cmap.ihmc.us/>.

The subsequent transition into an RDF graph (**Figure 29**) relies on the principles referred to before, i.e. each subject will be replaced by its respective HTTP URI, as will each predicate, so as to allow external linkage. The objects will include either a URI (blue rounded rectangles) or a literal value (orange rectangles), and URI-objects in a given statement may become a subject in another RDF statement, as can be seen both from Table 12 and Figure 28. In order to guarantee consistency, predicate URIs should stem from existing RDF vocabularies, which are collections of URIs that provide a set of classes and properties for the SW. Some examples include the Dublin Core Metadata, the Friend of a Friend (FOAF) project, Schema.org, SKOS, OWL, just to name a few⁸⁴.

The predicate URIs used in this example belong to FOAF, Schema.org and RDF Schema (namely `rdf:type`), whereas the subject and object URIs were collected from the Example Domain (`example.org`), from Dbpedia⁸⁵, and from Bioportal, an open repository of biomedical terminological resources⁸⁶.

The last two statements `<radical hysterectomy> <has place> <Florida Hospital East Orlando>` and `<Florida Hospital East Orlando> <has homepage> <www.floridahospital.com/east.orlando>` were included in order to demonstrate that these objects' URIs in the RDF graph refer to different resources: the former is related to the hospital itself (i.e. to its Dbpedia entry), whereas the latter is related to the hospital's homepage. The change in the date entry (from 15-08-2017 to 2017-08-15) has

⁸⁴ [a] The **Dublin Core Metadata Terms** are, as the name implies, a set of metadata terms used to describe resources such as videos, images, Web pages, books, CDs, etc. The initial set, the Dublin Core Metadata Element Set (DCMES), containing 15 terms, was also published as an ISO standard in 2009 (ISO 15836: 2009). The DCMES and the remaining sets are maintained by the Dublin Core Metadata Initiative (<http://dublincore.org>). [b] The **Friend of a Friend (FOAF)** vocabulary is a machine-readable RDFS/OWL ontology used to describe people, activities, as well as relations to other people and objects (<http://xmlns.com/foaf/spec>). [c] **Schema.org** is a set of schemas used for structured data markup on Web pages. It can be used with RDFa and JSON-LD, for example, to add information to one's Web content (<http://schema.org>). [d] **Simple Knowledge Organization System (SKOS)** is a data model used for expressing Knowledge Organization Systems such as thesauri, classification schemes, taxonomies, etc., in a machine-readable way, i.e. within the framework of the Semantic Web (<https://www.w3.org/2004/02/skos/intro>). [e] The **Web Ontology Language (OWL)** is a computational, logic-based language used to represent knowledge about things, groups of things and relations between things. It is part of the SW's technology stack, together with RDF and SPARQL, and constitutes the basis of a significant number of biomedical ontologies nowadays.

⁸⁵ Dbpedia is a knowledge base aimed at extracting structured, multilingual information from Wikipedia and making this information freely available on the Web. The main website is <http://wiki.dbpedia.org/>.

⁸⁶ Further information can be found at the homepage <http://bioportal.bioontology.org/>.

to do with the value allowed by `schema:date`, which needs to follow the ISO 8601 date format.

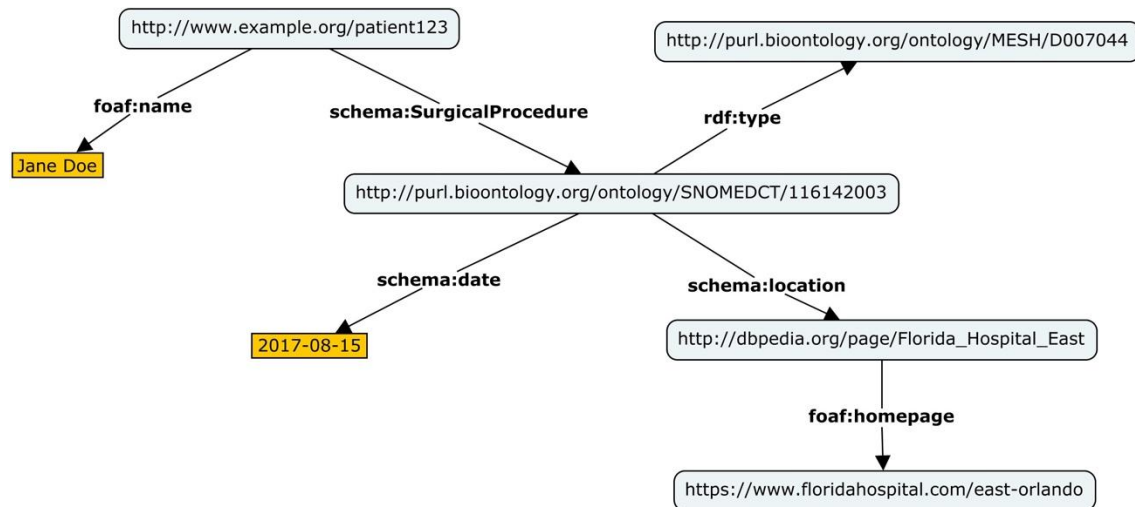


Figure 29 – Example of an RDF graph

The following figure (**Figure 30**) exemplifies the final step of this demonstration, which concerns the use of a serialization format to publish the information from the RDF graph on the Web. In this case, the RDF/XML format was chosen. The snippet below depicts the various statements and the description of each resource using this XML-based format.

```

<?xml version="1.0"?>

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1"
  xmlns:schema="http://schema.org"
  xmlns:ex="http://example.org/stuff/1.0/">

  <rdf:Description
    rdf:about="http://example.org/patient123">
    <foaf:name rdf:resource="Jane Doe"/>
    <schema:SurgicalProcedure rdf:resource="http://purl.bioontology.org/ontology/SNOMEDCT/116142003"/>
  </rdf:Description>

  <rdf:Description
    rdf:about="http://purl.bioontology.org/ontology/SNOMEDCT/116142003">
    <rdf:type rdf:resource="http://purl.bioontology.org/ontology/MESH/D007044"/>
    <schema:date rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2017-08-15</schema:date>
    <schema:location rdf:resource="http://dbpedia.org/page/Florida_Hospital_East"/>
  </rdf:Description>

  <rdf:Description
    rdf:about="http://dbpedia.org/page/Florida_Hospital_East">
    <foaf:homepage rdf:resource="https://www.floridahospital.com/east-orlando"/>
  </rdf:Description>
</rdf:RDF>

```

Figure 30 – Example of RDF/XML

This chapter has also demonstrated, however, a more accurate perception of the discrepancies in the various resources due to differing structures, principles, needs, and user types. This will, in turn, affect not only the concept of <Concept> depicted by these resources but also of other, related notions (cf. the variety of notions in the various resources, such as term, atom, descriptor, Fully Specified Name, name, etc.).

All in all, there is still a long way to go in what concerns system integration and only the future will tell if, within the wide range of available terminological resources, one can avoid or at least decrease conceptual inconsistencies and overlaps, while aiming to maximize the reuse of existing clinical information in different healthcare settings. The quest for consistency from a conceptual perspective must not overlook the relevance of its linguistic counterpart, though. As argued in the previous chapter, it is possible - and, indeed, desirable - to allow linguistic and conceptual information to coexist within a resource while maintaining their fundamental differences. The EndoTerm project aims to contribute to that objective.

Part III: EndoTerm – Rationale, Methodology, and Implementation

Chapter 4 – Endometriosis: the most common disease one has never heard of?⁸⁷

4.1. Endometriosis: facts and figures

As referred to in the introduction, endometriosis, which stems from Greek *éndon* (in, within), *métra* (womb), and *osis* (condition), is a chronic gynecologic disease which arises when tissue similar to the endometrium (i.e. the lining of the uterus) is found outside the uterine cavity (Dunselman et al., 2014). It affects an estimated 1 in every 10 women of reproductive age, i.e. approximately 176 million women worldwide (Adamson, Kennedy, & Hummelshoj, 2010; Rogers et al., 2009), and has a significant economic and social impact in today's society, as will be seen in subsequent sections in this chapter, accounting for severe absenteeism and productivity loss among millions of women each year, as well as for substantial costs worldwide in healthcare appointments, surgical procedures and imaging tests.

Despite its considerable prevalence, endometriosis is still a relatively unknown disease, even amidst healthcare professionals. Therefore, several initiatives have been undertaken in the last few years to help raise global awareness about this condition. One of the first actions at an institutional level was the “Written Declaration on Endometriosis”, introduced by a group of Members of the European Parliament (MEP) in March 2005, and which aimed at the recognition of endometriosis as a social disease. Supported by over one-third of MEP at that time (266 of them), the document also urged the European Commission to include prevention of endometriosis in future Action Programs for Public Health, to enable research into the causes, prevention, and treatment of this condition (Bianconi et al., 2007). Since results at this level have been taking some time to materialize, in September 2015 another “Written Declaration on Research and Prevention of Endometriosis” was launched by various MEP (Basterrechea et al., 2015), calling upon the EC to increase financial support concerning endometriosis research and prevention, as well as the drafting of clinical guidelines and the creation of reference centers throughout Europe.

⁸⁷ Adapted from the title of a 2017 TED Talk given by the American director and producer Shannon Cohn (<https://goo.gl/ezuWVh>).

As far as national governments are concerned, significant breakthroughs have been accomplished in the last few years, most notably in France, where partnership agreements aiming at awareness campaigns and healthcare professional training actions have been set up between the patient support platform Info Endométriose and the Ministry of Women's Rights (in March 2016), the Ministry of Education (in June 2016), and the Ministry of Health (in March 2017)⁸⁸. Moreover, just in December 2017, the Australian government, through its Ministry of Health, issued a public apology to all endometriosis patients in the country for the lack of attention paid to this condition and launched the first National Action Plan for Endometriosis, which will include several education, awareness, and research initiatives⁸⁹.

The increasing relevance of endometriosis has also become prominent within the expert community, with the number of PubMed publications containing endometriosis as part of their title or abstract rising from merely 2 in 1927 to 1,101 in December 2017⁹⁰. Moreover, a specialised, peer-reviewed publication, entitled *Journal of Endometriosis and Pelvic Pain Disorders* (JEPPD), was launched in 2009⁹¹, and there are several congresses dedicated to the disease, in particular the World Congress on Endometriosis, whose forthcoming edition (the 14th) will take place in 2020, in Shanghai, under the very pertinent title: "Endometriosis: cracking the enigma"⁹². In addition, and as will be further developed in this chapter, a considerable number of documents, particularly guidelines, consensus statements, classification proposals, and standardized data collection tools have been published over a relatively short period of time (5-10 years), especially by multidisciplinary *consortia*, Special Interest Groups (SIGs) and professional societies. Some examples of such organizations include, for instance, the Endometriosis SIG of the American Society for Reproductive Medicine (ASRM), the AAGL (formerly known as the American Association of Gynecologic Laparoscopists) Endometriosis/Reproductive Surgery SIG, as well as the SIG for Endometriosis and

⁸⁸ More information at <http://www.info-endometrioise.fr/nos-actions-avec-les-ministeres>.

⁸⁹ The Ministry of Health's Press Release can be consulted at: <https://goo.gl/aVvAr8>.

⁹⁰ The query, with the expression **endometriosis[Title/Abstract]**, was conducted on PubMed on December 20, 2017.

⁹¹ <http://www.j-endometriosis.com> – Some articles from this source were included in the EndoCorpus, as will be shown in Chapter 7.

⁹² More information at <http://endometriosis.ca/world-congress/wce2020/>.

Endometrial Disorders (SIGEED) of the European Society of Human Reproduction and Embryology⁹³.

In this regard, further reference should also be made to two international organizations that gather various stakeholders⁹⁴ and try to adopt a call-for-action approach concerning the disease. The World Endometriosis Society (WES), on the one hand, describes itself as the “global organisation for endometriosis clinical practice, research, education, and advocacy”. In addition to many other initiatives, it is responsible for organizing the stated World Congress on Endometriosis. On the other hand, WES collaborates extensively with the World Endometriosis Research Foundation (WERF), a “global charity which facilitates research into endometriosis to improve knowledge and treatment(s)”⁹⁵. Currently collaborating with 75 institutions in over 30 countries, WERF has sponsored a set of clinical studies that will be presented in more depth in subsequent parts of this chapter.

As for the patient side and given that empowerment is becoming more and more widespread (as referred to in Chapter 1), it is no wonder that the number of endometriosis support groups and patient associations has experienced a substantial growth in recent years. According to the platform endometriosis.org, there are at least 50 official endometriosis organizations around the world⁹⁶. As mentioned in the first chapter, patients are also more and more willing to take the driver’s seat concerning their own health, which has led to the advent of specialised endometriosis technological resources that rely on the information supplied by the patients themselves so as to compile as much data as possible, in a crowdsourcing perspective, hence contributing to an earlier diagnosis and a more effective management of the disease. One of these resources is the *Observatoire de l’Endométriose* (OZ2020)⁹⁷, a French consortium comprising subject field experts, patient organizations, the pharmaceutical, genetic, and

⁹³ Further institutional information can be found at <https://goo.gl/4yskxA>, <https://goo.gl/NstfVn>, and <https://goo.gl/LCmLio>, respectively.

⁹⁴ Not only gynaecologists/surgeons and physicians from other specialties, but also experts in health informatics, health economics, sociology, biology, genetics, and, last but not least, representatives of patient organizations.

⁹⁵ <https://endometriosisfoundation.org/>

⁹⁶ Cf. <https://goo.gl/ufRoZu>. The number of informal patient support groups is more difficult to determine, but a quick search on Facebook for “endometriosis support” on 28 December 2017 resulted in approximately 120 pages and 100 groups in total, some of the latter containing more than 10,000 members each.

⁹⁷ <http://www.oz2020.com/>

biology industry (e.g. ENDODIAG and the Genopole d'Evry), as well as eHealth technology developers (such as BePATIENT). OZ2020 aims to provide an integrated platform where patients can learn more about endometriosis, insert/consult/download their own health data, contact other patients and stakeholders, and enrich the existing datasets to improve research on the disease. Another initiative to bear in mind is Phendo, an observational research study conducted by Columbia university "to track, manage and understand endometriosis"⁹⁸. Similarly to OZ2020, Phendo promotes the collection of patient data through an app - downloaded mainly into smartphones and tablets -, with the purpose of creating a catalogue of endometriosis signs and symptoms in order to better understand the disease.

At a more global scale, one of the most successful and impactful events has been the Worldwide Endometriosis March (EndoMarch), founded in 2013 to empower, educate and effect change, and currently being organized in more than 50 countries, mainly by patient associations⁹⁹. Another relevant initiative included the worldwide release, in March 2016, of a documentary entitled "EndoWhat?"¹⁰⁰, produced and directed by Shannon Cohn, which aims to inform the general public about the disease and empower endometriosis patients. Moreover, and very recently (November 2017), the photo below (**Figure 31**), by the British photographer Georgie Wileman, depicting her endometriosis surgeries between 2014 and 2017, was awarded the Taylor Wessing Portrait Prize and displayed at the National Portrait Gallery in London¹⁰¹.

⁹⁸ <http://citizenendo.org/phendo/>

⁹⁹ The history of this event, as well as information concerning each year's edition, are available at <http://www.endomarch.org/>. 'March' refers here not only to the type of event itself, but also to the month in which it takes place, since March has been chosen as the endometriosis awareness month worldwide.

¹⁰⁰ <https://www.endowhat.com/>

¹⁰¹ Further information concerning both this photo and the overall project which it was part of, called 'Endometriosis', can be found at <https://www.georgiewileman.com/endo/>.



Figure 31 – ‘2014-2017’, by Georgie Wileman

All in all, it appears that major steps are being taken to bring this disease into the public eye, and in this regard, the EndoTerm project aims to become an additional contribution to this endeavor, particularly in what concerns knowledge representation, organization, and sharing in this domain.

Therefore, this chapter focuses on providing a comprehensive overview of endometriosis, with reference not only to its etiology, main signs and symptoms, diagnosis, treatment, and classification(s), but also to its history and current social and economic impact. Ultimately, by better grasping the chosen subject field, a terminologist will be better equipped to propose a systematization that may then culminate in a domain conceptualization.

4.2. Endometriosis: a riddle wrapped in a mystery inside an enigma¹⁰²

Often regarded as an enigmatic disease, endometriosis is defined as the presence of endometrial glands and stroma outside the uterine cavity (Giudice, 2010; Giudice, Evers, & Healy, 2012). In addition, this condition is described as estrogen-dependent, and since this tissue mimics the behavior of the endometrium within the context of monthly menstrual cycles, bleeding in response to exogenous, endogenous, and hormonal stimuli, an inflammatory process occurs, resulting in lesions and

¹⁰² Adapted from a quote by the physician Emery A. Wilson (1987).

adhesions (Bulun, 2009; Giudice, 2010; Kennedy et al., 2005). These adhesions (i.e. bands of fibrous scar tissue) can, in turn, bound the organs together and limit their function and mobility, thus causing pelvic pain and, in more advanced stages of endometriosis, a phenomenon known as “frozen pelvis”¹⁰³, in which the woman’s anatomy concerning reproductive organs and adjacent structures is seriously distorted. Endometriosis lesions can, thus, be either more superficial or highly invasive, i.e. involving vital structures, such as the bowel, the ureters or the bladder. These more invasive lesions are commonly known as deep infiltrating endometriosis or DIE.

The American gynecologist and surgeon John Sampson, who coined the term “endometriosis” in the 1920s, was one of the first to systematically categorize endometriosis lesions and to propose that these could have variable sizes, shapes and colors (Martin et al., 2017), depending on the evolution process of lesions and the length of that process¹⁰⁴. **Figure 32** depicts examples of lesions identified by Sampson himself.

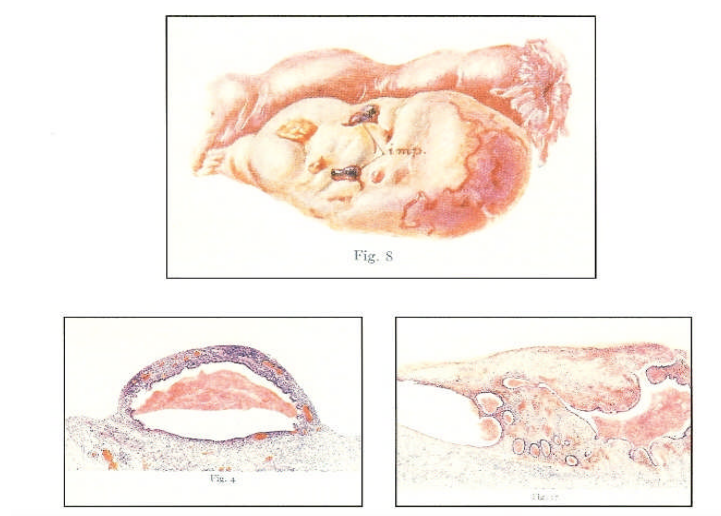


Figure 32 – Endometriosis lesions identified by Sampson (1924)

cited in (Martin et al., 2017)

¹⁰³ Cf. <http://www.medilexicon.com/dictionary/66552>.

¹⁰⁴ In one of his first papers on endometriosis (Sampson, 1921) introduced the notion of ‘chocolate cyst’ to refer to endometriomas, i.e. to endometriotic cysts in the ovaries, given that the old blood found inside these lesions is often a thick, dark, chocolate-like fluid.

Overall, endometriosis lesions are commonly located in the pelvic cavity - besides the ovaries, also the fallopian tubes, the peritoneum (i.e. the lining of the pelvis), the uterosacral ligaments, the round ligaments, the pouch of Douglas (also known by the French borrowing *cul-de-sac*), and the rectovaginal septum -, although they can correspondingly be found in the bowel, bladder, colon, rectum, or vagina, as well as in more remote locations, such as the brain or lungs (Dunselman et al., 2014). In **Figure 33** below, it is possible to see some possible sites for endometriosis lesions.

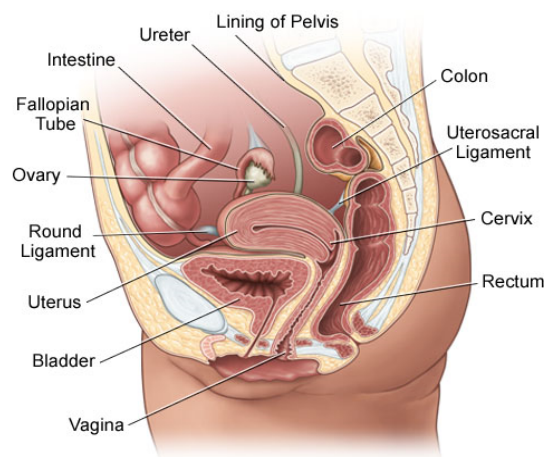


Figure 33 – Possible locations of endometriosis¹⁰⁵

On the other hand, it is difficult to reach a diagnosis, given the diversity of signs and symptoms¹⁰⁶, which are often mistakenly associated with other conditions, such as irritable bowel syndrome or interstitial cystitis. These signs and symptoms may not always correlate with the severity of the disease. In addition to infertility, women are affected predominantly by pain, which, in the most severe cases, can be excruciating to the point that it prevents regular day-to-day activities. The most common symptoms in that regard include dysmenorrhea (painful menstruation), deep dyspareunia (painful sexual intercourse), dyschezia (pain while defecating), dysuria (pain while urinating), and chronic pelvic pain (CPP), i.e. “pelvic pain lasting for six months or longer” (NICE,

¹⁰⁵ <https://goo.gl/hcpVuv>.

¹⁰⁶ Within the context of healthcare terminology, it is important to distinguish between **sign** and **symptom**. Whereas the former refers to objective evidence of a disease that can be perceptible to a physician as a result of a clinical examination or a given procedure (e.g. high blood pressure), the latter is related to any subjective evidence that can be perceived by a patient as being indicative of a disease (e.g. pain) (cf. <https://www.dorlandsonline.com/dorland/dictionary>).

2017). Other cyclical intestinal or urinary complaints are often present as well, such as bloating, diarrhea, and constipation, or hematuria (i.e. presence of blood in urine), respectively. This profusion of clinical manifestations may very likely be one of the factors contributing to an average diagnostic delay of almost a decade, as will be further explored in a subsequent section.

With the exponential growth and development of new tools and services, though, it is expected that substantial progress can be made as regards the early detection of this condition. At the moment, and in addition to the clinical interview and to the clinical/pelvic examination, there are several imaging procedures that support this diagnosis, namely transvaginal sonography (TVS), (pelvic) magnetic resonance imaging (MRI) examination, computer tomography (CT) scan, as well as transrectal sonography, colonoscopy, barium enema, or cystoscopy, especially in cases where there is a suspicion of bowel and/or bladder involvement. None of them is 100% accurate, though. In fact, the current 'gold standard' concerning the diagnosis of endometriosis comprises laparoscopic surgery and the following histological confirmation of the removed lesions (Dunselman et al., 2014). Other, less invasive options have been taken into account over the years, namely blood tests based on the CA-125 biomarker, usually recommended for the diagnosis of ovarian cancer, although its actual effectiveness may be debatable (Johnson et al., 2013). More recently, a French company has developed EndoDTect®, an innovative blood test based on disease-specific biomarkers which is expected to help drastically reduce the aforementioned diagnostic delay¹⁰⁷. However, it is currently undergoing validation studies, and no conclusive results have been reached so far.

Another key question regarding endometriosis is its inherent complexity also in what concerns its pathogenesis, an issue which remains widely debated. Over the past few decades, several theories to what might cause the disease have been put forward, yet none has been fully accepted, largely because the knowledge about the disease has been changing and evolving very rapidly. Furthermore, there is the notion, among the

¹⁰⁷ More information about this test and the company itself - EndoDiag - can be found at <https://www.endodiag.com/early-diagnosis>.

various stakeholders involved, that there is still a lot more to discover (Rogers et al., 2017).

Overall, these different theories can be grouped into three main categories, introduced by a comprehensive survey from the 1960s (Ridley, 1968): 1) *in situ* development theories; 2) transplantation theories; and 3) combination of both preceding categories. In the first category, it is argued that the lesions develop *in situ* (i.e. within a given location) from local tissues. One example of that category is the so-called coelomic metaplasia theory, according to which cells present in the mesothelial lining of the peritoneum would undergo metaplasia and transform into endometrial-like tissue as a result of undetermined factors (e.g. hormonal, environmental, etc.) (Nap, 2012). Another theory from the *in situ* category states that the disease would result from defective embryogenesis, more specifically from residual embryonic cells of either the Wolffian or the Müllerian ducts¹⁰⁸, which would develop into endometriosis lesions when subjected to certain stimuli (which, again, are yet to be clearly identified) (Sourial, Tempest, & Hapangama, 2014). A third approach belonging to the first category is the stem cell theory, which presupposes the existence of endometrial stem cells. Due to their undifferentiated nature and, hence, their ability to self-renew, these cells would play a key role in the cyclic regeneration of the endometrium during the menstrual cycle and, therefore, in the pathogenesis of several gynecological conditions, including endometriosis (Maruyama, 2014).

The second category, concerning transplantation theories, argues that the endometrial tissue is somehow transported from the uterus to other external locations. Within this group, the most influential theory - and still the most widely accepted today - was first proposed by (Sampson, 1927). Also known as the theory of retrograde menstruation, it states that menstrual tissue refluxes through the fallopian tubes and subsequently implants on pelvic structures, originating endometriosis lesions. Another approach has been put forward regarding the possible transportation of endometrial

¹⁰⁸ In the embryonic phase, the coelomic epithelium gives rise to the aforementioned Müllerian ducts, which, in turn, differentiate to form the female reproductive tract (i.e. the fallopian tubes, the uterus, the cervix, and the upper part of the vagina) (<https://goo.gl/aEyoVL>).

tissue into locations outside the uterine cavity, namely the dissemination via the lymphatic and vascular systems (Acién & Velasco, 2013; Nap, 2012).

Overall, the pathogenesis of endometriosis is still a controversial issue and the various proposals referred to above aim to provide answers to some of the enduring questions surrounding this enigmatic disease, albeit without unequivocal success: the existence of endometriosis lesions in prepubertal girls; the disease recurrence, even in cases where women have undergone radical surgery in order to remove all lesions and adhesions; the presence of endometriosis lesions in locations as remote as the brain; as well as the reason why only some women are affected by the disease if evidence has been shown that most of them experience retrograde menstruation. Despite the ongoing debate, current research has been progressively confirming that the disease etiology is very likely multifactorial, potentially involving a combination of genetic (Malinak et al., 1980; Simpson et al., 1980)¹⁰⁹, immunological, and environmental factors. Regarding the latter factors, several studies have claimed in the last two decades that the intense exposure to endocrine disruptors, such as dioxins, pesticides, phthalates, heavy metals, etc., which can mimic hormonal behaviour, can trigger a number of hormone-dependent pathologies, as is the case of endometriosis (Balabanič, Rupnik, & Klemenčič, 2011; Birnbaum & Cummings, 2002; Rier & Foster, 2002; Yao et al., 2017).

Bearing all the preceding information in mind, it is not surprising that finding a satisfactory treatment for endometriosis has been difficult. In fact, there is at present no known cure for this condition. Nowadays, the available options comprise essentially surgical treatment, hormonal therapy, and the use of analgesics, or painkillers. As stated earlier, laparoscopic surgery is currently the gold standard in what concerns the diagnosis of the disease, together with the histological confirmation of the removed tissue sample(s). However, laparoscopy can also be an effective first-line approach in terms of treatment. It is now increasingly supported by robotic tools and usually implies, when compared to laparotomy, less pain, a shorter hospitalization period, a quicker recovery and a better cosmetic outcome (Dunselman et al., 2014; Johnson &

¹⁰⁹ Back in the 1980s, the cited authors had already alerted to the familial tendency of the disease. They concluded that there was a 7-10-fold increased risk for women with an affected first-degree relative (e.g. mother or sister).

Hummelshoj, 2013; NICE, 2017). Although it does not represent a cure for endometriosis, a hysterectomy - possibly with bilateral salpingo-oophorectomy¹¹⁰ - may be necessary in extreme cases, especially for women who do not respond to more conservative treatment and who have either completed their family or do not wish to pursue (other) fertility options (Rogers et al., 2009).

Hormonal therapy, on the other hand, aims mainly to suppress the menstrual cycle, hence controlling the symptoms and, in some cases, actually leading to a temporary menopause. Contraceptives, either oral, vaginal, or transdermal (via patches), are widely used, often in a continuous way. In more severe cases, options may include gonadotropin-releasing hormone (GnRH) agonists (e.g. nafarelin, leuprolide, buserelin, goserelin, triptorelin), progestagens (e.g. medroxyprogesterone acetate, dienogest, cyproterone acetate, levonorgestrel¹¹¹, etc.), and the disease-specific Danazol¹¹², among several others (Johnson & Hummelshoj, 2013).

In addition, analgesics are frequently used to relieve the pain, namely paracetamol, opioids, and especially the group of non-steroidal anti-inflammatory drugs (NSAIDs), such as ibuprofen, naproxen, piroxicam, etc. Finally, there are also some complementary therapies, such as acupuncture, Chinese herbal medicine, pelvic floor rehabilitation, homeopathy, as well as the use of supplements like magnesium or primrose oil, the benefits of which still lack further study and subsequent evidence (Dunselman et al., 2014). As for endometriosis-related infertility, hormonal treatment (especially controlled ovarian stimulation - COS) is often used to support Assisted Reproductive Technology (ART) procedures, particularly In Vitro Fertilization (IVF) (Zegers-Hochschild et al., 2009).

¹¹⁰ A hysterectomy with bilateral salpingo-oophorectomy corresponds to the removal of the uterus, of both ovaries, and of both fallopian tubes (cf. <https://goo.gl/SbuQTH>).

¹¹¹ This hormonal medication can also be administered via a LNG-IUS, or levonorgestrel-releasing intrauterine system.

¹¹² Danazol is defined as “a synthetic steroid with antigonadotropic and anti-estrogenic activities that acts as an anterior pituitary suppressant by inhibiting the pituitary output of gonadotropins”. It possesses some androgenic properties and has been used in the treatment of endometriosis, as well as of some benign breast disorders (<https://www.drugbank.ca/drugs/DB01406>).

Given the inherent complexity that characterizes this disease, the several guidelines and consensus statements that have been issued by worldwide expert organizations in recent years focus on the need for an individualized approach, supported by multidisciplinary teams¹¹³ in centers of expertise (Dunselman et al., 2014; Johnson et al., 2017; Johnson & Hummelshoj, 2013; Rogers et al., 2017). Furthermore, they recommend concrete endometriosis awareness-raising actions/initiatives, destined not only to people in general but also to healthcare professionals. The implementation of such objectives would undoubtedly contribute to reduce the diagnostic delay referred to earlier, as well as improve the quality of life (QoL) of affected women and lead to a less significant economic and social impact, all of which will be discussed below.

4.3. The burdens of endometriosis: a delayed opportunity for millions of women?

As stated above, the prevalence of endometriosis is, indeed, difficult to determine: it is estimated to affect 10% of women in reproductive age worldwide, with a clear incidence among infertile women (25-50%) (Meuleman et al., 2009). Conversely, about 30-50% of women with endometriosis are infertile (Practice Committee of the American Society for Reproductive Medicine, 2012). However, it is still unclear whether those numbers are accurate: on the one hand, some women are asymptomatic (approximately 20% to one-third) (Bullett et al., 2010) and, on the other hand, the endometriosis-related signs and symptoms are often mistakenly attributed to other diseases, which may contribute to the significant diagnostic delay that was mentioned earlier. However, the culturally established idea that pain during menstruation is normal can also be partly responsible for this deferment of almost a decade.

In addition to this, some authors (Arruda et al., 2003; Ballard, Lowton, & Wright, 2006; Hadfield et al., 1996; Hudelist et al., 2012; Nnoaham et al., 2011) report that the lack of consistent information concerning the disease is still substantial, not only among affected women but also among physicians, especially in the primary care setting.

¹¹³ Including medical professionals specialised in gynaecology, urology, gastrointestinal tract, gynaecological imaging, (pelvic) pain management, and fertility, nurses, physical therapists, psychologists, etc.

Therefore, the development and implementation of public awareness initiatives are recommended, along with educational programs in public schools and specialized training, particularly for future gynecologists and general practitioners (GPs). (Johnson & Hummelshoj, 2013, p. 1557) actually argue that “endometriosis diagnosis and management should be incorporated into the primary health care of women worldwide”. **Table 13** contains an overview of some of the studies conducted in the last few years that aimed to calculate the timespan from the onset of symptoms to the definitive diagnosis of this condition.

Authors	Country/-ies	No. of women	Data gathered by	Average delay (in years)
Hadfield et al. [1996]	UK/US	218	Questionnaire	7.96-7.92 (UK) / 11.73-9.05 (US)
Arruda et al. [2003]	Brazil	200	Questionnaire	7
(Husby, Haugen, & Moen, 2003)	Norway	261	Questionnaire	6.7-6.2
Ballard et al. [2006]	England	32	Interview	8.5
Nnoaham et al. [2011]	10 countries	1,418	Questionnaire	6.7
Hudelist et al. [2012]	Austria & Germany	171	Questionnaire	10.4

Table 13 – Studies on the diagnostic delay of endometriosis

From the table, it becomes clear that the aforementioned diagnostic delay is, indeed, considerable (approximately 8 years, on average), even across different countries, with distinct socioeconomic profiles and healthcare systems - this is visible in Nnoaham et al. (2011), for instance. Moreover, these studies show that the situation has not been improving throughout the years, which also raises the question of the impact - both social and economic - that such a delay can have.

In fact, over the last decade, research has been undertaken regarding the economic burden of endometriosis, including both the direct costs, i.e. surgical procedures, diagnostic tests, physician visits, hospitalization periods, medication, and the indirect costs, comprising absenteeism - not being present at work - but also

presenteeism, defined as “coming to work despite illness, injury, anxiety, etc., often resulting in reduced productivity”¹¹⁴. These indirect costs can, nevertheless, be difficult to quantify, as stated by (Soliman et al., 2016), most notably due to the complexity of endometriosis, as well as to its challenging diagnosis, as referred to earlier. The results of some of the most cited studies on this topic have been summarized in **Table 14**:

Authors	Study design	Data collection	Country/-ies	Year of costing	Sample	Direct costs (average)	Indirect costs (average)
(Gao et al., 2006)	Review	Lit. review + national database an. [1990-2004]	USA	-----	Not specified	\$13,535 (total)	-----
(Nnoaham et al., 2011)	Cross-sectional study	Retrospective survey	10 countries	2007 hourly labour cost	1,418	Absenteeism: from US\$1/wk in Nigeria to US\$231/wk in Italy	Presenteeism: from US\$3/wk in Nigeria to US\$250/wk in the USA
(A. R. Levy et al., 2011)	Cross-sectional study	Retrospective survey	Canada	2009	27 patients 18 phys.	1,164 (Canadian \$) Per patient / per year	4,043 (Canadian \$) Per patient / per year
(Simoens et al., 2012)	Cross-sectional study	Prospective survey	10 countries	2009	909	3,281€ per woman/year	6,298€ per woman/year
(Prast et al., 2013)	Cross-sectional study	Retrospective survey	Austria	2009	73	5,606€ per patient/year	2,106€ per patient/year
(Soliman et al., 2016)	Review	Lit. review	Various	-----	12 primary studies	-----	-----
(Koltermann, et al., 2017)	Review	Lit. review [2004-2015]	Various	-----	11 articles	-----	-----

Table 14 – Studies concerning the economic burden of endometriosis

In general, and despite the significant variation across the different studies, namely regarding the methodology, the data sources, the timeframes, as well as the diverging notions of direct and indirect costs, the numbers point towards a substantial economic burden, which, according to the results presented by Simoens et al. (2012) (total costs of €9,579 per woman per year), can be comparable to the estimated annual

¹¹⁴ <http://www.dictionary.com/browse/presenteeism?s=t> (18.10.2017)

costs of other chronic diseases, such as diabetes mellitus, Crohn's disease, and rheumatoid arthritis. Therefore, and besides the sizable impact of the direct costs, mainly as regards surgical procedures, the studies above demonstrate that the indirect costs are particularly high. In this context, Nnoaham et al. (2011) calculated an overall work productivity loss of 10.8 hours a week, a considerable presenteeism-related impact which often represents an "invisible cost to organizations", as referred to by (Quazi, 2013). Most of the cited literature alerts to a possible undervaluation of the overall endometriosis economic impact. Approximately a decade ago, Gao et al. (2006: 1570) had already signaled the fact that the estimated direct costs for endometriosis at that time might be "the tip of the iceberg" concerning the true impact of the disease upon society. Taking this framework into consideration, Soliman et al. (2016) and Koltermann et al. (2017) highlight the need to update and further develop these studies, in order to obtain a more comprehensive and realistic glimpse of this economic burden and, in turn, contribute to improve the efficiency of healthcare provision and resource allocation.

In addition to these considerable financial repercussions, recent work has also demonstrated that the social and psychological impact of endometriosis pervades practically all aspects of women's lives. On the one hand, most of the women with endometriosis experience a significant reduction of their Quality of Life (QoL)¹¹⁵, mainly due to the previously stated chronic pelvic pain and dyspareunia (De Graaff et al., 2013; Facchin et al., 2015; Gilmour, Huntington, & Wilson, 2008). As a consequence, anxiety, emotional distress, as well as low self-esteem tend to arise, and, in some cases, even depression (Culley et al., 2013; Moradi et al., 2014). According to De Graaf et al. (2013), half of the women who responded to the questionnaire admitted a negative impact of the disease on their marital and intimate relationships and, from those, almost 20% considered endometriosis as a cause of their divorce. On the other hand, the crippling pain and the other symptoms very often prevent women from having a normal social life, from completing their education, or from actually being able to work full-time. In

¹¹⁵ **Quality of Life** is defined by the World Health Organization as "an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns". The WHO further develops this definition, highlighting the broad nature of this concept, which comprises "the person's physical health, psychological state, personal beliefs, social relationships and their relationship to salient features of their environment". (<http://www.who.int/healthinfo/survey/whogol-qualityoflife/en/>)

fact, in the study by De Graaf et al. (2013), 48% of the women who confirmed that endometriosis had a very negative impact on their professional life needed to reduce working hours. In light of this worrying background, Culley et al. (2013) and Facchin et al. (2015) recommend not only the implementation of psychological support actions for women living with this condition but also the development of additional and better training for healthcare professionals. All in all, a more patient-centered and holistic approach is needed in the management of endometriosis (cf. De Graaf et al. 2013).

In short, the preceding studies show that the current burden of endometriosis is, indeed, highly significant and multidimensional. Often described as a career woman's disease, endometriosis usually tends to be associated to contemporary lifestyles, more specifically of busy female professionals exposed to daily stress, dietary changes, and increased smoking and drinking habits, who delay motherhood or choose not to pursue it at all. However, the idea that endometriosis is a 20th-century disease has been proven not to be entirely accurate, as will be presented in the following section.

4.4. Endometriosis and its many 'faces': between concept and term

To contribute to a brighter future for new generations of women worldwide, recent research also believed it was important to delve into the past and find out more about this enigmatic disease. As mentioned before, the term itself was coined by Sampson, yet the question remained: when was endometriosis identified as a distinct clinical entity? Hence, a series of studies published at the onset of the 21st century launched a debate concerning this issue.

In fact, this renewed interest in the subject started in 1999, when Vincent Knapp argued that the first description consistent with endometriosis had been put forward 300 years earlier, by the German physician Daniel Christian Schrön, in his 1690 dissertation "*Disputatio Inauguralis Medica de Ulceribus Uteri*" (Knapp, 1999). More than a decade later, in a comprehensive reconstruction of the history of the disease, Ronald Batt (Batt, 2011) advocated that the Austrian physician and pathologist Karl Freiherr von Rokitansky was the first to provide a histological description of lesions compatible with endometriosis, in the 1860 paper "*Über Uterusdrüsen-Neubildung in Uterus und Tubenwandung*", where he was able to identify endometrial glands and

stroma in the myometrium and ovaries, designating the findings in the myometrium as “cystosarcoma adenoids uterinum”.

Both Knapp and Batt’s assertions have been rebutted by (Benagiano & Brosens, 2011; Benagiano, Brosens, & Lippi, 2015; Brosens & Benagiano, 2011), who maintain that the first clear description, in morphological and clinical terms, of what is currently known as endometriosis was presented by the Canadian gynecologist and surgeon Thomas Cullen, in the book “Adenomyoma of the Uterus” (Cullen, 1908).

As mentioned before, Sampson coined “endometriosis” in the 1920s, referring to the presence of endometrial tissue outside the uterus, whereas (Frankl, 1925) proposed the term “adenomyosis uteri” to designate the presence of endometrial tissue within the uterine cavity, namely the myometrium¹¹⁶.

To summarize, while the authorship of the term “endometriosis” is clear, the question of when it was first identified as a disease *per se* remains a subject of debate. This is partly due to the lack of indisputable evidence and documentation on the disease itself, especially before the 1850s. Indeed, the relatively well documented period following von Rokitansky’s findings differs from the sparse information that characterized the preceding centuries.

Nevertheless, a paper by (Nezhat, Nezhat, & Nezhat, 2012) not only claims to have found credible evidence supporting Knapp’s conclusions that Daniel Schrön’s document described, indeed, what is now known as endometriosis, but also posits that references to symptoms typically associated to this disease have been present since Antiquity. In their exhaustive review, these authors also put forward the hypothesis that many of the reported cases of hysteria in women throughout the centuries might have been, in fact, cases of undiagnosed endometriosis. Both of these postulates have also been shared by the physician David Redwine. In the book “Googling Endometriosis: The Lost Centuries” (Redwine, 2012), outlining the history of this condition from Antiquity until the 20th century, Redwine concludes that there are references to **tiAw**, i.e. pain of

¹¹⁶ A few decades later, Bird et al. (1972) provided what is regarded as the ‘modern’ definition of adenomyosis: “the benign invasion of endometrium into the myometrium, producing a diffusely enlarged uterus which microscopically exhibits ectopic non-neoplastic, endometrial glands and stroma surrounded by the hypertrophic and hyperplastic myometrium.” (cited in Benagiano & Brosens, 2006).

the womb, in the Kahun Medical Papyrus, one of the oldest known medical texts, presumably from around 1825 BC¹¹⁷. Furthermore, and similarly to Nezhat et al., Redwine also believes that hysteria could have been used as a way to catalogue symptoms and conditions that physicians were not able to explain in women's health at that time.

Despite the controversy above, it is indisputable that accounts of crippling periods, excruciating pain, and infertility have been reported since ancient times. Has endometriosis been affecting women for millennia, being, as stated by (Redwine, 2012, p. iii), "an incredible master of disguise"? Or could these symptoms refer to other gynecologic conditions? There has been, in fact, an ongoing debate among the expert community concerning the nature of the concept of <Endometriosis>: given its complexity, does it denote a single clinical entity or rather more than one? And is <Adenomyosis> a type of <Endometriosis> or another, independent disease? Within the scope of the present research work, it is believed that the preceding questions do play a key role, due to their implications in the delimitation of the studied domain, as well as in the development of a conceptual representation of this subject field.

In the 1990s, (J Donnez et al., 1996; M Nisolle & Donnez, 1997) first argued that <Endometriosis> might encompass distinct clinical entities, each with their own pathogenesis: 1) peritoneal endometriosis, due to the transplantation of endometrial tissue, with the different lesion colors representing particular stages of the disease evolution (red > black > white, respectively); 2) ovarian endometriosis, caused by coelomic metaplasia of invaginated epithelial inclusions; and 3) adenomyotic nodules of the rectovaginal septum, or deep infiltrating endometriosis of the rectovaginal septum, due to metaplasia of Müllerian remnants in this anatomical structure (cf. Later work by (Donnez, 2012; Donnez et al., 2012; Nisolle, 2002). Other perspectives (García-Velasco & Rizk, 2010; Leyendecker, 2000) regard <Endometriosis> as a single entity. Moreover, the question that this approach does not comprise or explain the existence of

¹¹⁷ Overall, this document deals with several aspects of women's health, such as fertility, contraception, pregnancy, and gynecologic diseases. It is divided into 34 sections, each dealing with a specific problem and its respective diagnosis and treatment. References to tiAw - pain of the womb, which can be found in section 24, and the remaining translation into English of the existing fragments are available at: <http://bit.ly/2EfiXh>.

endometriosis lesions in more distant locations, such as the brain, lungs, or the umbilicus, can also be raised.

To sum things up, although the expert community agrees that endometriosis is, indeed, complex and enigmatic, it remains to be confirmed whether the different types of lesions and their underlying pathogenesis actually correspond to distinct clinical entities or, instead, to several clinical manifestations of a single disease. The key to untangle this complex web may lie in the future results of the recent WERF Endometriosis Phenome and Biobanking Harmonisation (EPHect) Project, which aims to provide standardized tools for the collection of biological samples from women with endometriosis, in order to contribute to “large-scale, cross-centre, epidemiologically robust research” into the causes, diagnosis, and treatment of this condition¹¹⁸. The data collection tools have already been published (Becker et al., 2014; Fassbender et al., 2014; Rahmioglu et al., 2014; Vitonis et al., 2014) and there are some expectations concerning this project, despite its considerable challenges (Ahn, Singh, & Tayade, 2017).

As regards the potential distinction between <Adenomyosis> and <Endometriosis>, the issue is still controversial within the community and remains, to a certain extent, inconclusive, as can be seen from the following excerpts taken from medical dictionaries, textbooks, and expert articles:

“Adenomyosis is a benign condition characterized by endometrial glands and stroma within the myometrium, accompanied by hypertrophy of the myometrium. Called also endometriosis interna or uterina.” [Dorland’s Medical Dictionary Online]

“Adenomyosis was believed to be part of endometriosis and was historically referred to as endometriosis interna. However, (...) endometriosis and adenomyosis are clinically two different entities. The only common feature is the presence of ectopic endometrial glands and stroma.” (Bedaiwy & Falcone, 2007).

¹¹⁸ Cf. <https://endometriosisfoundation.org/ephect/>

“Recent MRI studies have shown that adenomyosis and endometriosis are probably two entities governed by a single pathophysiological and/or genetic process.” (G. Levy et al., 2013)

“In the 1920s, two separate conditions were identified: endometriosis and adenomyosis, with different clinical profiles (...). More recently, however, similarities between the two pathologies have led to a re-evaluation of the situation and to a theory that they may have a common origin.” (Benagiano, Brosens, & Habiba, 2015)

If, on the one hand, some authors argue that <Adenomyosis> and <Endometriosis> are, indeed, two distinct clinical entities, others state that the former is a subtype of the latter. Hence, and in order to gain additional insight on this matter, information was also collected from the biomedical terminological systems explored in chapter 2, and in fact, the three resources that further subdivide the concept of <Endometriosis> (the Disease Ontology, ICD-10 and SNOMED CT) all include <Adenomyosis> or <Endometriosis of uterus> as a subtype of <Endometriosis> within a topological approach. Bearing all of this in mind, it was decided that within the scope of the EndoTerm project, <Endometriosis of uterus>, i.e. <Adenomyosis>, would be considered a subtype of <Endometriosis>.

Nevertheless, this section demonstrates that the issues surrounding the nature of this concept as a clinical entity are still far from being solved, which raises further problems concerning the disease classifications proposed and used by the expert community, as will be seen in the following section.

4.5. The dilemmas of endometriosis: what to classify and how?

Over the last few decades, several endometriosis classification and staging proposals have been put forward, in an attempt to create a common framework that would facilitate data comparison and research (cf. (Adamson, 2011, 2013; Adamson & Pasta, 2010)). That has proven, however, to be a challenging task, and, as will be shown further below, the current solution recommended by the latest expert consensus

(Johnson et al., 2017) comprises a somehow “modular assembly system” (cf. Haas, Shebl, Shamiyeh, & Oppelt, 2013) rather than a unified classification.

In the following paragraphs, a short historical overview concerning the aforementioned proposals will be described, with a stronger emphasis being placed on what are currently the most widely known and used resources within this scope: the Revised American Society of Reproductive Medicine (rASRM) classification of endometriosis (ASRM, 1997), the Enzian classification (Tuttlies et al., 2005), and the Endometriosis Fertility Index (EFI) (Adamson & Pasta, 2010).

After Sampson’s initial endeavor to categorize hemorrhagic cysts, the first to propose an endometriosis staging system were (Acosta et al., 1973), most notably concerning pelvic endometriosis. Based on the extent and location of the disease, it began being categorized it as **mild** (referring to “peritoneal disease with no ovarian involvement other than minimal superficial endometrial implants”), **moderate** (defined as “ovarian disease with or without peritoneal involvement”, plus the presence of ovarian endometriomas with up to 2 cm), and **severe** (comprising significant ovarian or tubal adhesions, peritoneal endometriosis, as well as ovarian endometriomas larger than 2 cm) (Acosta et al., 1973; Buttram, 1987; Canis et al., 1993).

Since these and other early proposals struggled to meet broad acceptance among clinicians and surgeons and given the urgent need to devise a standardized tool, the American Fertility Society (AFS), known today as the American Society for Reproductive Medicine (ASRM), introduced a classification system for endometriosis in 1979. This document was later revised in 1985 and 1996, the latest version being republished in 1997 and still commonly used today. As can be seen from **Figure 34** below, the rASRM is a weighted point system, with scores assigned to endometriosis lesions according to their size (< 1 cm; 1-3 cm; > 3 cm) and location (peritoneum and/or right/left ovary). Adhesions on ovaries and/or fallopian tubes are also scored, as well as the partial or complete obliteration of the cul-de-sac. The endometriosis lesions are also described as either superficial or deep, whereas the adhesions are characterized as dense or filmy. The total points result in 4 grades of severity: Stage I - **minimal** (1-5 points); Stage II - **mild** (6-15 points), Stage III - **moderate** (16-40 points); and Stage IV - **severe** (> 40 points). Accordingly, visual examples and guidelines referring to each of

these stages are provided in the second part of the document. It should be noted, though, that in the early 1990s, (Canis et al., 1992, 1993) suggested that a fifth, most severe stage of endometriosis should be considered in the former AFS classification, especially for patients with bilateral dense adhesions. That was never implemented, and the four-stage-levels are still used to this day.



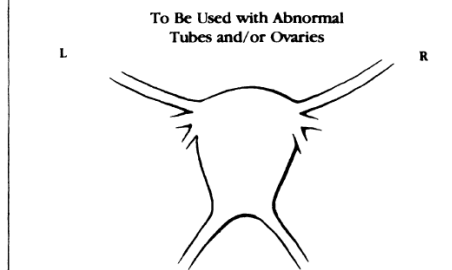
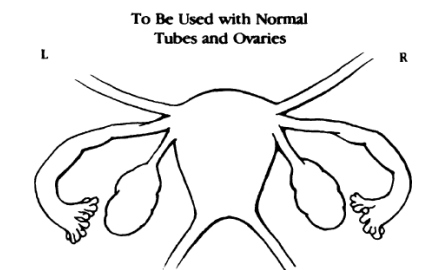
AMERICAN SOCIETY FOR REPRODUCTIVE MEDICINE REVISED CLASSIFICATION OF ENDOMETRIOSIS

Patient's Name _____ Date _____
 Stage I (Minimal) - 1-5 Laparoscopy _____ Laparotomy _____ Photography _____
 Stage II (Mild) - 6-15 Recommended Treatment _____
 Stage III (Moderate) - 16-40
 Stage IV (Severe) - > 40
 Total _____ Prognosis _____

PERITONEUM	ENDOMETRIOSIS	< 1cm	1-3cm	> 3cm
	Superficial	1	2	4
OVARY	Deep	2	4	6
	R Superficial	1	2	4
	Deep	4	16	20
	L Superficial	1	2	4
POSTERIOR CULDESAC OBLITERATION	Partial	Complete		
	4	40		
OVARY	ADHESIONS	< 1/3 Enclosure	1/3-2/3 Enclosure	> 2/3 Enclosure
	R Filmy	1	2	4
	Dense	4	8	16
	L Filmy	1	2	4
TUBE	Dense	4	8	16
	R Filmy	1	2	4
	Dense	4*	8*	16
	L Filmy	1	2	4
TUBE	Dense	4*	8*	16

*If the fimbriated end of the fallopian tube is completely enclosed, change the point assignment to 16.
 Denote appearance of superficial implant types as red [(R), red, red-pink, flamelike, vesicular blobs, clear vesicles], white [(W), opacifications, peritoneal defects, yellow-brown], or black [(B) black, hemosiderin deposits, blue]. Denote percent of total described as R____%, W____% and B____%. Total should equal 100%.

Additional Endometriosis: _____ Associated Pathology: _____



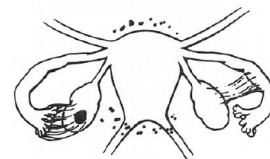
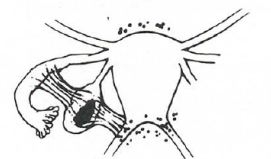
EXAMPLES & GUIDELINES		
STAGE I (MINIMAL)	STAGE II (MILD)	STAGE III (MODERATE)
		
PERITONEUM Superficial Endo - 1-3cm - 2 R. OVARY Superficial Endo - < 1cm - 1 Filmy Adhesions - < 1/3 - 1 TOTAL POINTS - 4	PERITONEUM Deep Endo - > 3cm - 6 R. OVARY Superficial Endo - < 1cm - 1 Filmy Adhesions - < 1/3 - 1 L. OVARY Superficial Endo - < 1cm - 1 TOTAL POINTS - 9	PERITONEUM Deep Endo - > 3cm - 6 CULDESAC Partial Obliteration - 4 L. OVARY Deep Endo - 1-3cm - 16 TOTAL POINTS - 26
STAGE III (MODERATE)	STAGE IV (SEVERE)	STAGE IV (SEVERE)
		
PERITONEUM Superficial Endo - > 3cm - 4 R. TUBE Filmy Adhesions - < 1/3 - 1 R. OVARY Filmy Adhesions - < 1/3 - 1 L. TUBE Dense Adhesions - < 1/3 - 16* L. OVARY Deep Endo - < 1 cm - 4 Dense Adhesions - < 1/3 - 4 TOTAL POINTS - 30	PERITONEUM Superficial Endo - > 3cm - 4 L. OVARY Deep Endo - 1-3cm - 32** Dense Adhesions - < 1/3 - 8** L. TUBE Dense Adhesions - < 1/3 - 8** TOTAL POINTS - 52 *Point assignment changed to 16 **Point assignment doubled	PERITONEUM Deep Endo - > 3cm - 6 CULDESAC Complete Obliteration - 40 R. OVARY Deep Endo - 1-3cm - 16 Dense Adhesions - < 1/3 - 4 L. TUBE Dense Adhesions - > 2/3 - 16 L. OVARY Deep Endo - 1-3cm - 16 Dense Adhesions - > 2/3 - 16 TOTAL POINTS - 114

Figure 34 – The rASRM classification of endometriosis (1997)

An additional feature of this 1997 revised version is the inclusion of images depicting the morphological appearance of the disease, i.e. examples of red, white, and black implants in each row, respectively (**Figure 35** below).

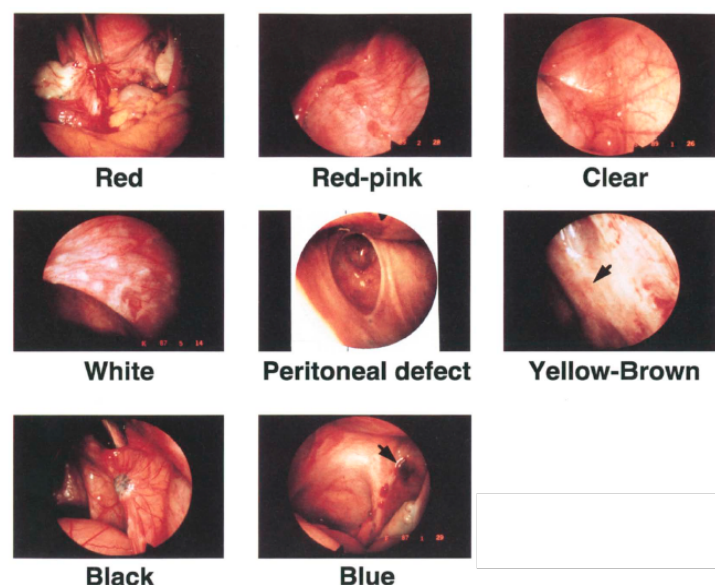


Figure 35 – Examples of endometriosis implants (ASRM, 1997)

Although it is still the most widely used endometriosis classification system among subject field experts today and is perceived as relatively easy to explain to patients, the rASRM has been heavily criticized due to its somewhat arbitrary scoring system, to the limited reproducibility of the staging, to its poor correlation with infertility and DIE and, essentially, to its lack of ability to effectively predict treatment outcomes (Adamson, 2011; Haas et al., 2013).

To address these limitations, other resources have been developed, but are yet to achieve more widespread use. One of them is the Enzian classification, originally proposed by (Tuttlies et al., 2005), firstly in German-speaking countries, and then worldwide, to supplement the rASRM concerning the description of DIE lesions. The document was further revised in 2010, 2011, and 2012 (SEF, 2012), with the latter version being shown below (**Figure 36**).

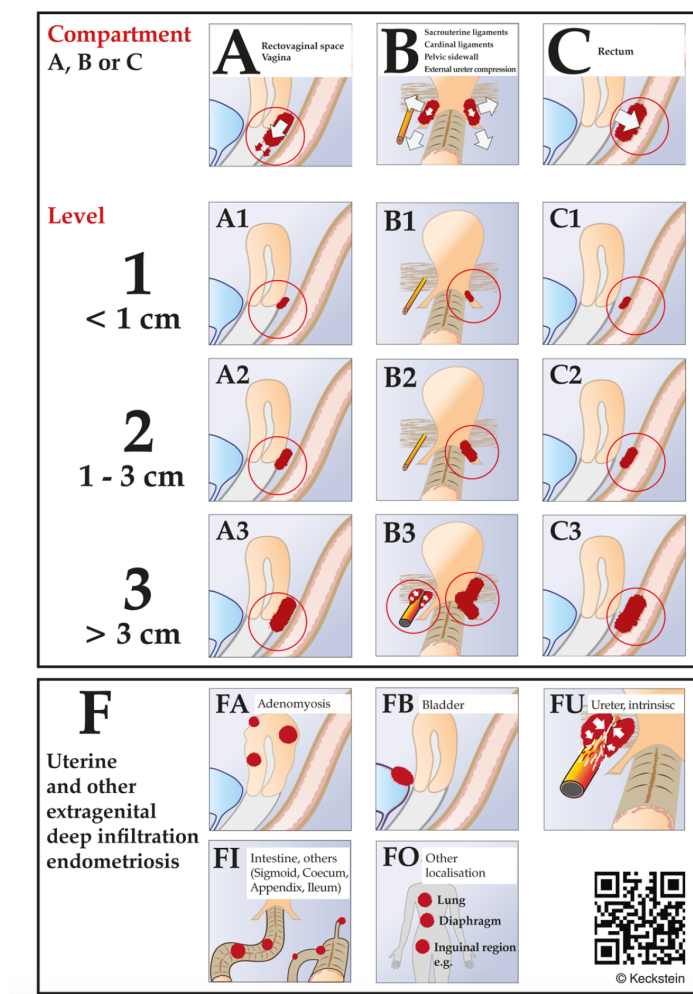


Figure 36 – The Enzian classification of Deep Infiltrating Endometriosis (SEF, 2012)

Generally speaking, this proposal combines morphological structures of the posterior pelvis into three 'compartments' where the appropriate surgical procedure(s) can be performed: Compartment A: comprises the recto-vaginal septum and vagina; Compartment B: includes the sacrouterine and cardinal ligaments, the pelvic sidewall, and also situations in which external ureter compression occurs; Compartment C: pertains to the rectum. Afterwards, each of the compartments above is graded regarding severity, based on the size of lesions. As depicted above, Level 1 refers to lesions under 1 cm, Level 2 to lesions between 1 and 3 cm, whereas Level 3 concerns lesions above 3 cm. Subsequently, and using a matrix-like structure, the classification is expressed as an alphanumeric string code, which may be shorter or longer depending on the complexity of the case (e.g. patient X can be assigned the Enzian classification A2, C1 if she has a 1-3 cm lesion in the recto-vaginal septum and vagina **plus** a < 1 cm lesion affecting the rectum).

In addition, the classification encompasses other DIE lesions described with the letter F, which stands for 'foreign/far' (Tuttlies et al., 2005), namely uterine lesions (i.e. adenomyosis) (FA) and/or other extragenital lesions, located in the bladder (FB); inside the ureter(s) (FU); in other parts of the intestine, such as the sigmoid, the coecum, the appendix and the ileum (FI); and in seemingly more unusual places, as the lungs, the diaphragm or the inguinal region.

Some reservations have been expressed concerning this classification, most notably the potential overlap with the rASRM and the complexity of the Enzian-related documentation, along with the difficulties in delimiting the stated compartments, especially in a realistic three-dimensional setting (Adamson, 2011; Haas et al., 2013; Tinneberg et al., 2014). Furthermore, and as mentioned before, the classification was predominantly used in German-speaking countries, which initially narrowed its international acceptance. However, this resource has been gaining ground and is now being taken into account in more recent initiatives (Johnson et al., 2017).

The third endometriosis classification that has been playing an increasingly important role among the expert community is the Endometriosis Fertility Index (EFI), proposed by Adamson & Pasta (2010). As the name indicates, it aims to estimate pregnancy rates in patients with surgically documented endometriosis who attempt a

non-IVF conception. In short, this resource consists of a form that should be completed in an endometriosis surgery where the patient still has expectations regarding fertility. The EFI can, therefore, provide useful support to treatment plans in infertile patients affected by endometriosis and results from the collection and analysis of real data, namely from 579 patients, with a subsequent validation involving 222 patients (ibid.).

As regards its overall structure, and as can be seen below (**Figure 37**), the EFI results from the total of two main factors - historical and surgical - and points are assigned to a set of sub-factors depending on the patient's situation. The historical factors comprise the patient's age at the time of surgery, the number of years of confirmed infertility, and whether or not there is a history of prior pregnancy, whereas the surgical factors include the so-called Least Function (LF) score, the patient's rASRM (former American Fertility Society (AFS)) Endometriosis score, as well as the patient's rASRM overall score. The LF score is, in turn, calculated by attributing a value to the level of functionality of the affected locations, namely the fallopian tubes, the fimbria¹¹⁹ and the ovaries (left and/or right). This level of functionality is expressed in a scale from 0 to 4, where 0 corresponds to absent or non-functional, and 4 to normal functionality. The resulting EFI total score, which ranges from 0 (poorest pregnancy prognosis) to 10 (best) can then be combined with the number of months after surgery (up until 36), to obtain an estimated pregnancy percentage. To sum things up, and just like Enzian, the EFI is not aimed to be a comprehensive endometriosis classification, but rather a complementary resource to support patients whose fertility options are still open.

¹¹⁹ The fimbria (or fimbriae, in the plural) is defined as "a bordering fringe, especially at the entrance of the fallopian tubes" (<https://www.merriam-webster.com/dictionary/fimbria>).

ENDOMETRIOSIS FERTILITY INDEX (EFI) SURGERY FORM

LEAST FUNCTION (LF) SCORE AT CONCLUSION OF SURGERY

Score	Description		Left		Right
4	= Normal	Fallopian Tube			
3	= Mild Dysfunction	Fimbria			
2	= Moderate Dysfunction	Ovary			
1	= Severe Dysfunction				
0	= Absent or Nonfunctional				

To calculate the LF score, add together the lowest score for the left side and the lowest score for the right side. If an ovary is absent on one side, the LF score is obtained by doubling the lowest score on the side with the ovary.

Lowest Score		+		=	
	Left		Right		LF Score

ENDOMETRIOSIS FERTILITY INDEX (EFI)

Historical Factors			Surgical Factors		
Factor	Description	Points	Factor	Description	Points
<u>Age</u>			<u>LF Score</u>		
	If age is ≤ 35 years	2		If LF Score = 7 to 8 (high score)	3
	If age is 36 to 39 years	1		If LF Score = 4 to 6 (moderate score)	2
	If age is ≥ 40 years	0		If LF Score = 1 to 3 (low score)	0
<u>Years Infertile</u>			<u>AFS Endometriosis Score</u>		
	If years infertile is ≤ 3	2		If AFS Endometriosis Lesion Score is < 16	1
	If years infertile is > 3	0		If AFS Endometriosis Lesion Score is ≥ 16	0
<u>Prior Pregnancy</u>			<u>AFS Total Score</u>		
	If there is a history of a prior pregnancy	1		If AFS total score is < 71	1
	If there is no history of prior pregnancy	0		If AFS total score is ≥ 71	0
Total Historical Factors			Total Surgical Factors		
EFI = TOTAL HISTORICAL FACTORS + TOTAL SURGICAL FACTORS:				+	
			Historical		Surgical
			=		
			EFI Score		

ESTIMATED PERCENT PREGNANT BY EFI SCORE

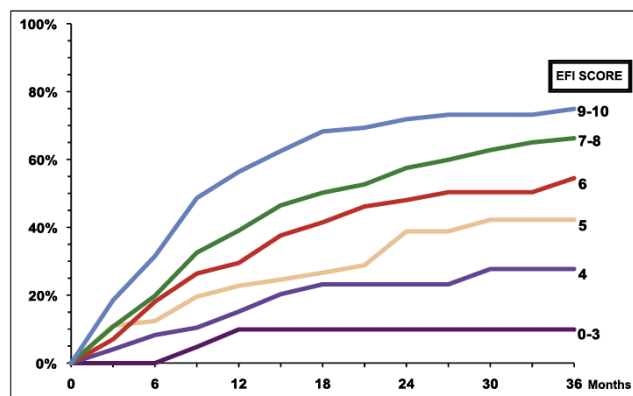


Figure 37 – The Endometriosis Fertility Index (Adamson & Pasta, 2010)

Due to the complexity that characterizes the disease and which, according to (Khazali, 2016), “defies a single all-encompassing classification system”, the most recent consensus on the classification of endometriosis, promoted by the World Endometriosis Society (WES) (Johnson et al., 2017) and involving various stakeholders from 19

countries, including patient organizations, recommended surgeons to use a “classification toolbox” based essentially on the rASRM and supplemented with either Enzian - in cases of women with DIE - or the EFI - in cases of women for whom fertility is a concern -, or both, whenever necessary. The consortium members also proposed a wider use of the World Endometriosis Research Foundation (WERF)’s EPHeC data collection tools, particularly when surgery does not take place.

In short, there is currently no standardized classification of the disease that can accurately reflect its complex nature. The resources that have been more exhaustively described throughout this section are mainly based on the location of the lesions and on their size and/or depth, although in the EFI classification, factors such as the functionality of the affected locations, as well as the patient’s age, previous pregnancy/-ies, and years of infertility, are also taken into account. Furthermore, there are still several questions concerning this pathology that require further clarification, namely, for instance, the boundaries between <Endometriosis> and <Adenomyosis>, and whether or not DIE should have its classification (Koninckx et al., 2011). Thus, it is unlikely that an ideal endometriosis classification system as the one advocated by (Adamson, 2011)¹²⁰ will be achieved any time soon. However, such a classification is urgently needed, both from a medical and from a patients’ standpoint (Haas et al., 2013), so it is very likely that this topic will continue to be a major concern, not only within the expert community but also among the various stakeholders.

¹²⁰ According to Adamson (2011), an ideal endometriosis classification system should be empirically and scientifically based; be comprehensive, useful, simple to use and easy to communicate with patients; enjoy wide consensus within the community; **have unambiguous definitions of terms**; and predict fertility and pain relief, as well as estimate the risk of recurrence (the segment was placed in bold by the author of this thesis).

Chapter 5 – An overview of EndoTerm

As referred to in the introduction, the third and final part of the thesis aims to describe the development and implementation of EndoTerm, a knowledge-based terminological resource focused on endometriosis (cf. (Carvalho et al., 2016a, 2016b, 2015)). As was shown in chapters 3 and 4, this medical condition is still underrepresented in biomedical terminological resources, despite its worldwide prevalence and significant economic and social impact. Therefore, EndoTerm seeks to contribute to a more effective way of organizing and sharing the existing knowledge in this particular subject field. Thus, the current chapter will start by providing an overview of the proposed resource, its objectives, examples of potential use cases, and target users, together with its underlying theoretical and methodological principles.

The following chapter will start by describing how subject field knowledge was acquired by the terminologist and will supply a first attempt to systematize the gathered information via a domain tree. At the core of that chapter is EndoTerm's conceptual backbone and the different steps that led to the conception of a semi-formal conceptualization, supported by non-hierarchical conceptual relations and a categorial structure.

Chapter 7, in turn, will be devoted to the linguistic dimension of this project, particularly to the collection, processing, and analysis of a specialized corpus devised especially for this purpose – the EndoCorpus. The final chapter of this third part aims to illustrate, via a case study involving a specific type of <Surgical procedure>, how the conceptual and the linguistic dimensions can interrelate and contribute to mutual enrichment, with special emphasis being placed on the drafting – or enhancement – of natural language definitions. Furthermore, Chapter 8 will show how EndoTerm content can be shared with both human- and machine-users, with the latter being in line with current trends in healthcare and biomedical terminological systems as regards interoperability.

Bearing in mind the theoretical background referred to in Chapter 2, this project's methodology seeks to integrate what is conceptual and what pertains to natural language by interrelating both dimensions throughout different stages of the work. Hence, the project relies on a mixed methodology underlying terminology work (both onomasiological and semasiological) (Santos & Costa, 2015), although the order is not arbitrary (*ibid.*). For knowledge organization purposes, such as those that help ground this research work, it is believed that a concept-based approach may constitute an adequate starting point.

On the one hand, a systematized conceptualization proposal resulting from: i) domain knowledge acquired through the contact with subject field experts, endometriosis-related texts and several multimedia resources; and ii) the analysis of endometriosis-related information available in biomedical terminological systems (*cf.* Chapter 3) was put forward **before** the actual corpus analysis (a top-down approach). At a later stage, the devised conceptual framework accompanied, albeit in the background, the corpus analysis stage, and was also used to support the drafting of natural language definitions. Hence, this resource can be placed under a knowledge-based approach to Terminology, following the notion put forward by (Meyer et al., 1997), in the sense that firstly, EndoTerm is grounded on subject field knowledge and, secondly, the data has been encoded as a “knowledge base” on an AI perspective (*ibid.*), since this resource relies on several micro-concept systems where the various conceptual relations are made explicit.

On the other hand, several texts were used within the scope of the current project, in particular stages and for different purposes: the terminologist initially consulted a series of popularized resources in order to better grasp the disease itself; next, more specialized resources (e.g. guidelines, scientific articles, textbooks, etc.) were accessed and analyzed, in order to help the terminologist get a more thorough understanding of how the expert community perceives the domain; and finally, due to the lack of pre-built corpora on the topic of endometriosis, EndoCorpus - a monolingual (English) specialized corpus – was created, in order to cater for the needs of the project. Following the semi-automatic corpus processing stage, the corpus analysis was carried out, supported by Meyer's notion of Knowledge-Rich Contexts (KRC) (Meyer, 2001), to

ascertain how – or if indeed – certain pre-defined conceptual information could be elicited from the EndoCorpus. The resulting semantic network, built from the corpus data, was, in turn, compared to the previous conceptualization, thus contributing to a new, enhanced, iteration of the latter (a bottom-up approach). Afterwards, the enhanced conceptual representations provided a framework of interrelated concepts, relationships, and domain constraints, and hence, an organized, clear, and consistent basis for the drafting of natural language definitions. This conceptualization was tested and validated via TeDI, or OntoTerminology Editor, a tool developed by Roche (2017), so as to verify the logical properties of the proposed conceptual systems and to enable the aforementioned operationalization.

Under this methodological framework, experts played a critical role by helping the terminologist identify the sources to be used, both in the domain knowledge acquisition stage and in the corpus collection stage. Regrettably, and although the validation of candidate terms and natural language definitions by subject field experts had been included as part of the initial project, in line with what has been proposed by (Costa et al., 2011, 2013, 2012; Silva, 2014), this was not possible to implement within the timeframe of this thesis.

In short, both the concept and the term permeated the whole project, giving rise to a tangled web of conceptual and linguistic information that aims to support better and more effective knowledge representation, organization and sharing within the domain of endometriosis. This is also illustrated by EndoTerm's core structure, composed of two modules that function in an autonomous, yet interrelated, way: [1] the conceptual module, described in Chapter 6 and conceived as EndoTerm's 'beating heart'; and [2] the linguistic module, explored in Chapter 7 and which aims to enrich the knowledge base with terms and their respective synonyms, when applicable. Finally, the linguistic module also comprises natural language definitions, either drafted or enhanced according to definitional templates grounded in EndoTerm's conceptual core structure and involving data from both the specific difference approach and the categorial structure. Despite this modular structure, which could misleadingly transmit the idea that EndoTerm would be grounded in two separate entities with no apparent connection, this project aims to demonstrate – especially through the case study to be

explored in Chapter 8 – that terminology work does, indeed, benefit from a methodology that can effectively integrate both dimensions.

As far as the use cases for EndoTerm are concerned, it is believed that this resource could make a valuable contribution at various levels, namely by supporting:

[1] the creation of other terminological and lexicographic resources dedicated to endometriosis, such as an online multilingual dictionary of endometriosis, which does not exist at the moment;

[2] the drafting of new (or the enhancement of existing) natural language definitions in this subject field, which may subsequently be incorporated not only in biomedical TS (a feature that they are currently lacking, as shown in chapter 3) but also in the aforementioned lexicographic and terminological resources;

[3] the development of initiatives integrated in the eHealth paradigm, e.g. training materials, in addition to other services targeted at the digital health literacy of patients and society in general;

[4] the optimization of information retrieval, namely by allowing human users to search for surgical procedures by approach, for instance;

[5] the aspired interoperability in this area, by resorting to W3C-compliant formats (RDF/XML) that follow the current trend, within the biomedical subject field, of making more and more content available via RDF and Linked Data;

Taking the aforementioned background into account, it is believed that the products supported by EndoTerm's core structure and content could mainly be destined to future healthcare professionals, to other healthcare professionals who wish to know more about the disease (e.g. nurses, physiotherapists, medical imaging physicians and technicians, General Practitioners, etc.), as well as to expert patients, patients, and respective caregivers. On the other hand, and with the nearly ubiquitous presence of technology and IT-related industries in healthcare, the end-user of medical terminological resources may no longer be exclusively human, which is why an example of EndoTerm content in RDF/OWL will also be provided. Lastly, due to its conceptual foundation and connection to current biomedical TS, EndoTerm's data may also be useful for medical coders and librarians.

Chapter 6 - EndoTerm's conceptual backbone

"A message to mapmakers: highways are not painted red, rivers don't have county lines running down the middle, and you can't see contour lines on a mountain."

(Kent, 2012, p. 27)

6.1. Getting to know the subject field: the essential first step in terminology work

Given the importance of getting to know the chosen subject field, especially at the onset of any terminology project, as advocated by (Cabr , 1999; Felber, 1984; ISO, 2009; J. Sager, 1990; Suonuuti, 2001), this subsection will address how the subject field content concerning this particular disease was gathered and systematized by the terminologist. Firstly, contacts were established with Portuguese subject field experts - senior gynecologists and surgeons specialized in the disease -, with one expert patient - more specifically, a EUPATI Fellow¹²¹ who has been working in this field - and with patient associations, namely the Portuguese Association for Women with Endometriosis (MulherEndo¹²²) and the North-American Endometriosis Association¹²³. Ultimately, those contacts also enabled the terminologist to identify some of these stakeholders' needs as regards the representation, organization, and dissemination of knowledge about this disease. On the other hand, the contact with experts at the onset of the work played a decisive role, not only due to their valuable input concerning the disease but also to their very helpful guidance as regards the resources to be consulted by the terminologist to better grasp the domain.

However, different sources were used in different stages of the work and for various purposes. Hence, and following expert recommendations, the very first resources explored by the terminologist were more generic and aimed to provide an overview of the disease itself to non-experts - in particular to patients with endometriosis - using popularization. The resources that were gathered and analyzed at this first stage are summarized below (**Figure 38**):

¹²¹ Cf. Chapter 1 for more information on EUPATI.

¹²² <http://mulherendo.pt/>

¹²³ <http://www.endometriosisassn.org/>

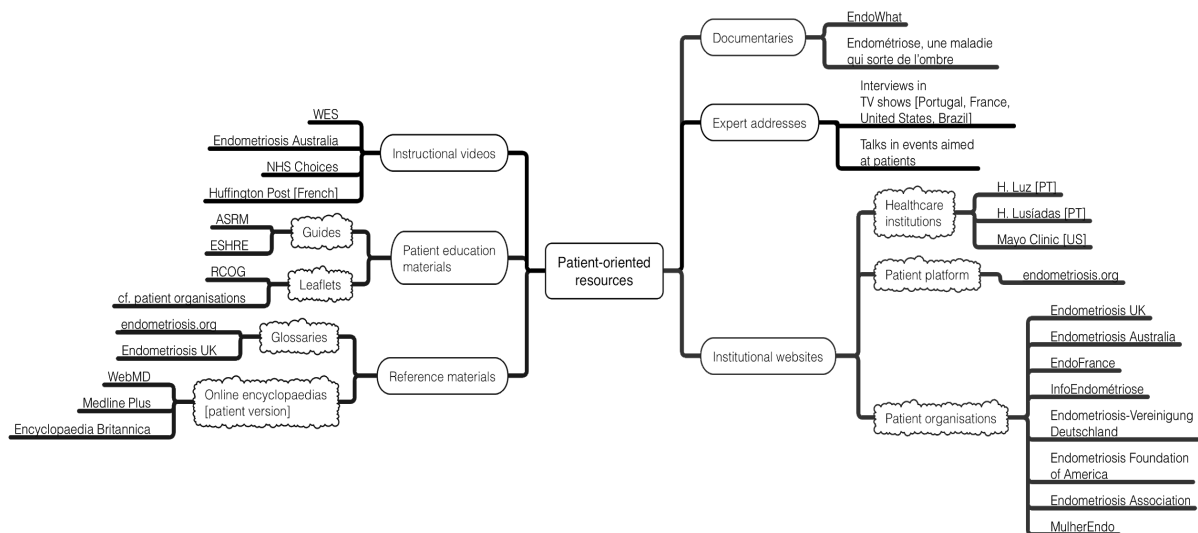


Figure 38 – Patient-oriented resources on endometriosis

Afterwards, and also following expert suggestions, more specialized resources were accessed¹²⁴, analyzed and categorized by MeSH's Publication Types [PT]¹²⁵, used to index medical literature (**Figure 39**). It should be noted that the preceding resource analysis did not entail corpus processing at this point, but rather aimed to help the terminologist get a more thorough understanding of how the expert community perceives the domain.

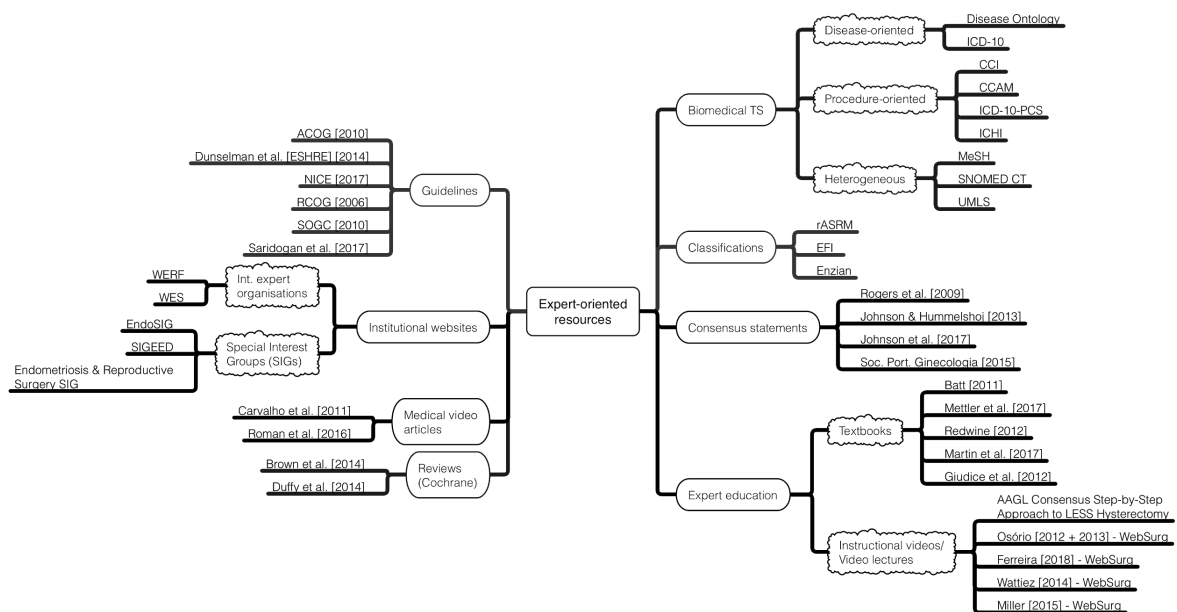


Figure 39 – Expert-oriented resources on endometriosis

¹²⁴ Since several of these resources require a subscription, only those that were freely available were consulted.

¹²⁵ The full list of MeSH's PT is available at <https://www.nlm.nih.gov/mesh/pubtypes.html>.

Based on the gathered information, particularly from the second set of resources, a preliminary systematization was put forward in the form of a **domain tree**. As shown in **Figure 40**, the purpose was mainly to represent the information schematically, to get an overview of the main areas that an endometriosis-related resource would encompass. Since this mind map did not aim to comprise terms, concepts, or relations, but merely to provide that aforementioned general outline, no specific or distinctive typographic features were included¹²⁶.

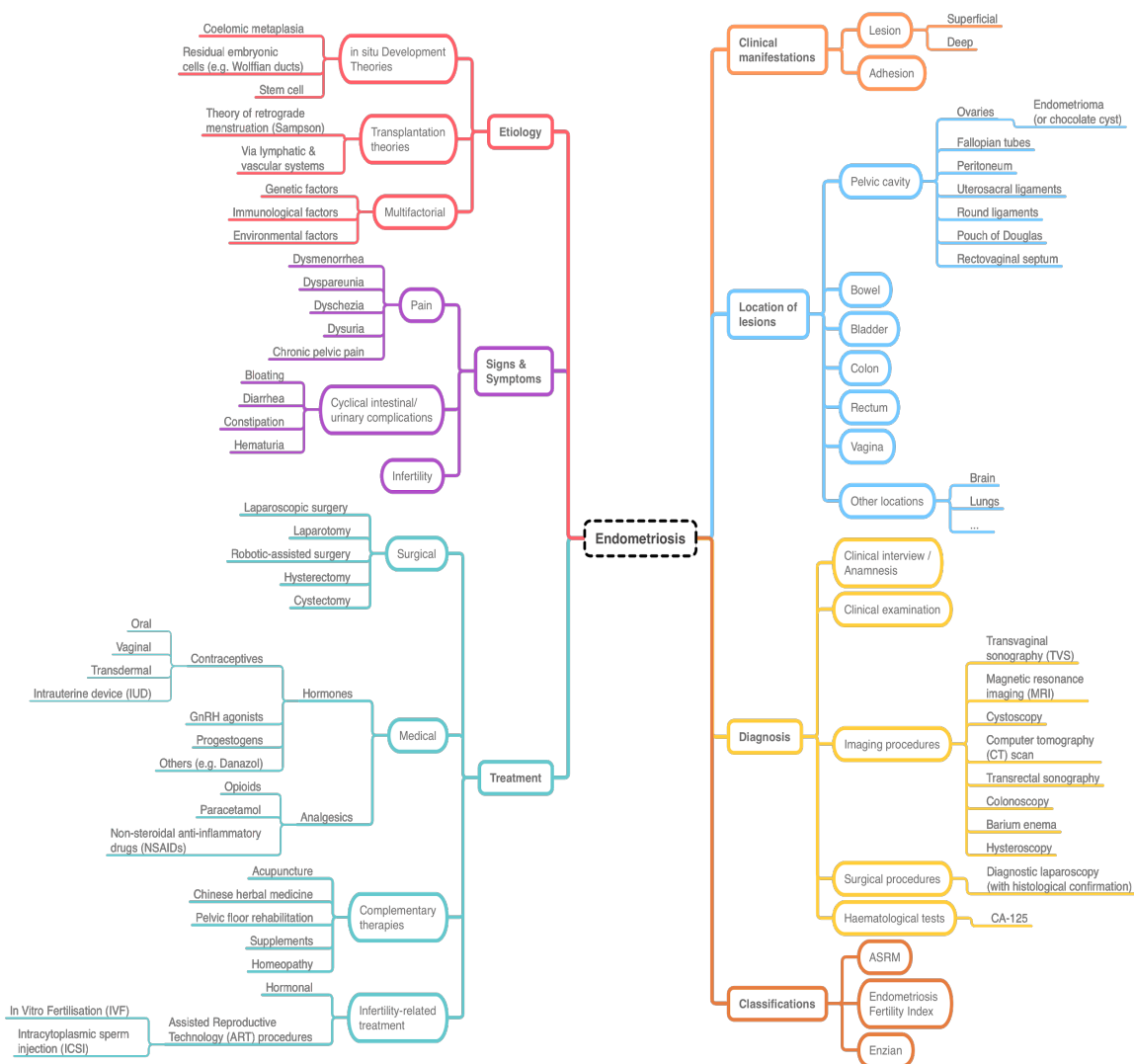


Figure 40 – Proposal of a domain tree for Endometriosis

¹²⁶ That will not be the case in the upcoming conceptualization proposals, which will resort to capitalization, punctuation marks, as well as colors, to distinguish between concepts, terms, and relationships (and, in the latter, between hierarchical and non-hierarchical relations).

Such a representation constituted a valuable starting point to display the enigmatic and intricate nature of endometriosis, as well as its broad scope as a subject field, which were shown in the previous chapter. Its social and economic relevance nowadays is unequivocal, as is the lack of knowledge that persists among the expert community, which constitutes one of the causes of a substantial diagnostic delay. The complexity that characterizes this disease also hampers a general consensus regarding its essence as a concept, and also concerning its classification and staging, with consequences in the way data, information, and knowledge about endometriosis are represented, organized, and shared. It is expected that the EndoTerm project can contribute, on the one hand, to systematize the current knowledge regarding this condition and, on the other hand, to enhance its yet rather marginal presence in existing biomedical terminological resources.

6.2. Identifying EndoTerm's concept of departure or finding the needle in the haystack

The ability to conceptualize is quintessential for humans as a species, as it enables us to better grasp whatever surrounds us and, in turn, organize that information to create and share knowledge. However, it is no easy task. As alluded to in the opening quote to this chapter, when one is creating a map (and, in this case, a knowledge organization proposal could be perceived as such), the paths, tracks, mountains and rivers (i.e. the concepts and their respective relationships) may already be there, but they are not fully 'visible' (as a whole) and connected yet. It is the mapmaker – the terminologist – who has to undertake that task of being able to see both the individual trees and the entire forest and then share that with everyone else.

Hence, and after gathering the necessary knowledge concerning <Endometriosis> and devising a systematization proposal via a domain tree, it was decided that a micro-concept-system-based structure would best suit EndoTerm's conceptual foundation. Subsequently, a survey of categories¹²⁷ in existing biomedical TS was carried out to ascertain which could be matched to each of the main areas directly

¹²⁷ In the various standards applied to the field of biomedical informatics (e.g. ISO 2007, 2012, in particular), the notion of 'category' is regarded as a synonym of 'generic concept' as defined in the ISO 1087-1 (2000).

linked to Endometriosis in the previous domain tree, with the goal of better substantiating the upcoming conceptualization process. For this purpose, both the UMLS's Semantic Types and SNOMED CT's Top-Level Hierarchy (cf. Chapter 3) were chosen, since these resources are widely regarded as the most comprehensive biomedical TS today, encompassing not only disease-related concepts but also those referring to procedures, substances, anatomy, devices, among many others. In light of the overall results of this categorization, which can be consulted below (**Table 15**), it was decided that EndoTerm's core micro-concept systems would focus on <Procedure>¹²⁸ concepts, particularly those related to the **surgical** treatment/management of endometriosis (identified in the UMLS as Therapeutic or Preventive Procedures).

TOP-LEVEL CONCEPT [SNOMED CT]	SEMANTIC TYPE [UMLS]	EXAMPLE FROM THE DOMAIN
Body structure (body structure)	Body Part, Organ, or Organ Component	<Ovary>
Clinical finding (clinical finding)	Sign or Symptom	<Dyschezia>
Disease (disorder)	Disease or Syndrome	<Endometriosis>
Morphologic abnormality (morphologic abnormality)	Finding	<Lesion>
	Neoplastic Process	<Endometrioma>
	Pathologic Function	<Adhesion>
Pharmaceutical/biologic product (pharmaceutical/biologic product)	Pharmacologic Substance	<Non-steroidal anti- inflammatory drug>
Physical object (physical object)	Medical Device	<Intrauterine device>
Procedure (procedure)	Diagnostic Procedure	<Cystoscopy>
	Therapeutic or Preventive Procedure	<Laparotomy>
Substance (substance)	Organic Chemical Hormone Pharmacologic Substance	<Progestogen>

Table 15 – EndoTerm's potential core categories

¹²⁸ It should be noted, once again, that EndoTerm concepts are capitalized (e.g. <Endometriosis>) and written between single chevrons, whereas terms are presented in lower case and between double quotation marks (e.g. "ovarian endometriosis"). As for the conceptual relations, they are represented in bold and with an underscore separating their components, when there is more than one (e.g. is_a).

The choice of <Surgical procedure> and related concepts as the basis of EndoTerm's micro-concept systems is supported by three main criteria: firstly, surgery is one of the fields where the current technological innovation has become more prominent, with the advent of cutting-edge, less invasive techniques, supported by increasingly smaller, yet highly effective medical devices (cf. Menvielle, Audrain, & Menvielle, 2017; Schlich & Crenner, 2017)¹²⁹; secondly, and to quote Jim Yong Kim, the President of the World Bank, surgery constitutes an “indivisible, indispensable part” of health care systems (cited in (Rose et al., 2015)), in such a way that a number of recent global initiatives - among them the Global Surgery 2030 report (Meara et al., 2015) and the World Health Assembly 68/15 Resolution, from May 2015 (WHA, 2015) - aim at contributing to a substantial improvement in the worldwide access to emergency and essential surgical care, thereby making it an essential component of universal health coverage (UHC)¹³⁰; and finally, surgery has been playing an increasingly fundamental role in the effective management of endometriosis, especially when combined with a medical approach, as stated by the several guidelines referred to in chapter 4. In fact, the combination of a diagnostic laparoscopy and the subsequent histological confirmation remains the gold standard as far as the diagnosis is concerned (Johnson & Hummelshoj, 2013; Kennedy et al., 2005; Rogers et al., 2017). Moreover, and despite the lack of accurate numbers at a global level, there seems to be a growing tendency towards minimally invasive surgical approaches.

Considering all of the above, the number of surgical procedures is expected to rise considerably in the upcoming years, with a visible impact not only on the way surgical data, information, and knowledge are represented and organized, but also on how these will then be shared and used in the interactions involving: i) healthcare

¹²⁹ Over the last few years, several postgraduate courses in Surgical Innovation have also been created in well-known universities worldwide. Some examples include the MSc in Surgical Innovation at the Imperial College London (<https://bit.ly/2Jh0EUO>) and, in the United States, the Duke Surgical Education and Innovation Program (<https://bit.ly/2HTinCn>) or the Harvard Surgical Program in Innovation (<https://bit.ly/2HCxH8T>).

¹³⁰ In the Global Surgery 2030 report, the Lancet Commission on Global Surgery defines a target of 80% coverage of essential surgical and anaesthesia services per country by 2030, as well as an increase in the overall surgical volume, reaching approximately 5,000 procedures per 100,000 people, particularly in low- and middle-income countries (or LMICs) (Meara et al., 2015).

providers (through EHRs, clinical research, quality assurance, resource management, etc.); ii) providers and other stakeholders (e.g. insurance companies, particularly concerning reimbursement procedures); iii) patients and providers (via PHRs or some of the health and fitness apps mentioned in Chapter 1); iv) and, last but not least, patients themselves (through crowdsourcing platforms and other social media-based communities).

But what is a <Surgical procedure>? According to Dorland's Medical Dictionary, it is "any act performed with instruments or by the hands of a surgeon"¹³¹, whereas the Merriam-Webster Medical Dictionary defines it as a "procedure performed on a living body usually with instruments for the repair of damage or the restoration of health and especially one that involves incision, excision, or suturing"¹³². For the OECD, on the other hand, it includes "all invasive therapies performed as inpatient surgery", with the latter being characterized as a surgical procedure that implies "an overnight stay in an inpatient institution"¹³³. However, and as will be further explored in this chapter, not all <Surgical procedures> necessarily require an overnight stay at the hospital (as, for instance, ambulatory surgery) or, indeed, the existence of an external skin incision. In fact, surgery is becoming less and less invasive over time. Other <Surgical procedures> may not involve excision, whereas others may not even be performed directly "by the hands of a surgeon", but via robotic assistance.

Since the preceding definitions have proven to be insufficient to accurately define the concept of <Surgical procedure>, and a general consensus is difficult to attain, as "there are neither specific criteria to define it nor specific criteria to define the limit between what a surgical procedure is and what it is not" (ISO 1828, 2012: v), another approach was pursued, namely the analysis of concept-based, procedure-related terminological resources, whose core principles were introduced in Chapter 3. It was believed these could complement the previous natural language definitions and help shed some light on the nature of a <Surgical procedure>, thereby supporting the development of a foundational, semi-formal conceptual system for EndoTerm.

¹³¹ <https://www.dorlandsonline.com> (14.02.2018).

¹³² <https://www.merriam-webster.com/dictionary/operation#medicalDictionary> (03.02.2018).

¹³³ <http://stats.oecd.org/glossary/detail.asp?ID=2617> (14.02.2018)

Hence, and in addition to **SNOMED CT**, widely used to represent different types of healthcare procedures in EHRs¹³⁴, other resources focusing entirely on <Procedures> were consulted and analyzed, namely the WHO's **ICHI**, the **ICD-10-Procedure Coding System (PCS)**, the **Canadian Classification of Health Interventions (CCI)**, and the **Classification Commune des Actes Médicaux (CCAM)**. While the ICHI seeks to achieve worldwide coverage (ICHI), the other three resources have a more national scope.

Currently in its Beta version - its full release is expected by the end of 2018 -, the ICHI has been developed as the third axis of the so-called WHO Family of Health Classifications (WHO-FIC), together with the ICD and the International Classification of Functioning, Disability and Health (ICF). Its international scope aims to: i) facilitate international comparisons of healthcare interventions across countries; ii) provide a standardized classification to countries which lack such a resource; iii) help complement existing national classifications; iv) support the monitoring, financing, and implementation of programs that can foster global healthcare and well-being programs (e.g. the UN's Sustainable Development Goals or the WHO's Universal Health Coverage action); and v) finally, contribute to a more effective performance assessment of healthcare interventions by supplying a common framework for analysis.

The ICD-10-PCS is the official system in the United States for coding inpatient (i.e. someone who has been admitted to a hospital) procedures, created for the Centers for Medicare and Medicaid Services (CMS)¹³⁵. Used mostly by health insurers for billing purposes, this classification has gradually replaced the ICD-9-CM (Clinical Modifications) - Volume 3, as it provides a more "flexible organization of concepts" and a framework that allows human users to assign and find codes more easily (Steindel, 2010). The CCI¹³⁶ has been developed and maintained by the Canadian Institute for Health Information to accompany the ICD-10-CA (the Canadian extension), whereas the CCAM¹³⁷ has been used, since 2005, as the *de facto* reimbursement classification for clinical procedures in France, mainly in hospitals.

¹³⁴ It should be noted that with the exceptions of SNOMED CT and the UMLS, the TS explored in Chapter 3 were mainly disease-oriented, and therefore, unsuitable for this particular stage of the work.

¹³⁵ Cf. <https://www.cms.gov/Medicare/Coding/ICD10/index.html>

¹³⁶ Cf. <https://www.cihi.ca/en/version-2018-icd-10-ca-and-cci-classifications>.

¹³⁷ The full classification is freely available online at <https://www.ameli.fr/accueil-de-la-ccam/index.php>.

Except SNOMED CT, all of the remaining procedure TS are classifications, supported by a multiaxial (and mainly hierarchical) core structure that becomes visible in their coding principles. While SNOMED CT's concept identifiers are non-semantic, the codes used in the ICHI, the ICD-10-PCS, the CCI and the CCAM have an intrinsic meaning, with each character of that code representing a given axis of classification that provides complementary information regarding the procedure in question.

The table below (**Table 16**) contains examples of each of the procedure TS above, specifically regarding the concept of <Total abdominal hysterectomy>, a type of <Hysterectomy>¹³⁸ in which the uterus and the cervix are completely excised¹³⁹, in this case via an abdominal approach, i.e. a laparotomy¹⁴⁰. As will be shown further ahead, <Hysterectomy> was the chosen example to illustrate EndoTerm's conceptualization proposal.

Procedure TS	Procedure	Code	WHERE? > Target	WHAT? > Type of Intervention; Action	HOW? > Approach
CCAM	Hystérectomie totale, par laparotomie	JKFA015	J Appareil urinaire et génital, procréation JK Utérus non gravide	F Exciser	A Abord ouvert
CCI	Excision total, uterus and surrounding structures	1.RM.89.LA	RM Therapeutic Interventions on the Uterus and Surrounding Structures	1 Physical/Physiol. Therapeutic Interventions 89 Excision total	LA using open approach
ICD-10-PCS	1. Resection of Uterus, Open Approach 2. Resection of Cervix, Open Approach	1. 0UT90ZZ 2. 0UTC0ZZ	OU Female reproductive system 9 Uterus C Cervix	0 Medical & surgical T Resection	0 Open approach Z No Device Z No Qualifier
ICHI	Total abdominal hysterectomy	NME JK AA	NME - Uterus	JK - Excision, total	AA - Open approach
SNOMED CT	Total abdominal hysterectomy	116143008	Procedure site - Direct: Entire uterus	Method: Excision - action	Surgical approach: Abdominal approach

Table 16 – The example of <Total abdominal hysterectomy> in current procedure TS

¹³⁸ Hysterectomy is a rather common type of surgical procedure in Gynecology, as stated before, and is often seen as a last resort in cases of severe endometriosis (Rogers et al., 2009, 2017).

¹³⁹ <https://www.dorlandonline.com/dorland/definition?id=100051982> (03.02.2018).

¹⁴⁰ In a laparotomy, an incision is made into the abdominal cavity, in such a way as to allow the surgeon to have a naked-eye view of the procedure site.
(<https://www.dorlandonline.com/dorland/definition?id=100057476> - 03.02.2018)

From a structural perspective, the various TS seem to be grounded on three core axes, systematized in the table's header row and in the figure below (**Figure 41**):

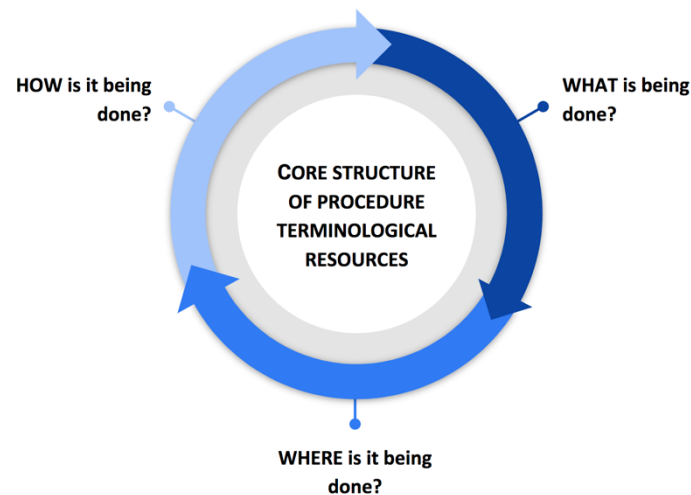


Figure 41 – The three axes of a <Surgical procedure>

1) the type of intervention or the type of action that is being / will be carried out: **WHAT?** 2) the location or target of that procedure in the human body: **WHERE?** 3) the approach to perform the procedure in question - in this case, “the technique used to reach the procedure site” (Butler et al., 2012) or the “mode d’accès” (ATIH, 2017, p. 36) - “type de pénétration de l’instrument dans l’organisme et le procédé technique employé”¹⁴¹, i.e. : **HOW?**

However, it also becomes clear that these procedure TS entail some of the problems already alluded to in Chapter 3 regarding biomedical TS in general. The first one refers to the wide variability of the collected data, particularly in what concerns the notion of <Procedure> and the types of healthcare interventions included. Due to significant differences concerning their foundational hierarchies - some related to the

¹⁴¹ This should not be confused with the CCAM’s “Voie d’accès”, i.e. “lieu de pénétration dans l’organisme d’un instrument pour la réalisation de l’acte” (p. 37). The former is the way a given instrument enters the body for a surgical procedure (i.e. via a small incision, a puncture, or a natural body orifice, for instance), whereas the latter is the place, in the body, where that instrument is inserted. For instance, an appendectomy and a thoracotomy share the same “mode d’accès”, but not the “voie d’accès”.

axis of classification chosen to ground the resource in, others regarding distinct levels of granularity -, comparisons across procedure TS can be as challenging as those concerning disease-oriented TS. As regards the first issue - the prime axis -, the various codes in the preceding table show very clearly that the initial core sections in ICD-10-PCS and CCI are based on the type of procedure (Action axis) (e.g. diagnostic, therapeutic, monitoring, or even administrative procedures), whereas CCAM and ICHI, for instance, are grounded in the location of the procedure (Target axis). On the other hand, SNOMED CT depicts a mixture of all three axes but is organized mainly by Action and Target. A quick consultation of the online browsers of the ICHI and SNOMED CT using the concept of <Hysterectomy>, for instance, confirms those discrepancies regarding the core axis, as well as the granularity (**Figure 42**).

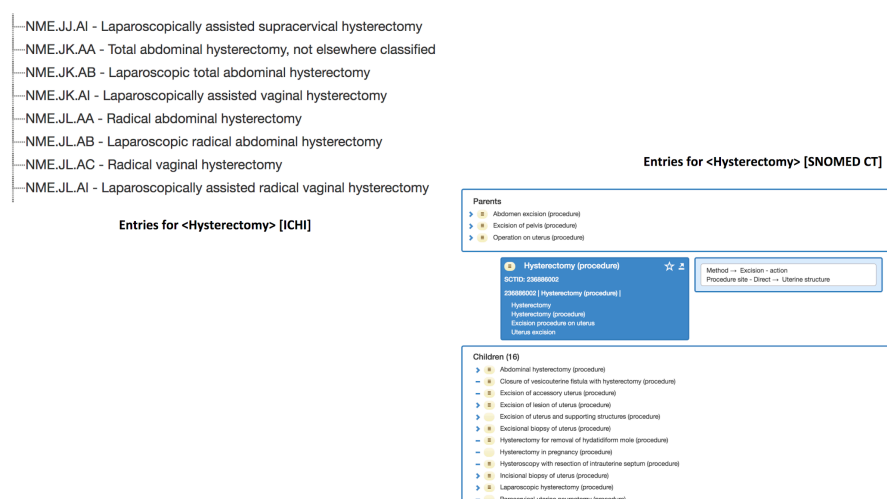


Figure 42 – The example of <Hysterectomy> and its sub-hierarchies (ICHI and SNOMED CT)

The second problem lies on the coexistence of several - and often incompatible - supporting classification schema and formats and the subsequent lack of interoperability. In fact, hardly any of the consulted procedure TS and their freely accessible browsers currently supports the exporting of their data in W3C-compliant formats. In SNOMED CT, the results are not directly exportable, but a local OWL version

can be generated by the user via the Perl script, and the ‘Stated Relationships’ file is available in the standard distribution of the terminology¹⁴².

In short, obtaining systematized data, information and knowledge on surgical interventions is not an easy task. Furthermore, and contrary to disease-oriented TS, where the ICD is regarded as the *de facto* standard, there is no unequivocal, standardized procedure classification system for surgical procedures, which makes the task of effectively representing, organizing, and sharing the related DIK even more challenging. Therefore, to achieve a certain degree of standardization and to facilitate semantic interoperability, the 1828: 2012 ISO standard was published, introducing a categorial structure (CAST) for terminological systems of surgical procedures. As referred to in Chapter 3 and as shown in the table below (**Table 17**), there are some pre-defined categories and relationships in the ISO 1828 standard, which comprise the framework that should support the creation of terminological resources that aim at some degree of operationalization and that focus on surgical procedures.

SOURCE CATEGORY	RELATION	TARGET CATEGORY
surgical deed	hasSubsurgicaldeed	surgical deed
surgical deed	hasObject	human anatomy
surgical deed	hasObject	lesion
surgical deed	hasMeans	human anatomy
surgical deed	hasMeans	interventional equipment
lesion	hasSite	human anatomy
interventional equipment	hasSite	human anatomy
interventional equipment	hasSite	lesion

Table 17 – The ISO 1828’s Categorial Structure

¹⁴² <https://bit.ly/2HS7LU7>. However, no information is available in the SNOMED International documentation section concerning the inverse process, i.e. importing content from OWL into SNOMED CT’s RF2 files.

The following UML¹⁴³ diagram further explores the ISO 1828 CAST by introducing the domain constraints (ISO, 2012, p. 7). Hence, a surgical procedure should contain at least one surgical deed or action. This surgical deed may, if necessary, be further subdivided into surgical subdeeds, yet one must be necessarily chosen as the main surgical deed. Every surgical deed must, in turn, be performed on a certain part of the human anatomy or a given lesion (located in the latter) and, moreover, it must resort to a given interventional equipment. To illustrate this, and resorting to EndoTerm's graphical notation scheme, the example of <Laparoscopic ovarian cystectomy> was chosen, since it is very often performed on patients with severe endometriosis: <Laparoscopic ovarian cystectomy> **is_a** <Surgical procedure> with <Excision> as the main <Surgical deed>. This <Procedure> **has_object** <Cyst> (a subtype of <Lesion>, this <Lesion> **has_site** <Ovary> (a subtype of <Human anatomy>), and **has_means** <Laparoscope>, i.e. it uses an access device which is a subtype of <Interventional equipment>.

¹⁴³ UML, or **Unified Modeling Language**, is a “general-purpose modeling language with a semantic specification, a graphical notation, an interchange format, and a repository query interface. It is designed for use in object-oriented software applications” (ISO/IEC, 2012, p. 2) – cf. also OMG, 2005; Booch et al., 1999.

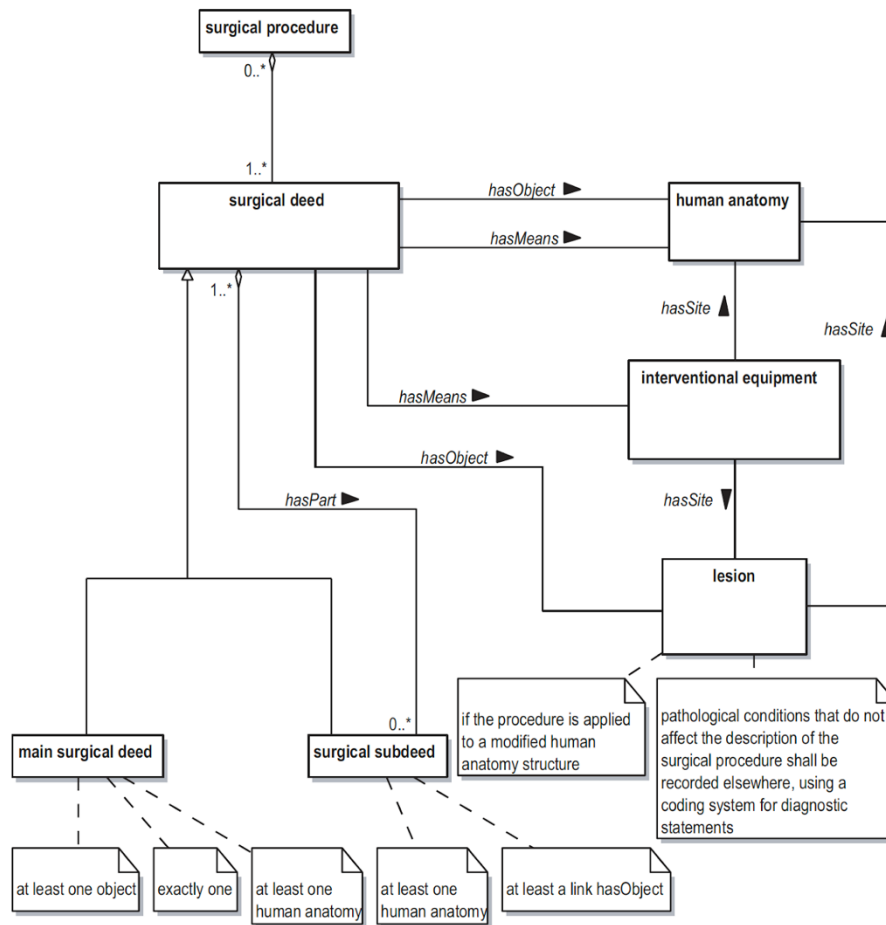


Figure 43 – UML diagram of the ISO 1828 Categorial Structure

In line with what has been stated earlier, the introduction of the ISO 1828 standard in 2012 aimed to bring some consistency and uniformity to the heterogeneous world of procedure classification systems, especially when that need for standardization has become more and more critical in global surgery (Costas-Chavarri & Meara, 2016). The WHO's attempt, with the IHCI, to devise a worldwide standardized resource still seems to fall short, not only due to its low compatibility with the ISO 1828 CAST, but also to its virtually non-existent semantic interoperability up to the present moment, as demonstrated by (Costas-Chavarri & Meara, 2016; Rodrigues et al., 2017; Rodrigues et al., 2011). With its logical, yet flexible, concept model and its increasing prevalence in national healthcare systems (the example of the UK's NHS was provided in Chapter 3), SNOMED CT has been recommended by (Millar, 2016; Rodrigues et al., 2013, 2015, 2017) as the future basis for documenting and reporting procedures and other

healthcare interventions worldwide. For these reasons, both the ISO CAST and SNOMED CT have played an important role in the conceptual foundation of EndoTerm, as will be further explored in the following section.

6.3. Knowing a concept by the company it keeps: EndoTerm's conceptual framework

Following what has been outlined in the preceding chapter, it was believed that EndoTerm should be supported by a core structure focused on the main concepts related to endometriosis and their respective relations, displayed in an explicit way. Moreover, and although <Surgical procedure> plays the central role in this conceptualization, this work aims to show that EndoTerm's framework has been conceived in such a way as to allow further expansion into other categories concerning this pathology, as the foundation that provides consistency to the resource as a whole.

Therefore, and bearing in mind both the previous systematization of procedure TS and the domain knowledge acquired by the terminologist, it was decided that the three axes in question - the WHAT, the WHERE, and the HOW -, perceived as the **essential characteristics**¹⁴⁴ of a <Surgical procedure>, would constitute the basis of EndoTerm's initial, semi-formal¹⁴⁵ conceptualization proposal. It was considered that any <Surgical procedure> regarding <Endometriosis> must entail a main ACTION (e.g. <Excision>, <Resection>, <Drainage>, <Inspection>, etc.), be performed on a given LOCATION, i.e. the anatomical sites that are affected by the disease, and in a certain way - in other words, using a predetermined SURGICAL APPROACH to access the procedure site.

The next critical step was to identify the knowledge organization principles underlying this conceptual framework, and the decision was made to combine a foundational hierarchy resorting to the specific difference, Aristotelian-based approach (cf. Chapter 2), with a categorial structure and respective set of conceptual relations and restrictions. On the one hand, and by relying on the identification of a concept's **delimiting** characteristics, i.e. those which are "used for distinguishing a concept from

¹⁴⁴ According to the ISO standard 1087-1 (2000), those which are "indispensable to understanding a concept" (p. 3).

¹⁴⁵ Semi-formal is used here in the sense that this first proposal was built without the help of automatic reasoning or the ability to undergo operationalization in the KE perspective.

related concepts” (ISO, 2000, p. 3), the hierarchical component would allow a more fine-grained analysis of the various <Surgical procedure> concepts within the scope of Endometriosis and, as a result, a logically consistent differentiation. Furthermore, and as mentioned earlier, it would provide valuable support to the terminologist in most of the tasks enumerated in the introductory section to this chapter, namely the identification of [1] candidate terms and [2] potential terminological synonyms, [3] the interaction with subject field experts, and [4] the drafting of natural language definitions. However, an Aristotelian-based approach, on its own, was considered insufficient to meet the specific needs of this project, since the same factor that contributes to its logical consistency - a somewhat rigid, mono-axial structure - would make a possible integration into resources such as SNOMED CT more challenging.

Accordingly, the inclusion of a categorial structure to complement the existing hierarchy would allow other, non-hierarchical and equally important concept relations to be represented in a more explicit way, thereby contributing to a more comprehensive conceptualization - in the sense that the three <Surgical procedure> core axes could be present -, in line with both the ISO 1828 CAST and SNOMED CT. In order to illustrate the aforementioned framework and its subsequent implementation, a specific <Surgical procedure> was selected due to its relevance in endometriosis, especially in severe cases when the symptoms cannot be alleviated, and the woman has made her decisions as far as fertility is concerned (cf. Chapter 4): the concept of <Hysterectomy>.

Although the number of hysterectomies performed each year has been decreasing in the United States (Desai et al., 2017; Wright et al., 2013), it is still one of the most common major gynecological procedures worldwide. According to the Centers for Disease Control and Prevention (CDC), from 2011 to 2015, 10.6% of women between the ages of 40-44 underwent a hysterectomy in the United States¹⁴⁶. And in Portugal, despite the overall reduction of the number of hysterectomies, including those due to endometriosis (1.1% of the total number of hysterectomies in 2000 to 0.9% in 2014), the procedures specifically related to adenomyosis (cf. Chapter 4), which were already

¹⁴⁶ <https://bit.ly/2r7U78j> (20.02.2018).

higher, experienced a considerable increase (3.4% in 2000 to 5.0% in 2014) (Gante, Medeiros-Borges, & Águas, 2017).

According to the online version of the Dorland's Medical Dictionary, a hysterectomy consists of the excision of the entire uterus and is "performed either through the abdominal wall (...) or the vagina"¹⁴⁷. As with many other medical terms, due to their Greek and Latin origins, the information about WHAT the procedure consists of and WHERE it is performed is already represented in the term's morphological structure in a very specific way: "hysterectomy" was coined from the Ancient Greek forms *hystéra* ('womb') and *-ektomía* ('cutting out from')¹⁴⁸. Thus, and from an etymological perspective, it means the surgical excision of the uterus. <Hysterectomy> can, nonetheless, imply the excision of other anatomical structures in addition to the uterus, as is shown in the following figure (**Figure 44**). A <Subtotal hysterectomy> involves, therefore, the <Excision> of the <Body of uterus>, whereas in a <Total hysterectomy>, the <Cervix> is also excised. In a <Radical hysterectomy>, on the other hand, and besides the <Body of uterus> and <Cervix>, the upper part of the <Vagina> is removed, together with the left and right <Parametrium>, comprising the <Round ligaments>, the <Broad ligaments>, the <Cardinal ligaments>, and the <Uterosacral ligaments> (Mann, 2017).

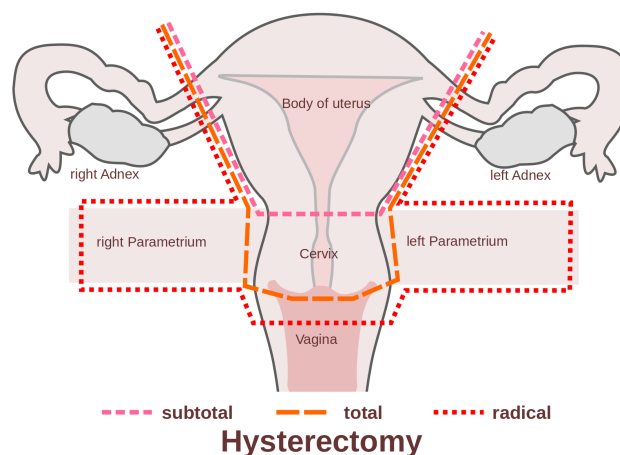


Figure 44 – Types of hysterectomy¹⁴⁹

¹⁴⁷ <https://www.dorlandsonline.com/dorland/definition?pid=100051975> (20.02.2018).

¹⁴⁸ <https://www.etymonline.com/word/hysterectomy> (20.02.2018).

¹⁴⁹ https://commons.wikimedia.org/wiki/File:Scheme_hysterectomy-en.svg (18.10.2017).

As for the third axis, the HOW, no further data is provided in “hysterectomy” as a simple term (i.e. with one lexical unit) concerning the technique used to access the procedure site (approach). Due to that, and also as an alternative to the existing procedure TS, which are organized either by Action or by Location/Target, the decision was made to use the HOW (Surgical approach) as the main axis of EndoTerm’s core hierarchy.

Before devising the conceptualization, the different surgical approaches of the ICHI, the CCI, the ICD-10-PCS and of the CCAM were collected and systematized (cf. **Annex 2**). SNOMED CT was not included in this process, since it relies on a different framework concerning the <Procedure> approaches, as well as on a different nomenclature: the notion of Approach used in this work, i.e. the technique to access the procedure site, corresponds to SNOMED CT’s |Access (attribute)|, whereas its |Surgical approach (attribute)| is closer to the CCAM’s “Voie d’accès”. Moreover, the |Access| attribute comprises only three qualifier values: |Open approach| (presumably matching the open approach of the remaining TS), |Closed approach| (it is not exactly clear what this refers to), and |Percutaneous approach| (again, presumably a match to the previously identified percutaneous approach).

Afterwards, the several values for each approach concerning the types of <Hysterectomy> in the aforementioned procedure TS were also systematized, and the following results were set up for EndoTerm as the basis of the conceptualization representing the main types of <Hysterectomy> according to <Surgical approach>: [1] <Open approach>, [2] <Laparoscopic approach>, [3] <Per orifice approach>, and [4] <Laparoscopic per orifice approach>. The latter, included only in the CCI and the ICHI, was incorporated into the initial conceptualization via polyhierarchy, thereby inheriting, via the generic-specific relation, the corresponding characteristics from both ‘parent concepts’.

The next step was to position the approaches above within the broader concept of <Surgical approach> by resorting to a specific difference, Aristotelian-based framework, as referred to earlier. To support this concept system, two main axes of analysis were set up and the specific differences outlined, as can be seen in the matrix-like structure below (cf. **Table 18**).¹⁵⁰

AXIS OF ANALYSIS	SPECIFIC DIFFERENCE	<Open approach>	<Laparoscopic approach>	<Per orifice approach>	<Laparoscopic per orifice approach>
degree of invasiveness	/invasive/	x			
	/minimally invasive/		x	x	x
existence of skin incision	/with skin incision/		x		x
	/without skin incision/			x	x

Table 18 – Specific-difference matrix supporting EndoTerm’s conceptualization by <Surgical approach>

The following phase consisted of creating a visual, semi-formal concept representation developed via CMap Tools. As can be seen in the **Figure 45** below, the specific differences are distinguishable as far as the colour scheme is concerned (in orange). Furthermore, the map also depicts what Roche calls concept name, which is “constructed in such a way that, simply by reading it, one understands the place the concept occupies inside the notional system” (Roche, 2015, p. 132). Therefore, he claims, it should not be confused with a random concept ID number (i.e. a non-semantic identifier, such as the ones in SNOMED CT).

¹⁵⁰ As referred to earlier regarding the concepts, terms, and relations, the specific differences also follow a typographic convention, being represented, in this case, between forward slashes and as lower case.

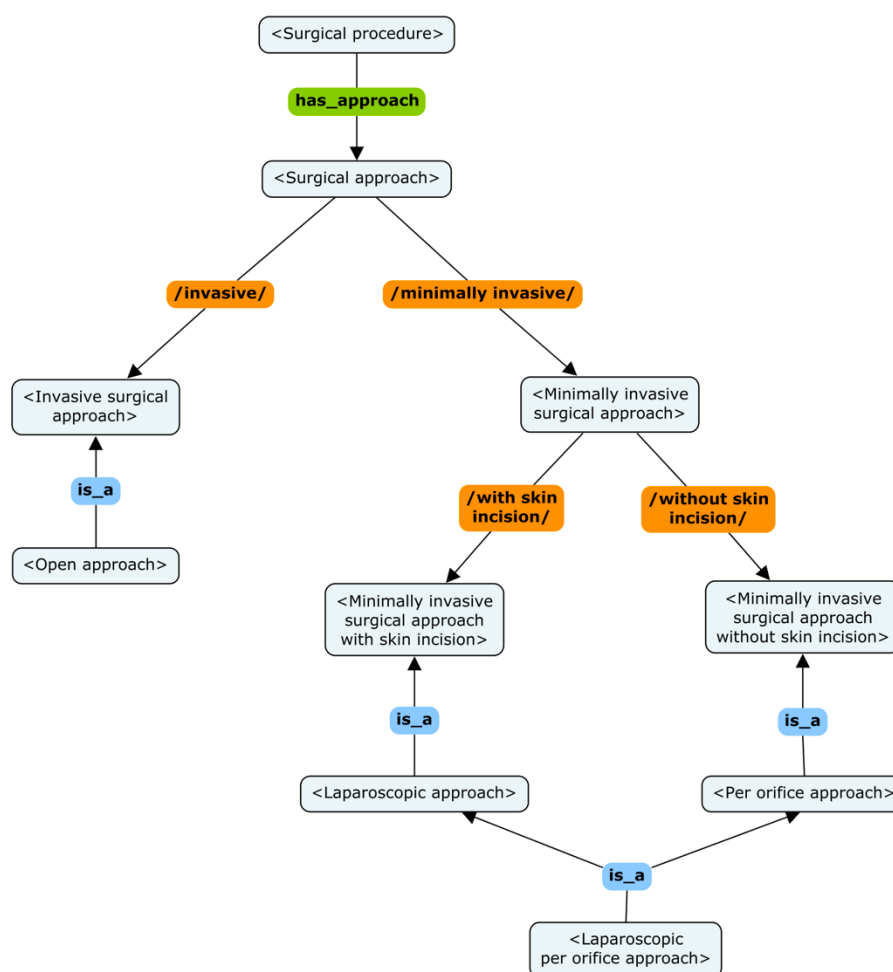


Figure 45 – EndoTerm’s conceptualization of <Surgical procedures> by surgical approach

Then, and in line with the health informatics standards focused on categorial structures (CEN, 2005; ISO, 2007, 2012), EndoTerm’s categorial structure for <Surgical procedures> was defined, as a way to enhance the preceding hierarchy-based system. In addition to the top-level concepts, which will constitute the basis of the remaining hierarchies within EndoTerm, this categorial structure also comprises a set of non-hierarchical relations to supplement the hierarchical **is_a**. The former were adapted from the comprehensive systematization of the relations available at the UMLS (cf. Chapter 3) and SNOMED CT¹⁵¹.

¹⁵¹ The overview of the authorized attribute relationships and respective target concepts as regards SNOMED CT’s |Procedure| concepts can be consulted in **Annex 3**.

Finally, EndoTerm’s categorial structure also includes the domain constraints, in other words, the authorized Source Concept (SC) - Relationship (Rel) - Target Concept (TC) combinations¹⁵².

SOURCE CONCEPT	RELATIONSHIP	TARGET CONCEPT
<Surgical procedure>	is_a	<Procedure>
<Surgical procedure>	has_method	<Surgical action>
<Surgical procedure>	has_direct_procedure_site	<Anatomical entity>
<Surgical procedure>	has_indirect_procedure_site	<Anatomical entity>
<Surgical procedure>	has_direct_morphology	<Morphologic abnormality>
<Surgical procedure>	has_approach	<Surgical approach>
<Surgical procedure>	uses_device	<Device>
<Surgical procedure>	uses_access_device	<Device>
<Surgical procedure>	has_direct_device	<Device>

Table 19 – EndoTerm’s categorial structure for <Surgical procedure> concepts

Moreover, the following table (**Table 20**) aims to demonstrate that this categorial structure is in line with both the ISO 1828: 2012 CAST and SNOMED CT’s concept model, especially concerning <Surgical procedure> concepts¹⁵³. EndoTerm’s contribution lies in the enhancement of the <Surgical approach> hierarchy, which is non-existent in the ISO 1828 CAST and rather incipient in SNOMED CT, with only three values, as shown earlier in this section.

¹⁵² In Description Logic, the source and target concepts are also known as domain and range, respectively, and they are also subject to constraints (Baader et al., 2003; Rector & Rogers, 2006).

¹⁵³ Once again, as following the same approach from Chapter 3, the summary table 20 displays the typographic and structural idiosyncrasies of each one of the cited resources.

ISO 1828: 2012 CAST	SNOMED CT	ENDOTERM
hasSubsurgicaldeed	Method (attribute)	has_method
surgical deed	Surgical action (qualifier value)	<Surgical action>
human anatomy	Body structure (body structure)	<Anatomical entity>
lesion	Morphologically abnormal structure (morphologic abnormality)	<Morphologic abnormality>
interventional equipment	Physical object (physical object)	<Device>
hasObject	Procedure site - Direct (attribute) Direct morphology (attribute) Direct device (attribute)	has_direct_procedure_site has_direct_morphology has_direct_device
hasSite	Procedure site - Indirect (attribute)	has_indirect_procedure_site
hasMeans	Using device (attribute) Using access device (attribute)	uses_device uses_access_device
-----	Surgical access values (qualifier value)	<Surgical approach>
-----	Access (attribute)	has_approach

Table 20 – Overview of ISO 1828 CAST, SNOMED CT's and EndoTerm's categorial structures

In addition to the <Surgical approach> core hierarchy, which may apply to several types of <Surgical procedure> within the scope of <Endometriosis>, two further hierarchies were created based on <Hysterectomy> and categorizing its different types according to the <Surgical approach> (the HOW) and to the <Anatomical entity> that is directly removed in the <Procedure> (the WHERE). These were complemented by EndoTerm's categorial approach, as shown in **Figures 46** and **47**, respectively).

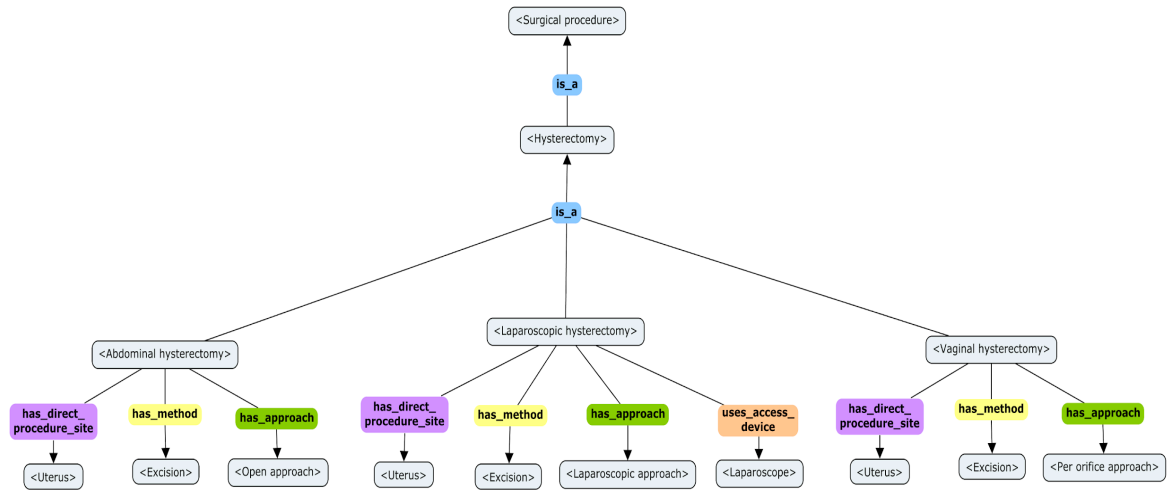


Figure 46 – Types of <Hysterectomy> according to the <Surgical approach>

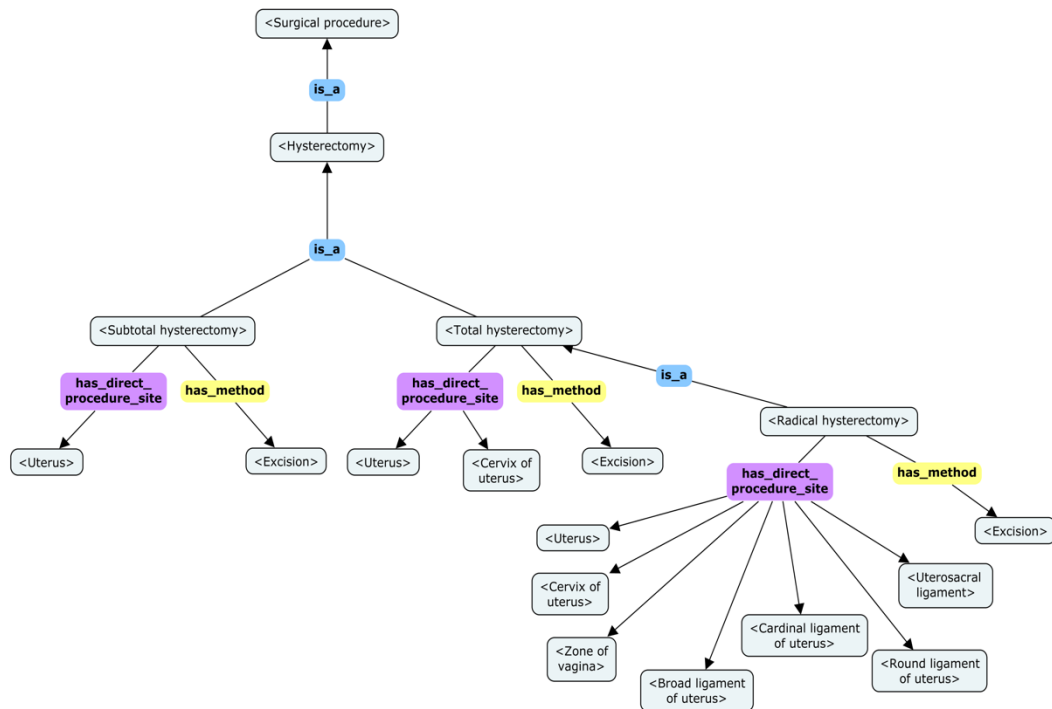


Figure 47 – Types of <Hysterectomy> according to the <Anatomical entity> removed via this procedure

The basis for the <Anatomical entity> hierarchy (**Figure 48**) was the Foundational Model of Anatomy (FMA) (Golbreich, Grosjean, & Darmoni, 2013; Noy et al., 2018; Rosse & Mejino, 2003), not only because this domain ontology is an authoritative source in the field, but also because it is available as OWL and RDF/XML, which can be useful in future work involving EndoTerm and Linked Data.

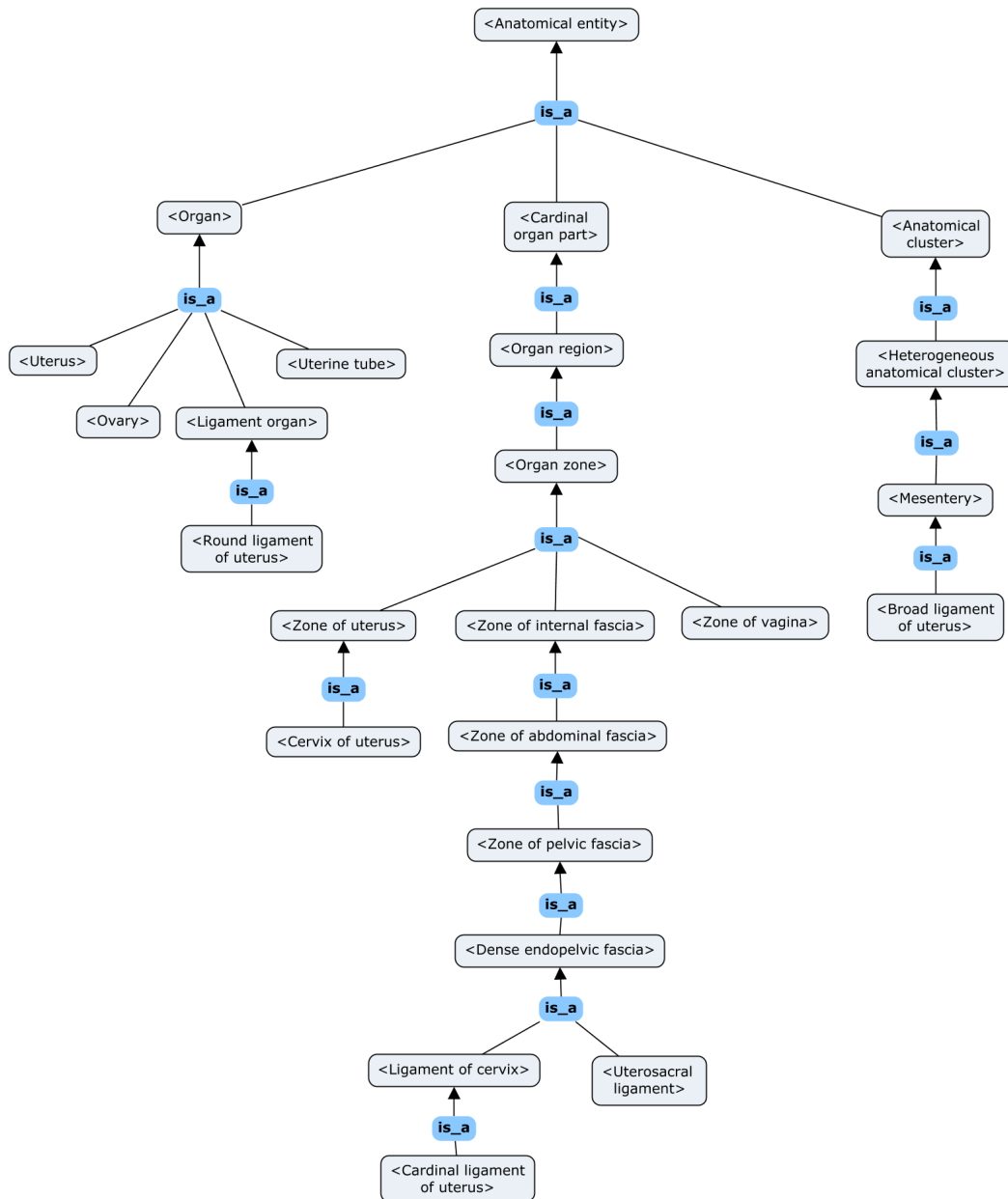


Figure 48 – EndoTerm’s hierarchy for <Anatomical entity>

As regards <Surgical action>, only one subordinate concept was included, given that, in the chosen example of <Hysterectomy>, the action (the WHAT) is always the same (<Excision>). Nevertheless, the core system has been conceived in such a way as to allow for other types of <Surgical action> to be added in the future. Concerning the <Morphologic abnormality> category, a simple hierarchy <Ovarian cyst> **is_a** <Cyst> **is_a** <Lesion> **is_a** <Morphologic abnormality> was devised, only to accommodate the example of another common <Surgical procedure> in endometriosis - <Ovarian cystectomy> - that will be explored in chapter 8, and to demonstrate how new content can be added to EndoTerm. Finally, as far as the category <Device> was concerned, the <Flexible laparoscope> **is_a** <Laparoscope> **is_a** <Endoscope> **is_a** <Device> / <Rigid laparoscope> **is_a** <Laparoscope> **is_a** <Endoscope> **is_a** <Device> hierarchy was included, mainly to demonstrate, in a more explicit way, that the <Laparoscopic approach> resorts to the <Laparoscope> as an access device via the conceptual relationship (<uses_access_device>).

Ultimately, this combination of hierarchical and non-hierarchical structures resulted in EndoTerm's core conceptual system, with a little under 70 concepts in total. The example below (**Figure 49**) is from one of the devised micro-concept systems, namely from the concept of <Total laparoscopic hysterectomy>, which **is_a** <Total hysterectomy> (in what concerns the WHERE) AND **is_a** <Laparoscopic hysterectomy> (regarding the HOW), thereby inheriting the conceptual relations and respective target concepts from both.

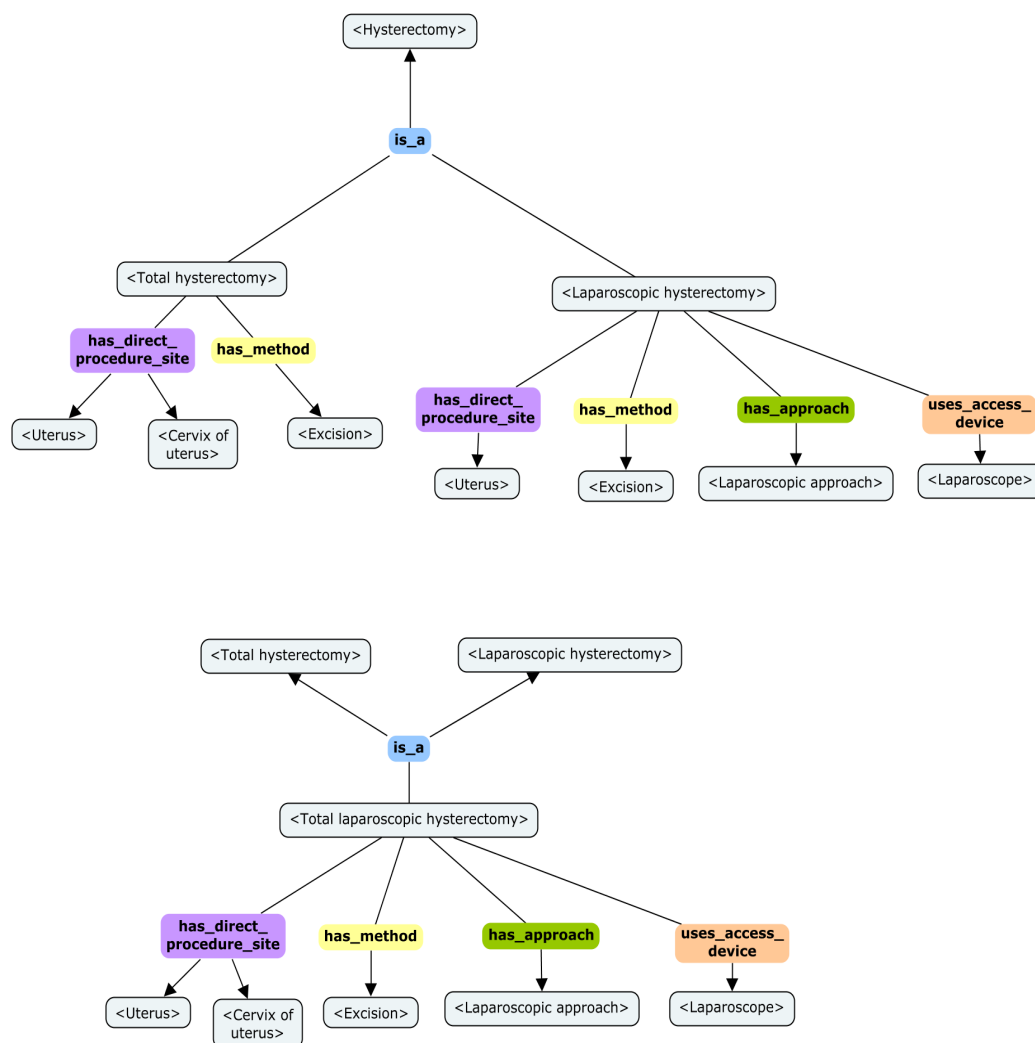


Figure 49 – Example of micro-concept system for <Total laparoscopic hysterectomy>

In short, this chapter has shown that despite the increasing difficulties in stabilizing biomedical concepts due to today's exponential innovation or to the idiosyncrasies of a given subject field, there are still core foundations and essential characteristics that one can rely on to underpin a conceptualization. In what concerns <Surgical procedures> within the scope of <Endometriosis>, for instance, the 3 pillars (What/Where/How) provided a key starting point from where several approaches emerged – specific difference, Aristotelian-based principles, hierarchical and non-hierarchical conceptual relations, together with a customized categorial structure.

The examples aim to illustrate how this combined approach allows EndoTerm to seemingly have the best of both worlds: on the one hand, the enhanced and logically solid concept differentiation features, provided by the specific difference approach, and, on the other hand, the comprehensiveness, flexibility and interoperability potential, supported by the categorial structure. Moreover, and if integrated into a more comprehensive resource, EndoTerm could enable the users to more effectively query the system by approach or by combining two or all three axes. Ultimately, it is believed that this complementarity can bring added value to the end product by helping overcome the limitations of 'plain' hierarchies while paving the way to interoperability.

As for the methodological option of resorting to the micro-concept system representation, it was mainly due to the need of guaranteeing a logical foundation that could prevent - or, at least, minimize - logical errors, especially at an initial, semi-formal stage where automatic reasoning was not yet available.

Finally, such a conceptual framework can also play a key role in developing and implementing consistent content development strategies, as well as in supporting the drafting of natural language definitions, as will be explored in Chapter 8.

Chapter 7 – Text analysis for onomasiological purposes: EndoTerm’s linguistic dimension

7.1. Introduction

Following what was presented earlier, this chapter will focus on the linguistic component of the project and on its contribution to enhance EndoTerm in that regard. However, and in line with the double dimensional approach to Terminology, the conceptual scope has not been ignored in this part of the work. In fact, the current chapter aims to show how the conceptual dimension can support its linguistic counterpart, and vice-versa. With EndoTerm’s conceptual framework stabilized, the current chapter will specifically address questions [4] and [5] outlined in the introduction:

[4] How can that conceptual framework support the *analysis of a specialized corpus*?

[5] On the other hand, how can pre-defined conceptual information be expressed – if it is, indeed, expressed – in natural language in that specialized corpus?

On the one hand, and as referred to in Chapters 2 and 5, a previous conceptual analysis can assist the terminologist in the preparation of term lists, in the identification of synonyms and interlingual equivalents, when applicable, as well as in the drafting and/or improvement of natural language definitions, namely by providing a framework that ensures logical and methodological consistency. On the other hand, and despite the pitfalls that characterize natural language, such as ambiguity, polysemy, or ellipses, even in specialized communication, resorting to corpus analysis allows a valuable insight into a given community of practice and into the way the experts view their domain. Moreover, if a term is perceived as a “verbal designation of a general concept in a specific subject field” (ISO, 2000, p. 6) and a concept is considered a “unit of knowledge” (*ibid.*, p. 2), then terms can be regarded as key access points to the knowledge of a given subject field. Hence, their ‘natural habitat’, i.e. specialized texts, should also be taken into account in terminology work.

In this particular case, the collection, processing, and analysis of the EndoCorpus, created specifically to meet the needs of this project, aimed to:

[1] propose a list of candidate terms (to illustrate the devised methodology, the examples to be shown regarding the corpus analysis have focused mostly on the term “hysterectomy”¹⁵⁴, in line with what was put forward in the preceding chapter);

[2] determine, again using “hysterectomy” as an example, which linguistic means were used to express semantic information regarding the Target of the procedure, i.e. which anatomical structures were removed (Where?), and the Approach, i.e. how the access to the procedure site took place (How?). In short, two of the three axes outlined in the preceding chapter were used as a sort of ‘conceptual compass’, to guide the terminologist in the stages referring to data processing and interpretation.

Furthermore, as will be explored below, previously defined categories within EndoTerm’s categorial structure were also used to support the identification and systematization of terms relating to <Device> and <Morphologic abnormality>, to help enhance the content of the resulting terminological resource.

Ultimately, and from a linguistic perspective¹⁵⁵, the corpus analysis relied on a hybrid approach comprising the identification and study of both internal and external patterns. The former focus on the **morphological** features of terms, hypothesizing a semantic connection between two or more terms based on their common morphemes (e.g. “-ectomy” for ‘excision’) – a semantic potential which can be especially relevant in medical terminology (cf. the work by Grabar & Zweigenbaum, 2004) –; and/or on the **formal** properties of multi-word terms, assuming that those that share the same head (“total hysterectomy” and “subtotal hysterectomy”, for instance) might also share semantic components (L’Homme & Marshman, 2006). As for the latter, they rely on the premise that semantic relations can be expressed through various explicit patterns, mostly lexical, having been successfully adapted into terminology work in terms of knowledge representation and organization by (Condamines & Rebeyrolle, 2001), with their “conceptual relation patterns”, and by the ground-breaking work from (Meyer,

¹⁵⁴ In line with the graphical notation principles outlined in previous chapters, terms will be represented in lower case and between quotation marks.

¹⁵⁵ The specific reference to a ‘linguistic’ approach has been made in order to clearly distinguish it from a distributional one (cf. Harris, 1954), which has not been used within the scope of this work.

2001; Meyer & Mackintosh, 1996), with their “Knowledge Rich Contexts” and their respective “Knowledge Patterns” (KPs).

The option for a ‘hybrid approach’, as mentioned earlier, was due to the fact that all the aforementioned perspectives have their limitations: in what concerns the formal features of terms, it is known that many related terms may not be similar in form or that, conversely, many similar forms may not be semantically related (e.g. polysemy) (L’Homme & Marshman, 2006); as for the lexical patterns, they are often unpredictable, may become ambiguous, especially if the link to anaphoric reference is no longer provided, or may change according to domain and language (Condamines et al., 2013). Hence, it was believed that this complementary approach would best suit the objectives of this part of the project, since, and to quote (L’Homme & Marshman, 2006, p. 77), a good grasp of these techniques by the terminologist can contribute to a better “understanding of a subject field and its terminological and conceptual structures”.

Thus, and as far as the structure of the chapter is concerned, the following sections will address the design and compilation of the EndoCorpus. Afterwards, the semi-automatic processing of the corpus data will be described, as well as the way that data was studied by resorting to a combination of corpus analysis tools and manual analysis. Subsequently, the data extracted from the corpus led to the creation of lexical networks which not only paved the way to EndoTerm’s linguistic module, but also contributed to the elicitation of conceptual information and, in turn, to the enhancement of the existing micro-conceptual systems.

7.2. The EndoCorpus: criteria for corpus design and collection

According to (Bowker & Pearson, 2002; Cabré, 1999; L’Homme, 2004), each corpus serves a particular purpose depending on the research project and the criteria underlying the corpus design and collection should reflect that main purpose. Specialized corpora are no exception. Consequently, and in order to understand whether or not existing corpora could be used within the scope of the EndoTerm project, one of the first tasks consisted of surveying the currently available medical corpora (and

subcorpora), including those focusing only on English¹⁵⁶. However, the results showed a rather modest presence of endometriosis – or, in some cases, none at all. Moreover, most of these resources relied on non-specialized sources, such as magazines, newspapers, or materials destined to popularization¹⁵⁷. As such, they did not fit the purposes of the EndoTerm project and were therefore not used. As a result, the EndoCorpus was created, mainly to address the need for a corpus specifically dedicated to endometriosis.

Firstly, it was necessary to outline the criteria supporting the design and compilation of the EndoCorpus. Hence, and following the principles put forward by (McEnery & Wilson, 2001; McEnery, Xiao, & Tono, 2006; Sinclair, 2005), the EndoCorpus was developed to be [1] machine-readable, [2] authentic, [3] sampled, and [4] representative. The latter, however, does not necessarily refer to statistical results, but to whether or not the text is acknowledged as ‘representative’ by the community by which and for which it has been produced, as stated by (Costa & Silva, 2008, p. 7). In this respect, expert collaboration played, once again, a decisive role, as will be shown below. In addition to the criteria above, it was decided that the corpus should comprise full texts that were freely available, written by subject field experts for subject field experts (expert to expert – cf. (Cabr , 1999; Pearson, 1998), as these would undoubtedly be more suitable for a terminological analysis.

Furthermore, one of the most important things to define at the corpus design stage was whether EndoCorpus would be monolingual or multilingual. While the initial goal was to develop a multilingual corpus which would include English, French, and European Portuguese, it quickly became clear that this objective would be somewhat difficult to attain, due to what (Fiormonte & Priego, 2016) entitle “knowledge monopolies” as regards global academic publishing, a paradigm which puts an increasing pressure on researchers – in Medicine as in other domains – to publish their results in

¹⁵⁶ The medical subcorpora of the British National Corpus (BNC), for instance, contains only 80 hits for “endometriosis”, whereas the Corpus of Contemporary American English (COCA) includes 302 hits, but only 52 from academic texts.

¹⁵⁷ That is the case, for example, of the more recent Medical Web Corpus from Sketch Engine, an English corpus made up of texts collected from the Internet. In fact, most of its 1,332 hits for “endometriosis” were taken from the Wikipedia and from patient-oriented websites, several of which had ceased to exist or had been deactivated at the moment this thesis was being written.

English, rather than in their native language¹⁵⁸. As a result, the EndoCorpus remained monolingual, in order to allow for a reasonable sample of endometriosis-related texts and their respective terms.

After having settled on the issue of language, a text typology (i.e. “la réunion et la classification d’un ensemble de textes sous une même étiquette” - (Costa & Silva, 2008, p. 6) was created, based on MeSH’s Publication Types¹⁵⁹. It included: [1] journal articles; [2] consensus statements; [3] guidelines; and, following expert suggestions, [4] reviews, more specifically the so-called Cochrane Reviews. Moreover, for the EndoCorpus to be ‘representative’ of the constant evolution that characterizes scientific and medical knowledge, it was decided to select more recent texts, published in the last 5 years (2013-2017). Nevertheless, a very small number of consensus statements, guidelines, and Cochrane reviews published before 2013 were also included, due to their relevance for the topic under analysis and to the fact that these types of documents, especially concerning endometriosis, are not issued as regularly as journal articles.

[1] Journal articles – Due to the massive amount of endometriosis publications in the last few years (1,000 articles alone with “endometriosis” as part of the title or the abstract were indexed in PubMed in 2017), it was decided that the journal articles would be selected from the *Journal of Endometriosis and Pelvic Pain Disorders* (JEPPD) – formerly known as *Journal of Endometriosis* –, a renowned international, peer-reviewed journal specifically dedicated to this condition. Given that the examples the EndoTerm project presented thus far focus predominantly on surgical procedures related to endometriosis, an online search was carried out for publicly available articles, published between 2013 and 2017, that had “endometriosis” and “surgery” as part of their title or abstract.

¹⁵⁸ This is becoming particularly prevalent in Portuguese Higher Education Institutions, with a noticeable impact on research funding and career progression (i.e. tenure). Given this clear prevalence of English as the *lingua franca* of science, it is believed that in the short term, and at least in some areas, expert-to-expert texts in European Portuguese, for example, will become increasingly scarce, and that will also have a major impact on terminological projects, especially those which aim at the creation of multilingual resources.

¹⁵⁹ Cf. <https://www.nlm.nih.gov/mesh/pubtypes.html>.

[2] Consensus statements – According to (NLM, 2018), these are “official statements of the findings or recommendations expressing the outcome of a meeting convened to evaluate current thought and research on a subject of interest.” In this respect, all consensus statements concerning endometriosis and issued between 2013 and 2017 were chosen.

[3] Guidelines – When compared with consensus statements, guidelines have a more normative scope, as they aim to present a set of principles, rules or policies regarding a particular topic or sector within a domain. Another relevant aspect is that guidelines are normally issued by either government agencies or organizations that are perceived as representative within a community of practice, such as professional societies, governing boards, or expert panels (cf. NLM, 2018). Thus, all the guidelines concerning endometriosis that have been published by the major professional societies at a global level were included in the EndoCorpus, even the most recent ones.

[4] Cochrane Reviews – The Cochrane Reviews are systematic reviews of primary research in human health care which focus on a given question and review all the corresponding data and results in a standardized, systematic way, providing “accurate and unbiased health information” in order to support evidence-based decision-making in that particular area (Cochrane, 2018). As stated earlier, this type of resource was not part of the initial typology proposal as regards the EndoCorpus. However, and following expert recommendation, it was included in the corpus, since Cochrane Reviews are acknowledged by the community as a benchmark regarding high-quality and up-to-date health care information. Hence, a search was carried out in the Cochrane Database of Systematic Reviews¹⁶⁰ for freely available reviews, published between 2013 and 2017, that included “endometriosis” and “surgery” in their title, abstract, or keywords. As the number of hits was relatively small, it was decided to widen the search to a period of 10 years, under the same criteria.

¹⁶⁰ Cf. <http://www.cochranelibrary.com/>.

Overall, 30 texts were included in the EndoCorpus, with noteworthy samples from all the text types. After being converted into .txt and ‘cleaned’ regarding formatting, the documents were labelled, and the total number of types and tokens was calculated, as summarized below (**Table 21**):

TYPE	FILE IDENTIFICATION	TYPE	FILE IDENTIFICATION
1. Cochrane Reviews	1.1_Aarts_etal_2015	4. Journal of Endometriosis and Pelvic Pain Disorders	4.1_Arendas_etal_2015
	1.2_Ahmad_etal_2015		4.2_Cela_etal_2014
	1.3_Brown_etal_2014		4.3_Farfaras_etal_2014
	1.4_Duffy_etal_2014		4.4_Gingold_etal_2016
	1.5_Lethaby_etal_2012		4.5_Gizzo_etal_2014
	1.6_Liu_etal_2014		4.6_Luisi_etal_2013
	1.7_Medeiros_etal_2009		4.7_Muzzli_etal_2017
	1.8_Orozco_etal_2014		4.8_Ottolina_etal_2017
2. Consensus statements	2.1_Rogers_etal_2009		4.9_Pissetti_etal_2017
	2.2_Johnson_Hummelshoj_2013		4.10_Rampinelli_etal_2017
	2.3_ACOG_2017		4.11_Rato_etal_2016
	2.4_Johnson_etal_2017		4.12_Siebert_etal_2016
3. Guidelines	3.1_ACOG_2010		4.13_Wills_etal_2017
	3.2_ESHRE_2014		
	3.3_NICE_2017	TOTAL WORD TYPES: 7,775	
	3.4_RCOG_2006	TOTAL WORD TOKENS: 152,309	
	3.5_Saridogan_etal_2017		

Table 21 – Overview of the files that constitute the EndoCorpus

7.3. The EndoCorpus: the semi-automatic processing stage

The following step included the semi-automatic processing of the corpus data. Two different tools were used for this purpose: AntConc and TermoStat. Developed by Laurence Anthony, from Waseda University (Japan), AntConc is a freeware corpus analysis toolkit used mostly for concordancing and text analysis. For this part of the project, the most recent version for Macintosh OS X (v. 3.5.7) was used (Anthony, 2018). TermoStat 3.0, in turn, was developed by Patrick Drouin, from the Université de

Montréal (Drouin, 2010), and is an online term extractor that uses hybrid methods to identify candidate terms, i.e. methods that combine statistical and linguistic approaches to term extraction (cf. Pazienza, Pennacchiotti, & Zanzotto, 2005), by comparing results from an analysis corpus (in this case, the EndoCorpus) with those from a non-technical reference corpus (Drouin, 2003).

Firstly, and to get an overview of the EndoCorpus’s content and its terminological potential, AntConc’s N-gram function was explored. As described by (Anthony, 2014), this function allows the user to find “common expressions in a corpus without needing to specify a search term”. In this case, and as illustrated below (**Figure 50**), the entire corpus was scanned for N-Grams between 2 and 4 elements, with a minimum frequency of 2 hits and occurring in at least one of the texts (range).

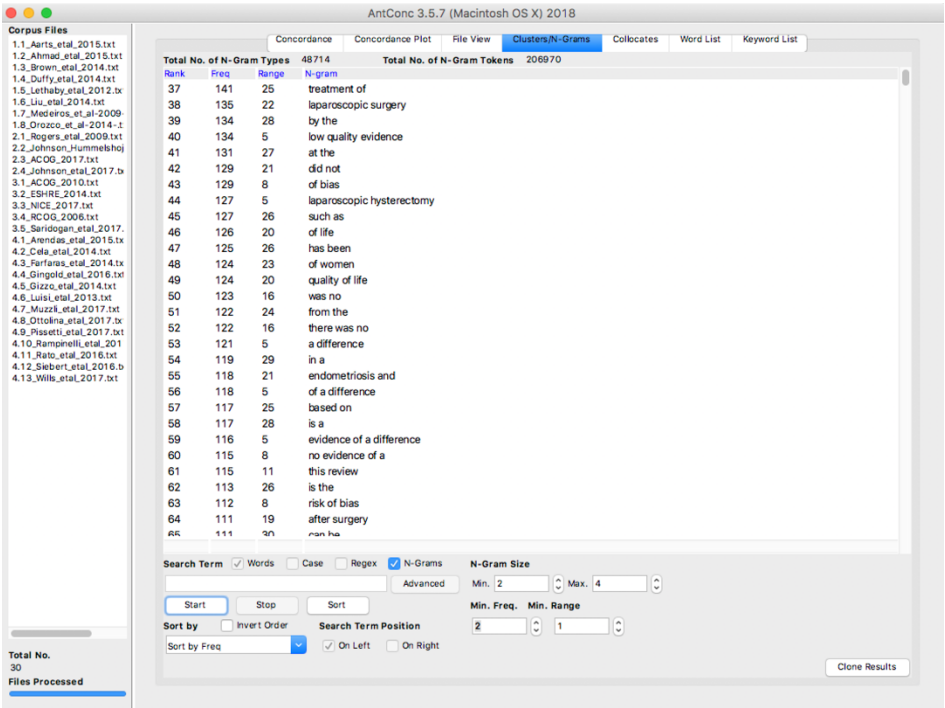


Figure 50 – The EndoCorpus and AntConc’s N-Gram function

The resulting 48,714 N-gram types were then copied into a MS Excel file, to expedite the manual selection process and facilitate consultation via the use of filters. Given that the validation of the candidate terms by subject field experts was not possible in the iteration of the EndoTerm project described in this thesis, a ‘double validation’ strategy was applied, namely by resorting to TermoStat. Therefore, all of the 30 texts

that comprise the EndoCorpus were processed via TermoStat (an example of which is shown in **Figure 51**), and this tool’s list of candidate terms was compared with the preceding AntConc list from the MS Excel file, so that only the forms that appeared on both tools remained on the list. In addition, TermoStat’s analysis allowed the inclusion, in the spreadsheet, of simple candidate terms (one word form)¹⁶¹, as well as of morphosyntactic information (identified in the right hand side as “Matrice”), as a result of Termostat’s integration with the TreeTagger annotator (Schmid, n.d.).

Candidat de regroupement	Fréquence	Score (Spécificité)	Variantes orthographiques	Matrice
hysterectomy	311	335.08	hysterectomy hysterectomies	Nom
laparoscopic hysterectomy	87	177.91	laparoscopic hysterectomy	Adjectif Nom
quality evidence	71	160.51	quality evidence	Nom Nom
low quality evidence	56	142.28	low quality evidence	Adjectif Nom Nom
low quality	57	139.91	low quality	Adjectif Nom
vaginal hysterectomy	53	138.34	vaginal hysterectomy	Adjectif Nom
abdominal hysterectomy	45	127.26	abdominal hysterectomy abdominal hysterectomies	Adjectif Nom
risk of bias	33	108.53	risk of bias	Nom Préposition Nom
operation time	26	95.94	operation time operation times	Nom Nom
bias	60	95.77	bias biases	Nom
trial	152	94.94	trial trials	Nom
outcome	105	93.98	outcome outcomes	Nom
surgeon	59	90.04	surgeon surgeons	Nom
normal activity	25	88.79	normal activities	Adjectif Nom
urinary tract	29	84.26	urinary tract	Adjectif Nom
meta-analysis	19	81.42	meta-analysis	Nom
visceral injury	18	79.13	visceral injury	Adjectif Nom
total laparoscopic hysterectomy	18	79.13	total laparoscopic hysterectomy	Adjectif Adjectif Nom

Figure 51 – Example of how TermoStat was used to process the EndoCorpus

It should be noted that the ongoing systematization of the potential candidate terms was supported by both conceptual and linguistic criteria. As regards the concept-based criteria, it was decided that the word form selection would only include those forms that might correspond to EndoTerm’s main categories stemming from the domain tree and presented in Chapter 6. In the examples that follow, special emphasis will be given to those categories related to the previously described categorial structure, i.e. <Surgical procedure>, <Surgical action>, <Surgical approach>, <Device>, <Anatomical entity>, and <Morphologic abnormality>. Furthermore, linguistic criteria were undertaken in the narrowing down of the MS Excel initial list, thereby complementing the previous approach. These criteria were mostly based on internal patterns, as mentioned earlier: on the one hand, forms which contained morphological features that

¹⁶¹ It was decided not to search for 1-word n-grams via AntConc, as this would contribute to an exponential growth in the number of results and, thus, make the selection process lengthier and even more complex.

could relate to the preceding categories were to be kept on the list (e.g. -ion and -ectomy for nouns referring to a <Surgical action>; -oma for nouns which might address a <Morphologic abnormality>; oopho-, hyster(o), and salping- for examples of an <Anatomical entity>); on the other hand, it was hypothesized that compound nouns sharing the same head and following the Adjective + Noun morphosyntactic structure could also share semantic features¹⁶². All this information was then registered on the spreadsheet, as some form of ‘semantic annotation’, and the end-result was a list with approximately 200 candidate terms, a snippet of which is shown below. The picture also aims to illustrate the usefulness of the filter approach by displaying, on the right-hand side, the selection of forms via their morphosyntactic structure (**Figure 52**):

	A	B	C	D	E
	Candidate Term	Frequency	Morphosyntactic structure	EndoTerm categories + Categorial structure	Term, synonyms?
64	laparoscopic ablation	2	Adj + N	<Surgical action> [<Surgical approach>]	laparoscopic route
65	laparoscopic approach	21	Adj + N	<Surgical approach>	
66	laparoscopic cystectomy	1	Adj + N	<Surgical procedure> [<Surgical approach> + <Morphologic abnormality> + <Surgical action>]	
67	laparoscopic drainage	1	Adj + N	<Surgical action> [<Surgical approach>]	
68	laparoscopic excision	8	Adj + N	<Surgical action> [<Surgical approach>]	
69	laparoscopic hysterectomy	127	Adj + N	<Surgical procedure> [<Surgical approach> + <Anatomical entity> + <Surgical action>]	
70	laparoscopic operation	2	Adj + N	<Surgical approach>	laparoscopic procedure laparoscopic surgery laparoscopic surgical procedure
71	laparoscopic ovarian cystectomy	4	Adj + Adj + N	<Surgical procedure> [<Surgical approach> + <Anatomical entity> + <Morphologic abnormality> + <Surgical action>]	
72	laparoscopic procedure	10	Adj + N	<Surgical approach>	laparoscopic operation laparoscopic surgery laparoscopic surgical procedure
73	laparoscopic radical hysterectomy	1	Adj + Adj + N	<Surgical procedure> [<Surgical approach> + subtype (Target) + <Anatomical entity> + <Surgical action>]	
74	laparoscopic radical surgery	1	Adj + Adj + N	<Surgical procedure> [<Surgical approach> + subtype (Target)]	
75	laparoscopic resection	2	Adj + N	<Surgical action> [<Surgical approach>]	
76	laparoscopic route	6	Adj + N	<Surgical approach>	
77	laparoscopic surgery	58	Adj + N	<Surgical procedure> [<Surgical approach>]	laparoscopic approach laparoscopic operation laparoscopic procedure laparoscopic surgical procedure laparoscopic surgery
78	laparoscopic surgical procedure	2	Adj + Adj + N	<Surgical procedure> [<Surgical approach>]	
79	laparoscopic uterine nerve ablation	4	Adj + Adj + N + N	<Surgical action> [<Surgical approach> + <Anatomical entity>]	
80	laparoscopic uterosacral nerve ablation	1	Adj + Adj + N + N	<Surgical action> [<Surgical approach> + subtype (Target) + <Anatomical entity> + <Surgical action>]	laparoscopic-assisted vaginal hysterectomy laparoscopically assisted vaginal hysterectomy
81	laparoscopic vaginal hysterectomy	1	Adj + Adj + N	<Surgical procedure> [<Surgical approach> + subtype (Target) + <Anatomical entity> + <Surgical action>]	laparoscopic vaginal hysterectomy laparoscopically assisted vaginal hysterectomy
82	laparoscopic-assisted vaginal hysterectomy	10	Adj + Adj + N	<Surgical procedure> [<Surgical approach> + subtype (Target) + <Anatomical entity> + <Surgical action>]	
88	malignant disease	4	Adj + N	<Disease>	
89	malignant gynaecological condition	1	Adj + Adj + N	<Disease>	malignant gynaecological disease
90	malignant gynaecological disease	6	Adj + Adj + N	<Disease>	malignant gynaecological condition
91	mini-laparoscopic hysterectomy	9	Adj + N	<Surgical procedure> [<Surgical approach> + incision size + <Anatomical entity> + <Surgical action>]	
94	minimally invasive procedure	2	Adj + N	<Surgical procedure> [<Surgical approach>]	
95	minimally invasive surgery	6	Adj + N	<Surgical procedure> [<Surgical approach>]	
96	minimally invasive surgical approach	3	Adj + Adj + N	<Surgical approach>	minimally invasive approach minimally invasive surgical technique minimally invasive procedure minimally invasive approach

Figure 52 – Snippet of the list of candidate terms with its annotations

¹⁶² The results stemming from TermoStat as regards the candidate terms’ morphosyntactic structure were then confirmed and two inconsistencies were found: in “total hysterectomy”, the form “total” was correctly tagged as an adjective, whereas “subtotal” in “subtotal hysterectomy” was tagged as a noun; secondly, “single-port” in “single-port laparoscopic hysterectomy” was tagged as an adjective in one of the texts and as a noun in another. Within the scope of the EndoTerm project, “subtotal” and “single-port” in those candidate terms were classified as an adjective and a noun with an adjectival function, respectively.

Moreover, an extra column was added to the right, with the goal of registering potential terminological synonyms, i.e. terms that, within a given language, represent the same concept (cf. Chapter 2; (Costa, 2017; ISO, 2000, 2009).

Due to the considerable amount of candidate terms, and for exemplification purposes only, the upcoming section devoted to the in-depth corpus analysis will focus on “hysterectomy”, either as simple term, or integrating multi-word terms either on the left or on the right, hence on 42 from the almost 200 candidate terms, with a partial view being displayed in the following figure (**Figure 53**):

	A	B	C	D	E
	Candidate Term	Candidate Term		EndoTerm categories + Categorial structure	Term. synonyms?
1					
56	hysterectomy			<Surgical procedure> [<Anatomical entity> + <Surgical action>]	
57	hysterectomy approach			<Surgical approach>	
69	laparoscopic hysterectomy			<Surgical procedure> [<Surgical approach> + <Anatomical entity> + <Surgical action>]	
73	laparoscopic radical hysterectomy			<Surgical procedure> [<Surgical approach> + subtype (Target) + <Anatomical entity> + <Surgical action>]	
81	laparoscopic vaginal hysterectomy			<Surgical procedure> [<Surgical approach> + subtype (Target) + <Anatomical entity> + <Surgical action>]	laparoscopic-assisted vaginal hysterectomy
82	laparoscopic-assisted vaginal hysterectomy			<Surgical procedure> [<Surgical approach> + subtype (Target) + <Anatomical entity> + <Surgical action>]	laparoscopically assisted vaginal hysterectomy
83	laparoscopically assisted vaginal hysterectomy			<Surgical procedure> [<Surgical approach> + subtype (Target) + <Anatomical entity> + <Surgical action>]	laparoscopic vaginal hysterectomy
91	mini-laparoscopic hysterectomy			<Surgical procedure> [<Surgical approach> + subtype (Target) + <Anatomical entity> + <Surgical action>]	laparoscopically assisted vaginal hysterectomy
99	multi-port laparoscopic hysterectomy			<Surgical procedure> [<Surgical approach> + incision size + <Anatomical entity> + <Surgical action>]	laparoscopic-assisted vaginal hysterectomy
105	open abdominal hysterectomy			<Surgical procedure> [number of entry ports + <Surgical approach> + <Anatomical entity> + <Surgical action>]	
122	prior hysterectomy			<Surgical procedure> [<Surgical approach> + <Anatomical entity> + <Surgical action>]	
138	radical hysterectomy			<Surgical procedure> [<Surgical approach> + <Anatomical entity> + <Surgical action>]	
139	robot-assisted hysterectomy			<Surgical procedure> [subtype (Target) + <Anatomical entity> + <Surgical action>]	robotic hysterectomy
141	robot-assisted laparoscopic hysterectomy			<Surgical procedure> [<Device> + <Anatomical entity> + <Surgical action>]	robotic-assisted hysterectomy
147	robot-assisted laparoscopic radical hysterectomy			<Surgical procedure> [<Device> + <Surgical approach> + <Anatomical entity> + <Surgical action>]	robotically assisted hysterectomy
149	robotic hysterectomy	3	Adj + N	<Surgical procedure> [<Device> + <Anatomical entity> + <Surgical action>]	robot-assisted hysterectomy
152	robotic-assisted hysterectomy	13	Adj + N	<Surgical procedure> [<Device> + <Anatomical entity> + <Surgical action>]	robotically assisted hysterectomy
153	robotically assisted hysterectomy	1	Adv + Adj + N	<Surgical procedure> [<Device> + <Anatomical entity> + <Surgical action>]	robot-assisted hysterectomy

Figure 53 – Candidate terms containing “hysterectomy”

After the selection of the candidate terms regarding “hysterectomy”, a quick search of the entire EndoCorpus was made using AntConc’s Concordance Plot. This feature allows the user to know which texts from the corpus contain a given search term, how many hits of that form can be found in each text, and where they can be specifically located in the text. As illustrated below (**Figure 54**), only 14 of the 30 texts specifically contained the search term “hysterectom*”¹⁶³, so those were the ones that were then subjected to AntConc’s Concordance feature.

¹⁶³ The * operator (wildcard) was used in “hysterectom*” so that both singular and plural forms of the noun in question could be identified by the software.

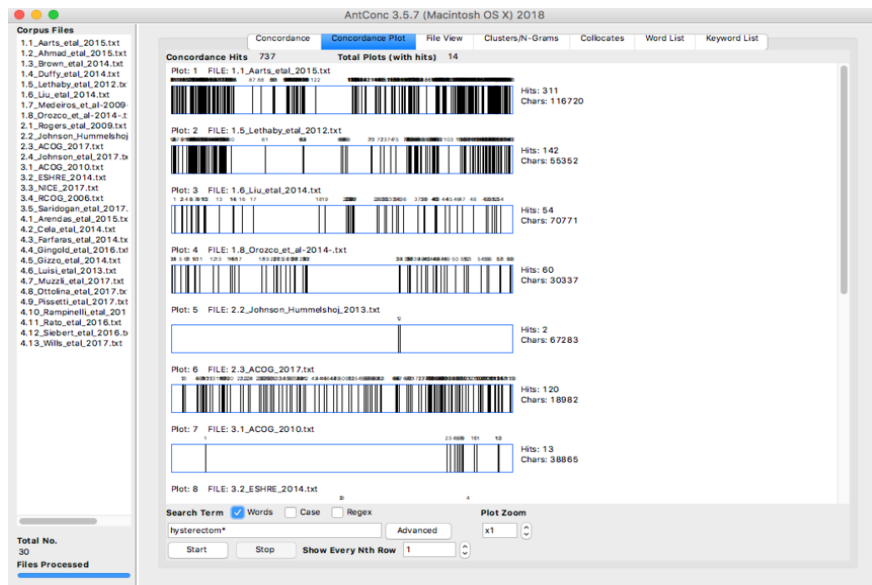


Figure 54 – Concordance Plot for “hysterectomy*”

Next, each one of the 737 concordance hits were displayed via a KWIC (Key Word in Context) format (cf. Figure 55), not only to support the data analysis and discussion – to be explored in the following section – but also to help filter some ‘noise’ that might subsist.

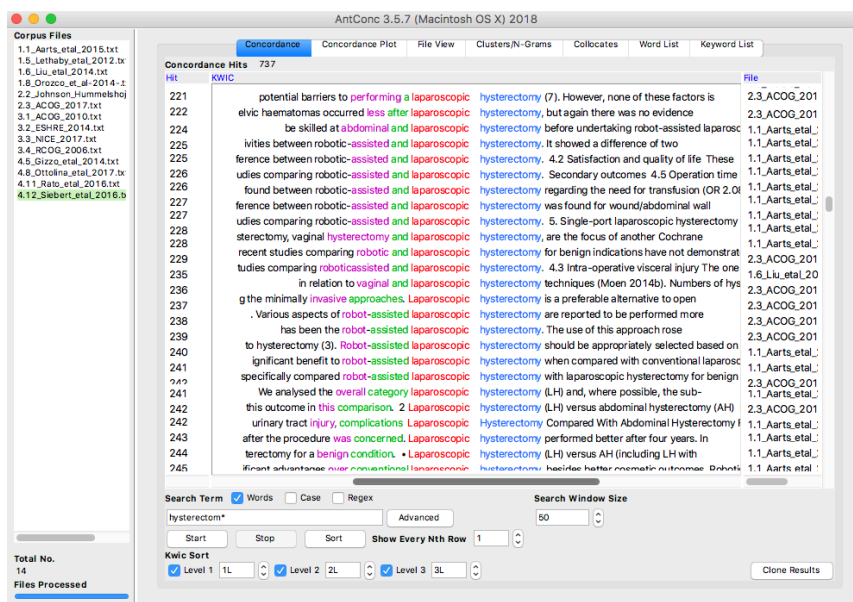


Figure 55 – Concordance results for “hysterectomy*” in KWIC format, sorted by forms on the left of the Search Term

A total of 10 forms were removed from the existing list of candidate terms after the results of the KWIC concordance – thus, the final list of candidate terms with “hysterectom*” was reduced to 32 forms. Some examples of ‘noise’ included:

[a] “invasive hysterectomy”, which does not refer to a type of “hysterectomy” – “Of all minimally invasive hysterectomy procedures, the vaginal approach has been the only one demonstrating a consistent decrease in use from 25% of cases in 1998 to 17% of cases in 2010 [2.3.]”);

[b] the same happened with “alternative hysterectomy” – “Well-designed studies that compare the outcomes of the robot-assisted laparoscopic approach with **alternative hysterectomy routes** will further elucidate...” [2.3.];

[c] finally, the case of “entire abdominal hysterectomy” – “(...) that is not representing the **entire abdominal hysterectomy population** in our clinical practice, but a population of only those women who were eligible for vaginal or laparoscopic hysterectomy as well.” [1.1.]

7.4. The example of “hysterectomy” in the EndoCorpus: data analysis and systematization

As could be seen in the preceding section, nowadays there are plenty of computer programs available in the market that can quickly and effectively process corpora by resorting to a wide range of functionalities. However, the critical part is, in fact, how to analyze those results. Besides the operational know-how and linguistic knowledge, it is indispensable that the terminologist gets to know the chosen subject field extensively before the corpus collection and processing stage. These conditions, when supplemented by a prior conceptual analysis, can provide a valuable support to what can be regarded as the core task of a terminologist when working with corpora, and that is text interpretation.

Following the research questions and the methodological principles put forward in section 7.1., the data extracted from the several iterations of the corpus processing stage was then analyzed and systematized through the development of lexical networks. The examples of such networks, to be further explored in this section, focus primarily

on “hysterectomy” and aim to represent the semantic relation of hypernymy. As previously stated, the corpus data analysis relied on a set of internal and external patterns. Both can play a decisive role in a more fine-grained examination of the results, yet in order to take maximum advantage of all the semantic potential that the internal patterns can have, especially in situations where there aren’t many external patterns available or when these are ambiguous, it is particularly important to have a solid conceptual foundation underlying the adopted methodology. In addition, such an approach can be critical in the terminologist’s interaction with the subject field experts, or in cases where an expert validation is permanently or temporarily unfeasible.

Hence, and in what concerns the **internal** patterns, the example of “hysterectomy” demonstrates that the simple term itself contains morphological features that already provide semantic information – it is the “excision” (-ectomy) of the “uterus” (hyster(o)) –, thereby allowing the terminologist to infer that this <Procedure> possibly involves a <Surgical action> targeting an <Anatomical entity>. As regards the terms’ formal (i.e. morphosyntactic) structure, it was possible to identify, in the EndoCorpus, among others, a pattern consisting of Adj + N forms that could be relatively productive regarding hypernymy (e.g. “abdominal hysterectomy”, “concomitant hysterectomy”, “total hysterectomy”, etc.). However, it is important to be aware not only of possible cases of ‘noise’ (cf. the end of the previous section), but also of the fact that this pattern may refer to different types of knowledge, in the sense that “abdominal hysterectomy” would refer to the surgical approach used by the physician to access the procedure site, “concomitant hysterectomy” to the timing of the surgery (it would occur simultaneously with another procedure), whereas “total hysterectomy” would apply to the body structures to be excised (both the “uterine body” and the “cervix”). Furthermore, and as can be seen in the table below (**Table 22**), several examples of expansion of the Adj + N pattern into Adj + Adj + N or even into Adj + Adj + Adj + N have been found. In the Adj + Adj + N pattern, very often one of the adjectives refers to the surgical approach and the other to the target body structure(s) (e.g. “laparoscopic radical hysterectomy”, respectively), and although the reference to the surgical approach seems to appear more frequently before rather than after the one concerning the body structure(s), examples of the latter have been found as well.

As regards the Adj + Adj + Adj + N pattern, the first adjective of the examples indicates either the use of robotic assistance or the number of entry ports for the instruments used in the procedure.

Candidate Term	Morphosyntactic structure
"hysterectomy"	N
"abdominal hysterectomy" "concomitant hysterectomy" "elective hysterectomy" "laparoscopic hysterectomy" "mini-laparoscopic hysterectomy" "prior hysterectomy" "radical hysterectomy" "robotic-assisted hysterectomy" "subtotal hysterectomy" "supracervical hysterectomy" "total hysterectomy" "vaginal hysterectomy"	Adj + N
"abdominal subtotal hysterectomy" "conventional laparoscopic hysterectomy" "laparoscopic radical hysterectomy" "laparoscopic-assisted vaginal hysterectomy" "open abdominal hysterectomy" "robot-assisted laparoscopic hysterectomy" "standard abdominal hysterectomy" "subtotal abdominal hysterectomy" "subtotal laparoscopic hysterectomy" "total abdominal hysterectomy" "total laparoscopic hysterectomy"	Adj + Adj + N
"single-port laparoscopic hysterectomy"	N + Adj + N
"robot-assisted laparoscopic radical hysterectomy"	Adj + Adj + Adj + N
"single-port total laparoscopic hysterectomy"	N + Adj + Adj + N

Table 22 – Morphosyntactic patterns for “hysterectomy”

The initial stage of the corpus analysis, namely the one regarding the N-Gram results, also allowed for the identification of possible cases of ellipsis. The following forms (**Table 23**), all referring to the undertaken surgical approach to this surgical procedure, seem to point towards a more ‘linguistically economic’ designation (first row). Although quite pertinent, this topic will not be further explored within the scope of this thesis. Still, it can be taken into account in future research work.

abdominal hysterectomy	laparoscopic hysterectomy	minimally invasive hysterectomy	robot-assisted hysterectomy	vaginal hysterectomy
abdominal approach to hysterectomy	laparoscopic approach to hysterectomy	minimally invasive approach to hysterectomy	robot-assisted laparoscopic hysterectomy	vaginal approach to hysterectomy
		minimally invasive route to hysterectomy	robotic hysterectomy	vaginal surgical approach for hysterectomy
			robotic-assisted hysterectomy	
			robotically assisted hysterectomy	

Table 23 – Examples of forms displaying potential cases of ellipsis

Several external patterns were also present in the corpus, mainly via lexical relations, but also through what Meyer (2001) categorizes as grammatical and paralinguistic KPs. The following examples (**Table 24**) focus on **hyponymy** and regard the types of surgical approaches to “hysterectomy”. Different colours have been used to emphasize the components of this semantic relation: hypernyms in orange, hyponyms in blue, and the KPs that signal the relation itself in red. As is shown below, examples 1-8 refer to lexical patterns, some via the verb “to be”, but not all of them (e.g. “refers to”, “subdivided into”, “may be categorized into”, etc.); in examples 9 through 13, the hyponymy relation is expressed by either a Verb + Adverb (9 and 10) or a Verb + Preposition + Noun Phrase (11, 12, and 13) combination, which can also be a relevant pattern to extract semantic information, considering that reference is being made to the way (How?) an action is performed; in examples 15 and 16, on the other hand, the semantic relation is expressed not by a verb or verb combination, but by a preposition (“including”) and a noun “subcategories”, respectively. Finally, examples 17 to 20 depict the so-called paralinguistic KPs (Meyer, 2001), i.e. where the terminologist elicits semantic information from punctuation, the use of bold or italics, supplemented by the use of abbreviations such as “e.g.”. In some cases, the layout of the actual text (like placing the *definiendum* and the *definiens* in different lines, for instance) can also contribute to shed some light into the terminologist’s analysis. If that is the case, AntConc’s File View can be used to visualize the information more accurately (cf. **Figure 56** below).

1	"The four approaches to hysterectomy for benign disease are abdominal hysterectomy (AH), vaginal hysterectomy (VH), laparoscopic hysterectomy (LH) and robotic-assisted hysterectomy (RH)" [1.1.]
2	"For this document, vaginal and laparoscopic procedures (...) are considered minimally invasive surgical approaches." [2.3.]
3	"The vaginal approach is preferred among the minimally invasive approaches." [2.3]
4	"Mini-AH refers to an approach to hysterectomy where the abdominal incision does not exceed 7 cm." [1.1.]
5	"Approaches to hysterectomy may be (...) categorised into four options: abdominal hysterectomy (AH); vaginal hysterectomy (VH); laparoscopic hysterectomy (LH) (...); and robotic-assisted hysterectomy (RH)" [1.1.]
6	"Approaches were as follows: abdominal hysterectomy (...); vaginal hysterectomy (...); laparoscopic hysterectomy (...); robotic hysterectomy (...)" [1.1.]
7	"We defined SP-LH as LH through one single port." [1.1.]
8	"... we further subdivided LH (...) into LAVH, LH(a), TLH and non-categorisable LH" [1.1.]
9	"Hysterectomies are performed vaginally, laparoscopically, (...) or abdominally." [2.3.]
10	"... usually refers to a hysterectomy where at least part of the operation is undertaken laparoscopically..." [1.1.]
11	"Hysterectomies (...) can be performed effectively via a laparoscopic approach." [1.6.]
12	"... would be different when performed by the laparoscopic route..." [1.5.]
13	"... hysterectomy was performed through the abdominal or laparoscopic route..." [1.5.]
14	"VH involves removal of the uterus via the vagina, with no abdominal incision." [1.1.]
15	"(Laparoscopic hysterectomy (LH, including mini-LH and single-port (SP)-LH)" [1.1.]
16	"... report their laparoscopic approach to hysterectomy according to our defined subcategories: laparoscopic-assisted vaginal, LH(a), total laparoscopic and subtotal laparoscopic hysterectomy." [1.1.]
17	"We performed subgroup analyses according to the type of surgical procedure (e.g. hysterectomy, sacrocolpopexy,...) [1.6.]
18	"Minimally invasive approaches to hysterectomy (vaginal or laparoscopic, including robot-assisted laparoscopy) should be performed (...)" [2.3.]
19	"c) hysterectomy approach (abdominal, laparoscopic or vaginal)" [1.5.]
20	"Approach to hysterectomy — abdominal, vaginal, laparovaginal, laparoscopic" [1.5.]

Table 24 – KPs used to express the types of surgical approaches to “hysterectomy”

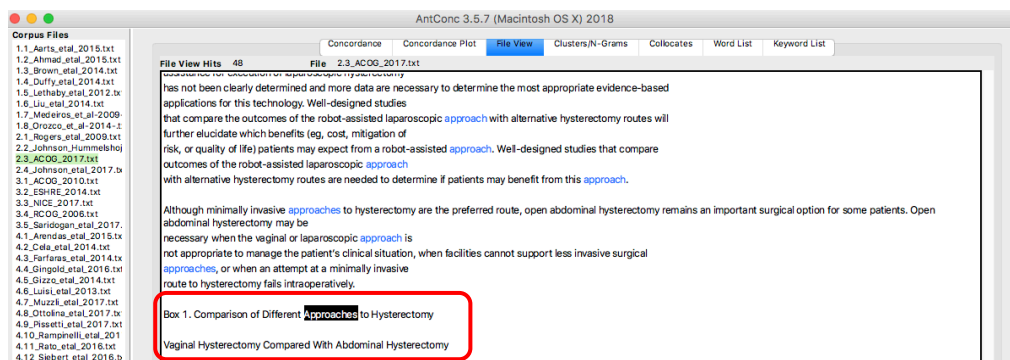


Figure 56 – Example of AntConc’s File View functionality

As far as the types of “hysterectomy” by targeted body structures are concerned, the corpus provided useful insights (cf. **Table 25** below), with explicit reference to the surgical action (except for example 6), as well as to the anatomical structures to be removed. In examples 5 to 7, the parentheses were also used to supplement the linguistic information.

1	"A total hysterectomy involves the removal of both the uterine body and the cervix." [1.5.]
2	"A total hysterectomy is the removal of the entire uterus including the cervix." [1.1.]
3	"... a subtotal hysterectomy involves the removal of only the uterine body, leaving the cervix intact." [1.5.]
4	"When the cervix is not removed this is known as a subtotal or supracervical hysterectomy." [1.1.]
5	"Hysterectomy using an abdominal approach removes either the uterus alone (subtotal hysterectomy) or both the uterus and the cervix (total hysterectomy)." [1.5.]
6	"... a subtotal operation (where the cervix was conserved), was performed..." [1.1.]
7	"...surgery for early cervical cancer involves a radical hysterectomy (removal of the uterus, cervix, upper vagina, parametria and pelvic lymph nodes)." [1.6.]

Table 25 – KPs used to express the types of “hysterectomy” by targeted body structure

Finally, it should be noted that some information extracted from the corpus also displays lexical (examples 1-4) and paralinguistic (examples 5-7) patterns which could help elicit semantic information on potential synonyms (from a discursive perspective, but not necessarily from a terminological one) (**Table 26**).

1	"Robot-assisted surgery (RAS), also known as robotic surgery, robot-assisted laparoscopic surgery or computer-assisted surgery, is a recent innovation in the field of minimally invasive surgery." [1.6.]
2	"Subtotal hysterectomy is also referred to as supracervical hysterectomy." [1.5.]
3	"When the cervix is not removed this is known as a subtotal or supracervical hysterectomy." [1.1.]
4	"Laparo-endoscopic single site (LESS) surgery (...) is another common term which is often used interchangeably with SILS..." [1.2.]
5	Vaginal hysterectomy (VH) [1.1.]
6	"The SILS/LESS approach (...) involves a single intraumbilical 12 mm incision..." [1.2.]
7	"Laparoscopic surgery is increasingly preferred to laparotomy (open abdominal surgery) for several gynaecological procedures." [1.6.]

Table 26 – KPs used to express synonymy

In order to systematize the semantic information stemming from the corpus analysis and then assess how that information could contribute to the enhancement of the existing micro-concept systems as well as of EndoTerm as a resource, a set of lexical networks were developed, with two examples – one depicting the types of “hysterectomy” by targeted body structures (Figure 57) and the other by surgical approach (Figure 58) – being shown below.

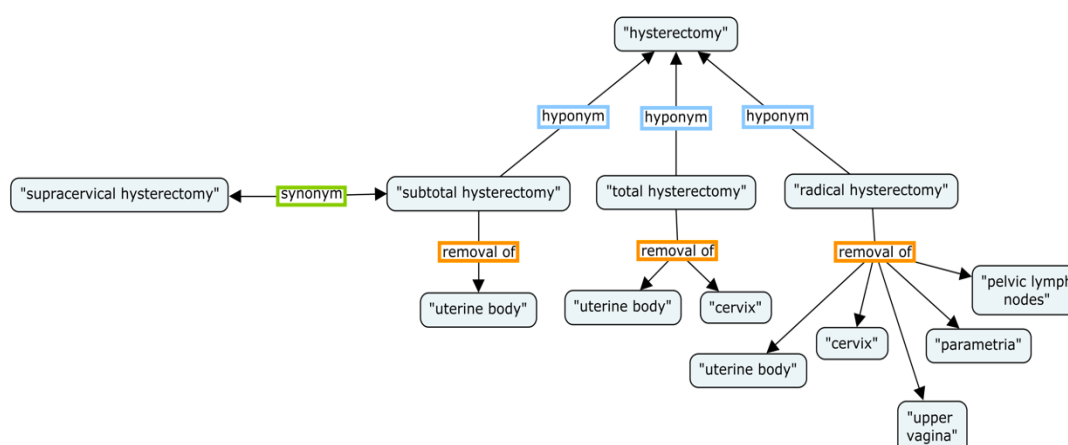


Figure 57 – Lexical network 1: types of “hysterectomy” by targeted body structure

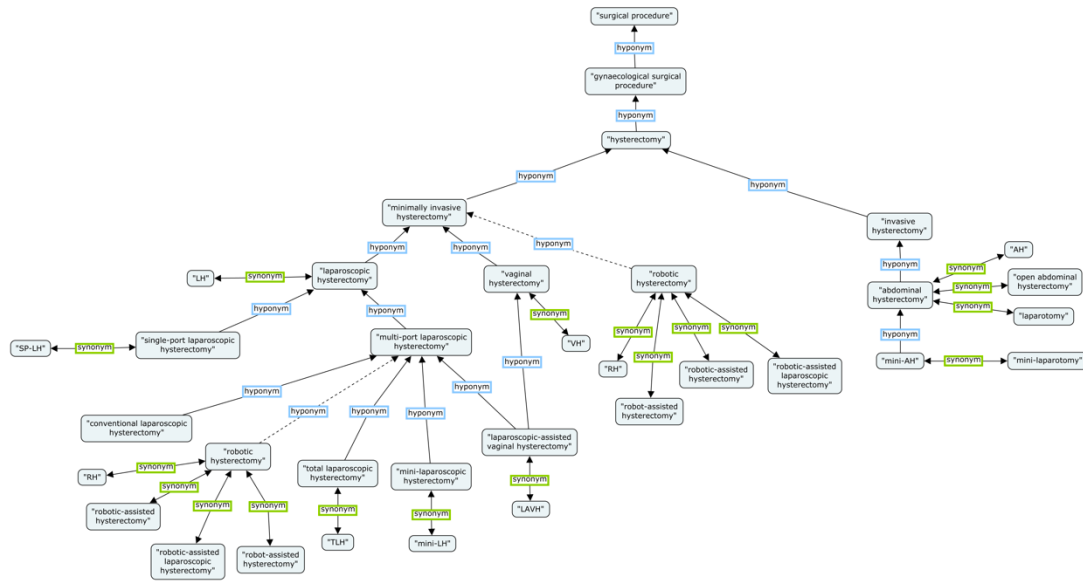


Figure 58 – Lexical network 2: types of “hysterectomy” by surgical approach

About this second lexical network, reference should also be made to the issue of “robotic hysterectomy”, since the corpus data is not clear on whether or not this is an approach to “hysterectomy” on its own or a subtype of “laparoscopic hysterectomy”. If one extrapolates the semantic information concerning the synonymy example (number 1) from **Table 26** earlier, it could be assumed that “robotic surgery” and “robotic-assisted laparoscopic surgery”, by being discursive synonyms, would refer to a subtype of “laparoscopic approach”, but then the corpus presents diverging information in this respect, as is shown below (cf. **Table 27**). Examples 1 to 8 depict “robotic hysterectomy” as a subtype of “laparoscopic approach”, whereas the examples 9 to 11 consider it to be a separate branch. Although it could be argued that the latter perspective might not be representative, as all the examples were taken from the same text, as opposed to the former, it is still not unequivocal which one of the approaches is regarded as prevalent within the community of practice. Widening the EndoCorpus sample could be something to undertake in the future in this respect. Due to all of this, “robotic hysterectomy” was represented under both perspectives in the preceding lexical network, with the dashed lines indicating that level of uncertainty embodied in the corpus information.

1	"Hysterectomies are performed vaginally, laparoscopically (total laparoscopic hysterectomy [with or without robotic assistance] or laparoscopically assisted vaginal hysterectomy), or abdominally." [2.3.]
2	"Approach to hysterectomy - abdominal, vaginal, laparovaginal, laparoscopic" [1.5.]
3	"c) hysterectomy approach (abdominal, laparoscopic or vaginal)" [1.5.]
4	"Hysterectomy is the surgical removal of the uterus and there are three main surgical approaches: abdominal, vaginal and laparoscopic." [1.8.]
5	"The role of robotic assistance for execution of laparoscopic hysterectomy has not been clearly determined..." [2.3.]
6	"Minimally invasive approaches to hysterectomy (vaginal or laparoscopic, including robot-assisted laparoscopy) should be performed (...)" [2.3.]
7	"... compared outcomes of the robot-assisted laparoscopic approach with alternative hysterectomy routes..." [2.3.]
8	"Robot-assisted surgery (RAS) is a technological advance in conventional laparoscopic surgery (CLS)..." [1.6.]
9	"The four approaches to hysterectomy for benign disease are abdominal hysterectomy (AH), vaginal hysterectomy (VH), laparoscopic hysterectomy (LH) and robotic-assisted hysterectomy (RH)." [1.1.]
10	"Approaches were as follows: abdominal hysterectomy (...); vaginal hysterectomy (...); laparoscopic hysterectomy (...); robotic hysterectomy (...)" [1.1.]
11	"Approaches to hysterectomy may be broadly categorised into four options: abdominal hysterectomy (AH); vaginal hysterectomy (VH); laparoscopic hysterectomy (LH) where at least some of the operation is conducted laparoscopically (...) and robotic-assisted hysterectomy (RH)." [1.1.]

Table 27 – Examples from the corpus regarding “robotic hysterectomy” and other related terms

7.5. Saying vs modelling? The added value of a complementary approach

When comparing these lexical networks with the micro-concept systems developed in the preceding chapter, it becomes clear that the structures are not isomorphic (an expanded version of each of the representations is available in **Annexes 4 and 5**). As regards the maps (lexical and conceptual) referring to the affected body structures, the differences exist, though they are not substantial. Indeed, they are related to two main aspects: firstly, that it does not become clear, from the corpus analysis, that “radical hysterectomy” is a subtype of “total hysterectomy”, as depicted in the conceptual system; and secondly, that the anatomical entities seem to be more detailed in the conceptual system, particularly in what concerns the different <Uterus

ligaments>. Conversely, the term “pelvic lymph nodes” does not specifically appear in any of the studied procedure terminological systems that provided the basis for the conceptual system.

In the maps concerning the surgical approaches, on the other hand, the differences are more substantial, with a wider range of options being put forward in the lexical network: in addition to the aforementioned “robotic approach”, other approaches are based on the number of access ports (“single-port” vs. “multi-port”/ “multiple port”), as well as on the size of the incisions (“mini-AH” or “mini-laparoscopic hysterectomy”).

Indeed, and as stated by (Roche, 2007), “saying is not modelling”, in the sense that talking about knowledge is not the same as knowledge itself. However, acknowledging these differences and the idiosyncrasies that characterize each of these dimensions does not mean one of them should be disregarded in detriment of the other. As stated throughout this work, it is believed that a terminological approach to knowledge representation as the one being presented here can greatly benefit from the presence of – and the interaction between – both dimensions, thereby leading to mutual enhancement and to an improvement in the quality of terminology work.

If, on the one hand, a conceptual framework offers the necessary stability and consistency to support the collection, processing, and analysis of a specialized corpus, enabling a more effective development of multilingual resources, the thorough linguistic analysis of that corpus, on the other hand, when anchored in solid methodological principles, also provides privileged access into a given subject field and its respective community of practice. To a certain extent, talking about knowledge and sharing it with other fellow experts or with future experts leads to a co-creation of that knowledge, since the recipients of a specialized text will, in turn, build up on the transmitted knowledge to further develop it, thus contributing to ‘new’ knowledge.

This becomes particularly relevant in fields where technological innovation is constant and critically important, such as Medicine and, within it, everything related to surgical procedures. In this respect, the examples from the current and preceding chapters show that in situations where frequent access to experts is difficult or even impossible, specialized corpora can provide the terminologist with access to ground-

breaking – yet relatively stabilized – knowledge at a faster pace, hence complementing information from concept-based platforms and resources, which may take longer to integrate such content. Furthermore, the task of modelling directly from expert feedback also entails some degree of subjectivity, as the outcome may differ depending on the group of experts and on their scientific and academic backgrounds. In short, a compromise needs to be reached among the members of that group, with the resulting product – e.g. an ontology – necessarily depicting an “intersubjectivity” (Roche, 2015, p. 148).

Taking this lack of isomorphism between the lexical and the conceptual representations into account, and in order to systematize what has been explored thus far, it was decided that the following – and final – chapter would present a case study around the term “single-port laparoscopic hysterectomy”, to illustrate how the linguistic and the conceptual dimensions can interrelate and integrate. Ultimately, chapter 8 will also show how EndoTerm content can be expanded and shared, both with human and machine users.

Chapter 8 – LESS can indeed be more: linguistic and conceptual challenges in the age of interoperability

“For any nascent field to develop and grow, defined terminology is a primary prerequisite for allowing a comprehensive flow of information among specialists and specialties.” (Gill et al., 2010, p. 763)

8.1. Introduction

Following what has been presented in the previous chapter, the development of the EndoTerm project led to the identification of the term “single-port laparoscopic hysterectomy”. Moreover, further searches in the EndoCorpus encompassed the terms “laparo-endoscopic single-site surgery” and “single incision laparoscopic surgery”. This, in turn, prompted the study of laparoendoscopic single-site (LESS) surgery, a type of minimally invasive surgical approach that “allows laparoscopic surgery through a single incision in the umbilicus” (Chern et al., 2012) and that has been gaining ground in various gynecological procedures, including those related to endometriosis. Thus, it was decided to further develop this notion of LESS and select it as the basis of a case study, since it comprises a set of linguistic and conceptual challenges that are relevant to analyze within the scope of the EndoTerm project.

Through the case study, this final chapter aims to provide some answers to the last set of research questions put forward in the introduction to this thesis, namely:

[6] How can the aforementioned conceptual framework support the drafting of natural language definitions?

[7] How do the conceptual and the linguistic dimensions interrelate and how can that be systematized and shared among both human and machine users?

[8] How can a conceptualization be operationalized in order to achieve the aspired interoperability in this field?

Hence, this chapter will start by providing a more comprehensive description of LESS, including its respective usage in the gynecology community. Next, the integration of the conceptual information regarding LESS into the existing EndoTerm content will be addressed, followed by the overall validation of the whole core conceptual structure via

TEDi. Afterwards, a proposal of a formal concept definition will be put forward, as well as of a natural language definition, the latter supported by a definitional template. Then, the issue of how to share EndoTerm content will be presented, in formats targeted at human and machine users. Finally, an example will be shown of how content concerning LESS may be integrated into SNOMED CT, thereby contributing to its enrichment.

8.2. LESS: a Brave New World for surgery?

The latest innovations in what concerns surgical techniques and instrumentation have considerably impacted surgical procedures in recent decades. This has led to the development of a set of surgical approaches known as minimally invasive, with a spectrum ranging from the standard laparoscopic approach to an approach where no surgical incision is actually made, as the surgery is performed solely via natural orifices¹⁶⁴. In order to optimize the benefits of the more traditional laparoscopic approach, surgeons have attempted to decrease the number of trocars used in the procedure, as well as the size of the ports and even the number of incisions. As a result, the notion of LESS was introduced in gynecology about four decades ago, when Wheeless and Thompson, in 1969, first reported the technique and respective results of a series of laparoscopic tubal ligation with single-trocar laparoscopy. In 1991, Pelosi and Pelosi performed a total hysterectomy with bilateral salpingo-oophorectomy using the so-called single puncture technique (Liliana et al., 2011).

An extended search within specialized literature on this topic, including multimedia content, pointed towards a lack of terminological consensus among the expert community. In fact, more than 30 different designations have been identified in the literature, as shown in **Table 28**.

¹⁶⁴ This is currently known as Natural Orifice Transluminal Endoscopic Surgery or NOTES (cf. Box *et al.*, 2008).

Abbreviated form	Full form
E-NOTES	embryonic natural orifice transumbilical endoscopic surgery
LESS	laparoendoscopic single-site surgery
NOTUS	natural orifice transumbilical surgery
OPLS	one-port laparoscopic surgery
OPUS	one-port umbilical surgery
R-LESS	robotic laparoendoscopic single-site surgery
R-NOTES	robotic-assisted natural orifice transumbilical endoscopic surgery
RSS	robotic single-site surgery
S3	single-site surgery
SAES	single-access endoscopic surgery
SALS	single-access laparoscopic surgery
SAS	single access surgery
SAVES	single-access video endoscopic surgery
SIL	single-incision laparoscopy
SILS	single-incision laparoscopic surgery
SIMPL	single-instrument port laparoscopic surgery
SIMPLE	single-incision multi-port laparoendoscopic surgery
SIRS	single-incision robotic surgery
SIS	single-incision surgery
SLaPP	single laparoscopic port procedure
SLiPP	single laparoscopic incision and port procedure
SLIT	single laparoscopic incision transabdominal surgery
SPA	single-port access
SPAL	single-port access laparoscopy
SPICES	single-port incisionless conventional equipment-utilizing surgery
SPL	single-port laparoscopy
SPLS	single-port laparoscopic surgery
SPRS	single-port robotic surgery
SPS	single-port surgery
SSA	single-site-access laparoscopic surgery
SSL	single-site laparoscopy
SSULS	single-site umbilical laparoscopic surgery
STLS	single-trocar laparoscopic surgery

TUES	transumbilical endoscopic surgery
TULA	transumbilical laparoscopic assisted surgery
U-NOTES	umbilical natural orifice transluminal endoscopic surgery
Box et al. [2008]; Romanelli & Earle [2009]; Autorino et al. [2011]; Escobar & Ramirez [2011]; Mereu et al. [2011]; Rao et al. [2011]; Sarkissian & Irwin [2013]; Naitoh [2014]; Ramesh et al. [2014]; Samarasekera et al. [2015]	

Table 28 – Terms referring to LESS surgery

In fact, the dilemmas concerning the nomenclature of LESS surgery are pertinent: on the one hand, the idea of accessing a given operating site via a single incision may lead to misunderstandings, since most open surgeries (e.g. laparotomies) are also performed through a single incision, the difference being in its size. On the other hand, the notion of “single port” may apply to several situations where the different instruments are passed through multiple channels into a single port, but it is no longer valid in cases where multiple ports are inserted via a single incision (Rao et al., 2009; Rao, Rao, & Bhagwat, 2011).

To eliminate or at least minimize this terminological dispersion, the multidisciplinary Laparoendoscopic Single-Site Surgery Consortium for Assessment and Research (LESSCAR) issued a White Paper (Gill et al., 2010) that aimed to standardize the terminology in the field. Interestingly, the expert consortium decided to use the concept and its respective characteristics as the starting point to then be able to reach a consensus regarding the term to be chosen:

To select the most appropriate name/designation for this field, LESSCAR thought the selected name should accurately encompass the following broad concepts: (1) a single entry port (or incision), (2) applicability to multiple locations (abdomen, pelvis, thorax), (3) laparoscopic, endoscopic, or robotic surgery, (4) umbilical or extraumbilical access, (5) intra- and transluminal (...) approach, and (6) and a broad reach so as to be inclusive, not exclusive. (Gill et al., 2010, p. 763)

As a result, LESSCAR proposed the term “laparoendoscopic single-site surgery” as the one that most accurately designated and encompassed all the laparoscopic techniques that use a **single incision** to access the operating site. Therefore, within the

scope of the EndoTerm project, the gathered data concerning LESS provided sufficient ‘food for thought’ to constitute a ‘terminological case study’. Some questions interconnecting the linguistic and conceptual dimensions arose from the data analysis: i) are all the gathered terms actual synonyms from a terminological standpoint, i.e. representing the same concept?; ii) what usage has the expert community in the field of Gynecology (and endometriosis, in particular) been making of these various designations?; iii) knowing that texts do not contain concepts themselves, but the linguistic usages of the terms that designate those concepts, what type of information could be extracted regarding LESS, both from the EndoCorpus and from the consulted literature listed in Table 28?; iv) and what information concerning LESS can be found in the current biomedical TS?

As regards the first point, it is a fact that in the literature, the designations used to refer to LESS are often depicted as synonyms, which, from a terminological perspective, raises the question of whether apples are, indeed, being compared to apples. An analysis of the structure of the terms outlined in Table 28 shows that the notion of a single incision/access to the body (“single incision”; “single access”; “single port”; “single site”) seems to play a central role in this type of surgical approach. One of the terms, however, refers to “incisionless” and four others mention “natural orifice”, which points towards the inexistence of a skin incision and, hence, opposes the notion prevailing in the remaining designations. Additional information is provided by most terms concerning the location of the incision (“umbilical” or the more specific “transumbilical”), the type of surgical approach (“endoscopic” or the more specific “laparoscopic”), the use of some type of equipment (“video”, “trocar”, “equipment-utilizing”), and the presence – or not – of robotic assistance. All of the above seems to indicate that not all these designations are in fact representing the same concept, but expert validation, which was not conducted within the timeframe of this thesis, would be necessary in order to confirm this hypothesis and further develop it.

Secondly, to get a glimpse of the actual usage of these terms among the community of subject field experts, a search was conducted in MEDLINE/PubMed® with the full forms of all the 36 designations and resorting to the following search expression: (“Term” [All Fields] AND gynecology [All Fields] AND (“2010/01/01” [Date-Publication]:

“2017-12-20” [Date-Publication]). The aim was to see which terms had been more widely used in scientific, peer-reviewed papers within Gynecology since 2010 (the date of publication of the LESSCAR White Paper). The results showed that “laparoendoscopic single-site surgery” was the most commonly used designation in the aforementioned time period (with 113 results), followed by “single-port access” (55), “single-port laparoscopy” (45), “single-port surgery” and “single-incision laparoscopic surgery” (with 33 each), and “single-port laparoscopic surgery” (31), just to name those examples with more than 30 hits. When changing the query from ‘gynecology’ into ‘endometriosis’, there were far less hits, but the results were relatively similar, with a prevalence of “laparoendoscopic single-site surgery” (10) over “single-incision laparoscopic surgery” (6), “single-port access” (5), “single-port laparoscopic surgery” (4), in addition to “single-port laparoscopy”, “single-port surgery” and “single-site laparoscopy” (with 3 each).

Thirdly, after a manual analysis of the example of LESS (and related terms) occurring both in the EndoCorpus and in the literature listed in Table 28, it was possible to compile the data into the following lexical network (**Figure 59**):

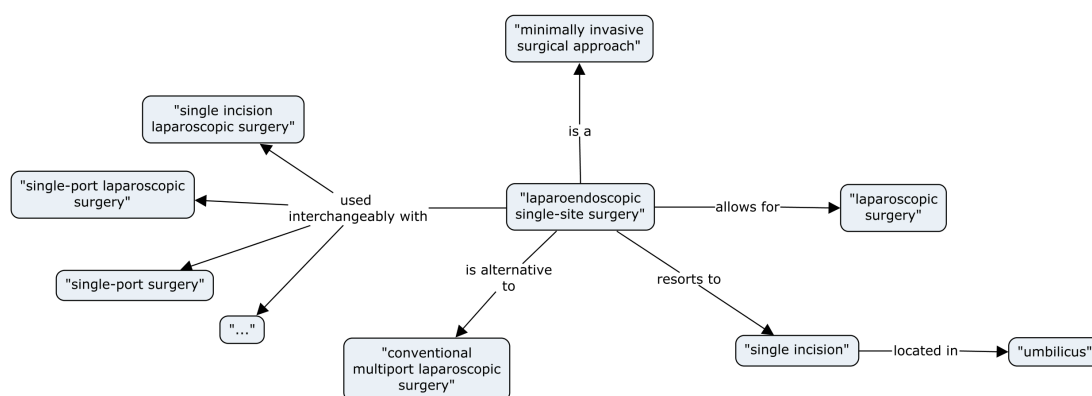


Figure 59 – Lexical network for “laparoendoscopic single-site surgery”

It should be noted that all the presented relations (even the “is a”) were actually present in the texts and thus express lexical relations. The lexical *is a* should therefore not be confused with the conceptual *is_a* relation. Finally, a search was carried out in the disease-oriented and procedure-oriented TS used in chapters 3 and 6, in order to obtain further information about LESS. Being a type of surgical approach that was

estimated, back in 2010, to already account for 50-80% of current surgeries in some specialties (particularly urology and gynecology) (cf. Gill et al., 2010), it was surprising to realize that the concept of <Laparoendoscopic single-site ... [type of surgical procedure]> does not currently exist as such in any of the consulted resources. Although data on <Laparoscopy>, <Laparoscopic approach>, <Percutaneous endoscopic approach> or <Minimally invasive surgical approach> can be found, there is nothing that refers specifically to the existence of a single skin incision, which would allow concept differentiation. The only available references to anything of that sort are in SNOMED CT, with |Single-port laparoscopic cholecystectomy (procedure)| and |Single port laparoscopic nephrectomy (procedure)|¹⁶⁵. In both cases, there is no difference between these SNOMED CT concept definitions and those of |Laparoscopic cholecystectomy (procedure)| and |Laparoscopic nephrectomy (procedure)|. Moreover, they are subtypes of |Laparoscopic procedure (procedure)|.

Due to the terminological pertinence of these results, it was decided to incorporate the gathered information – both linguistic and conceptual – into EndoTerm and apply that information to an example involving <Hysterectomy>. In short, the texts provided relevant semantic information that, together with the conceptual data from the biomedical TS, allowed the terminologist to systematize the acquired knowledge and propose an enhanced conceptualization. Hence, what will be explored in the following sections refers to the concept of <Laparoendoscopic single-site total hysterectomy>.

8.3. The conceptual side of LESS

Bearing in mind all of the above, as well as the conceptual framework presented in Chapter 6, it was considered that the information concerning the use of a single incision to access the operating site within the scope of a laparoscopic procedure refers to the HOW (<Surgical approach>), not to the WHAT (<Surgical action>) or to the WHERE (<Anatomical entity>). Therefore, an attempt was made to integrate this conceptual

¹⁶⁵ The different spellings (with and without the hyphen in single port) correspond to the actual results in SNOMED CT).

information into the existing <Surgical approach> hierarchy, thus generating a new iteration of this micro-conceptual system. To maintain the specific difference, Aristotelian-based approach that would guarantee the required consistency, another axis of analysis was devised, based on what was perceived as the **delimiting** characteristic (cf. ISO 1087-1, 2000) of this approach. In this case, as is shown in **Figure 60**, the new axis corresponds to the number of skin incisions and is accompanied by its respective differences (/single skin incision/ vs. /multiple skin incisions/), so as to allow concept differentiation within the <Minimally invasive surgical approach with skin incision> (or the <Laparoscopic approach> depicted in the previous concept system).

Hence, the <Minimally invasive surgical approach with single skin incision> would correspond to the notion of LESS, whereas the <Minimally invasive surgical approach with multiple skin incisions> would allow the system to accommodate the conventional, multi-port laparoscopic surgical approach. The concepts of <Laparoendoscopic single-site approach> and <Conventional laparoscopic approach> were also added to the semi-formal micro-concept system, not only to pave the way to other future concepts that might also be subtypes of <Minimally invasive surgical approach with single skin incision> and <Minimally invasive surgical approach with multiple skin incisions>, respectively, but also to facilitate a possible integration into biomedical TS, since the resulting 'concept name' (cf. Roche 2012, 2015), as such, would not match the current nomenclature used by these biomedical resources.

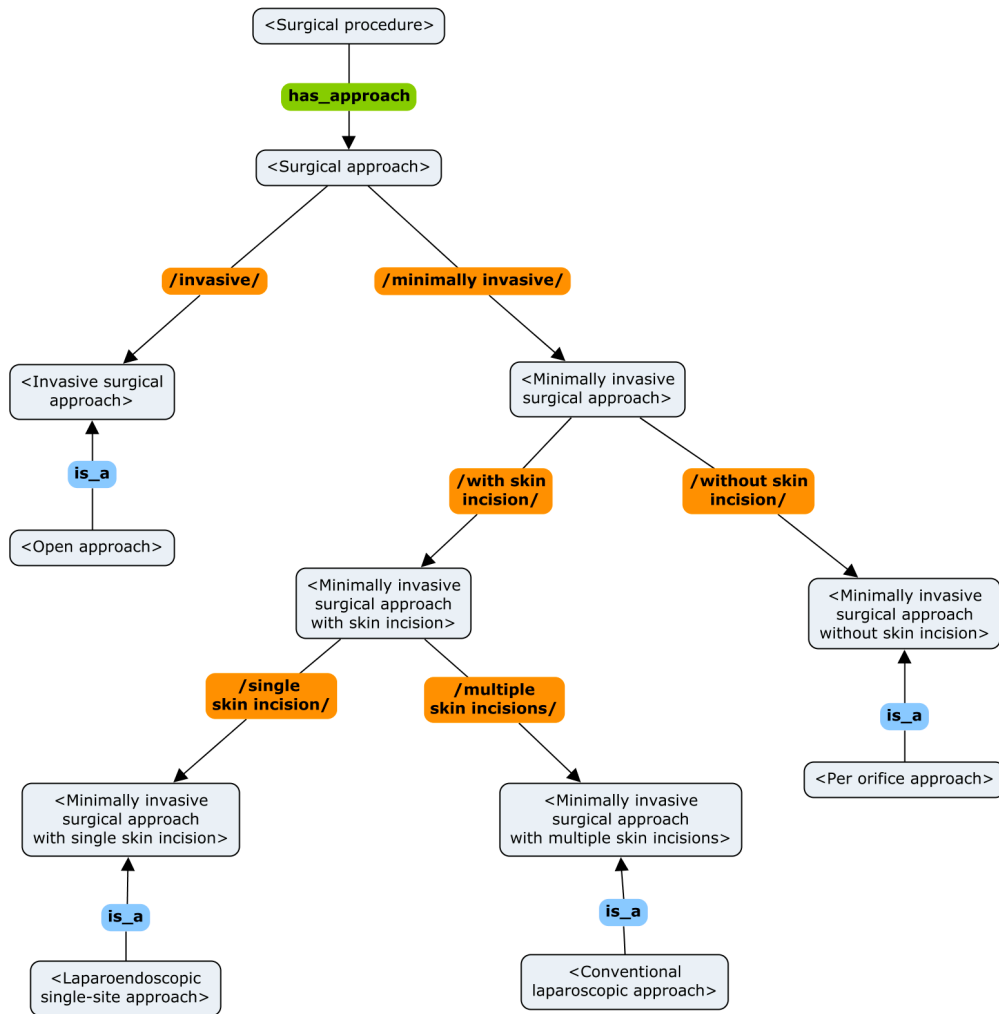


Figure 60 – New iteration of the micro-concept system on <Surgical approach>

However, when applying this micro-conceptual system to a concrete <Surgical procedure>, such as <Hysterectomy>, one needs to find a way to connect the <Surgical approach> hierarchy to the remaining hierarchies that is still logically consistent. Since a <Laparoendoscopic single-site hysterectomy> is a subtype of <Laparoscopic hysterectomy> via their <Surgical approach>, and since there is no other <Minimally invasive surgical approach with single skin incision> at the moment besides the <Laparoendoscopic single-site approach>, the information would be redundant if it was introduced into the system only via the <Surgical approach> hierarchy. In fact, the procedure <Laparoendoscopic single-site hysterectomy> would inherit **both** <Surgical approach> values under the has_approach conceptual relation without any particular

need for it to happen. For those reasons, it was decided that under the <Surgical approach> hierarchy, and for now, only the ‘concept name’ stemming from the specific difference exercise would be used, i.e. <Minimally invasive surgical approach with single skin incision>. In turn, the concept of <Conventional laparoscopic hysterectomy> would be added to the <Laparoscopic hysterectomy> / <Hysterectomy> / <Surgical procedure> hierarchy instead, in order to acquire the <Surgical approach> value of <Minimally invasive surgical approach with multiple skin incisions>, thereby allowing a differentiation towards the aforementioned <Laparoendoscopic single-site hysterectomy>. Moreover, it is believed this helps avoid logical discrepancies now and in the future. As a result, and when applying this new iteration to our example of <Laparoendoscopic single-site total hysterectomy>, it becomes clear that this concept has two ‘parents’, or supertype concepts (<Laparoendoscopic single-site hysterectomy> and <Total hysterectomy>), thereby inheriting their specific differences, as well as their attribute relationships and respective values (cf. Figure 61).

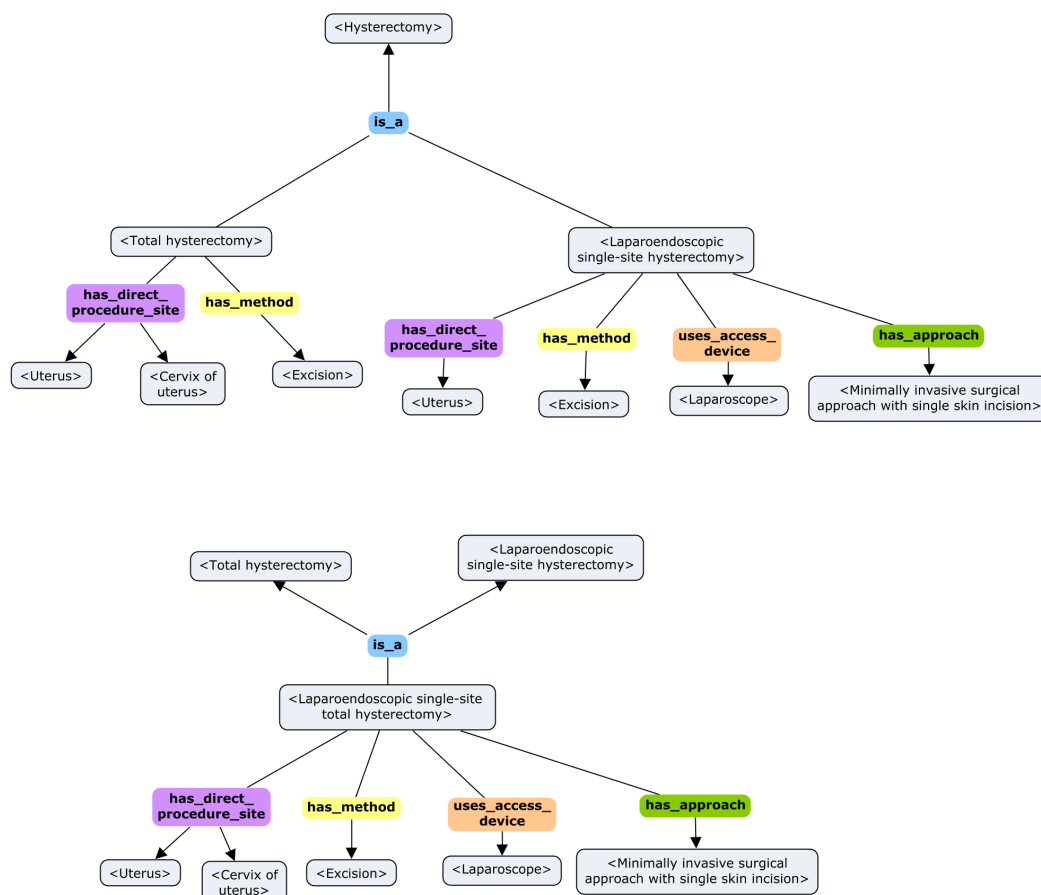


Figure 61 – Micro-concept system for <Laparoendoscopic single-site total hysterectomy>

It should also be noted that due to the synergies involving the specific difference approach, the categorial structure, and the non-hierarchical relations, the conceptualization acquires a flexibility and a logical consistency that enables content expansion. The figure below (**Figure 62**) depicts an example concerning the concept of <Laparoscopic single-site ovarian cystectomy>. An <Ovarian cystectomy> is also a very common procedure in cases of endometriosis and is usually performed via minimally invasive approaches. Again, such a concept does not exist in the several biomedical TS at the present time, but the framework provided by EndoTerm could contribute to its integration.

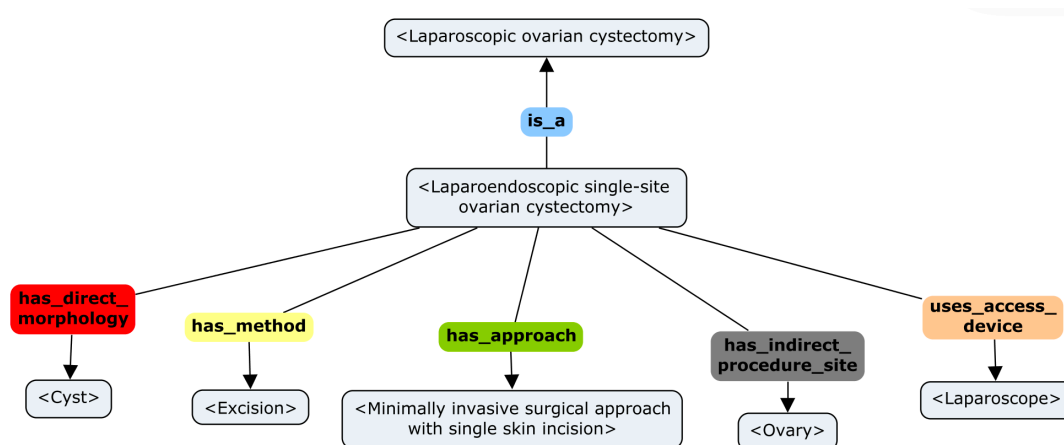


Figure 62 – Micro-concept system for <Laparoendoscopic single-site ovarian cystectomy>

Also anchored in EndoTerm’s categorial structure, this concept representation differs from the previous one in the sense that what is being directly removed is a <Morphologic abnormality>, or lesion (<Cyst>) (hence the has_direct_morphology relation), which, in turn, is located in the <Ovary> (that will expectedly not be removed), which is why the has_indirect_procedure_site relation is used.

After having stabilized those micro-concept systems, the work proceeded to their validation – from a logical perspective – and implementation through TeDI (short for OntoTerminology **EDI**tor). Created by (Roche, 2017), TeDI is a software environment dedicated to the development of multilingual ontoterminologies (cf. Chapter 2), which

allows the user to develop a formal ontology that can be shared by different sets of terms in various languages.

In order to make that implementation possible, TeDI is structured around two independent, yet interrelated modules: on the one hand, the Concept Editor is supported by both essential (axes of analysis) and descriptive (attributes) characteristics, providing a set of epistemological principles based on an Aristotelian, specific difference approach (cf. Chapter 2). Hence, logical properties can be verified throughout the ontology building process, thereby preventing non-sense representations. On the other hand, the Term Editor is where the terms and other related information (e.g. notes, sources, spelling variants, term status, etc.) are inserted.

By making the connection between the terms and the common conceptual framework in an explicit way, the system is able to automatically calculate **terminological equivalents** (i.e. when two terms from different languages denote the same concept) and **terminological synonyms** (in other words, when two terms within the same language denote the same concept) (cf. (Costa, 2017)).

Due to all these features, TeDI was the chosen tool for the validation and implementation of EndoTerm's conceptual module. Thus, it was possible to insert into TeDI the information concerning EndoTerm's semi-formal micro-concept systems (the concepts, relations – hierarchical and non-hierarchical, and the categorial structure) and convert them into a formal ontology, also benefiting from the tool's built-in reasoner and from the subsequent logical verification that takes place during the ontology development process. It is believed, though, that TeDI's important role at this stage of the project was also greatly supported and enhanced by the 'conceptual homework' that upheld the development of EndoTerm's conceptual foundation and the conception of its semi-formal micro-concept systems. In fact, due to this previous work, the inserted content was successfully validated without the detection of any logical errors. It should also be noted that because TeDI was still under development at the time this part of the project was carried out, the stand-alone version that was used was 0.98, created on July 7, 2017.

The image below (**Figure 63**) shows a glimpse of TeDI's concept editor, in this case focusing on the aforementioned concept of <Laparoendoscopic single-site total hysterectomy>, namely: [1] its position in the overall hierarchy of <Surgical procedures>; [2] the fact that TeDI supports polyhierarchy; [3] one of the non-hierarchical relationships (**has_approach**); and [4] the Value concepts inherited from the <Surgical approach> hierarchy. In this case, because a <Laparoendoscopic single-site total hysterectomy> is_a <Laparoscopic hysterectomy>, it inherits the <Surgical approach> value of <Minimally invasive surgical approach with skin incision>. Moreover, because it is a <Laparoendoscopic single-site hysterectomy>, it also inherits the value of <Minimally invasive surgical approach with single skin incision>.

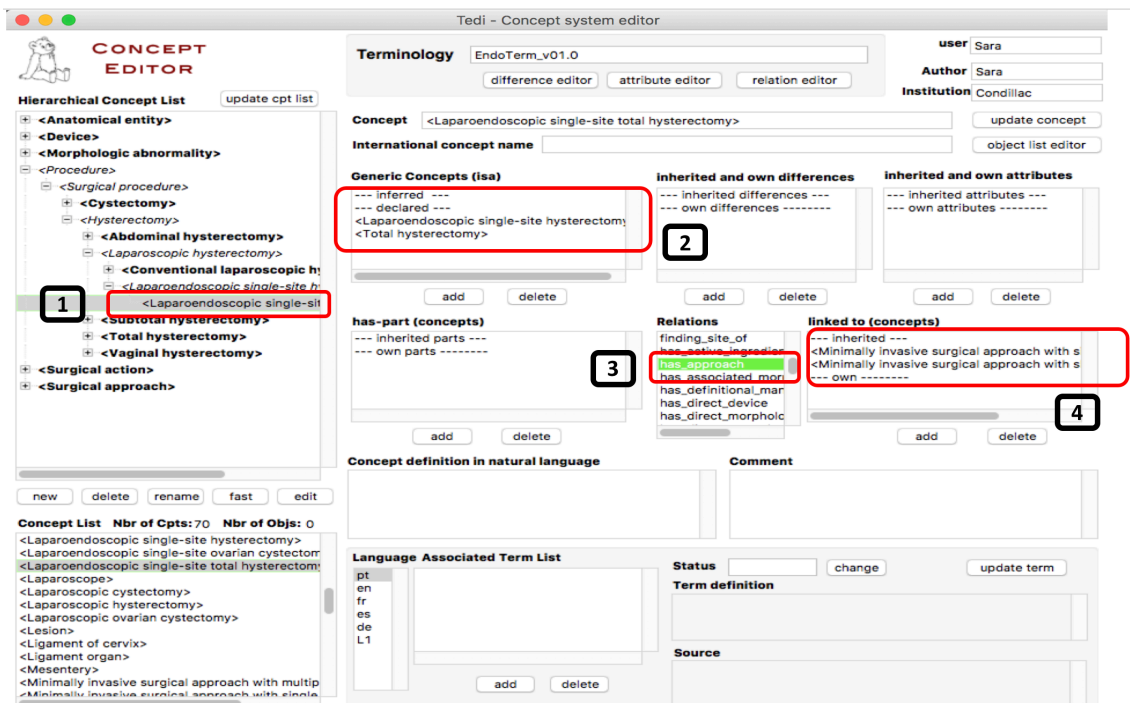


Figure 63 – <Laparoendoscopic single-site total hysterectomy> in TeDI's concept editor

The following image (**Figure 64**) illustrates, in turn, how the concept of <Minimally invasive surgical approach with single skin incision> integrates the specific differences depicted in the preceding CMap, namely that this <Surgical approach> is /minimally invasive/, /with skin incision/ and with /single skin incision/. These specific differences and their respective axes of analysis were previously inserted in the difference editor available at TeDI [right-hand side of the image].

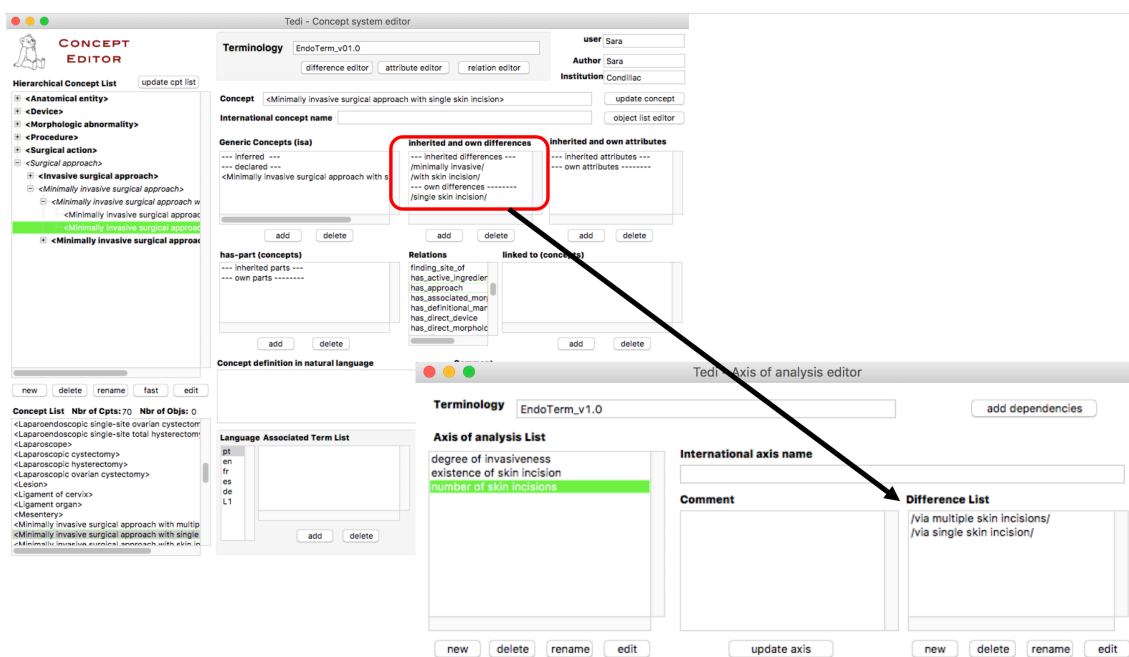


Figure 64 – <Minimally invasive surgical approach with single skin incision> and its specific differences

To set the ground for future work regarding EndoTerm, namely the expansion of the knowledge base into other categories related to <Endometriosis>, the initial non-hierarchical conceptual relations list was extended, based, once again, on the systematization of those currently included in both the UMLS and SNOMED CT. As a result, 23 relations and their respective inverse labels (so, a total of 46 relations) were added to TeDI's version specifically developed for this project (TeDI for EndoTerm) (cf. **Table 29**).

RELATION	INVERSE
access_device_used_by	uses_access_device
active_ingredient_of	has_active_ingredient
approach_of	has_approach
associated_morphology_of	has_associated_morphology
caused_by	causes
definitional_manifestation_of	has_definitional_manifestation

device_used_by	uses_device
diagnosed_by	diagnoses
direct_device_of	has_direct_device
direct_morphology_of	has_direct_morphology
direct_procedure_site_of	has_direct_procedure_site
finding_site_of	has_finding_site
indirect_procedure_site_of	has_indirect_procedure_site
method_of	has_method
procedure_morphology_of	has_procedure_morphology
specimen_of	has_specimen
specimen_procedure_of	has_specimen_procedure
specimen_source_morphology_of	has_specimen_source_morphology
specimen_source_topography_of	has_specimen_source_topography
specimen_substance_of	has_specimen_substance
performed_by	performs
prevented_by	prevents
treated_by	treats

Table 29 – EndoTerm’s full list of non-hierarchical conceptual relations

The screenshot below (**Figure 65**) illustrates this extended list in TeDI’s relation editor, as well as an example of how the categorial structure was implemented. As can be seen on the right-hand side, the inverse relation to the one selected from the relation list on the left is identified in [1]. [2] and [3] refer to the source and target concept, respectively. In other words, domain and range, as explored in Chapter 3. By previously letting the system know which source concept can relate to which target concept via relation X, it is possible to prevent logical errors when inserting new content since TeDI then only lists the ‘authorized’ target concepts for that specific relation.

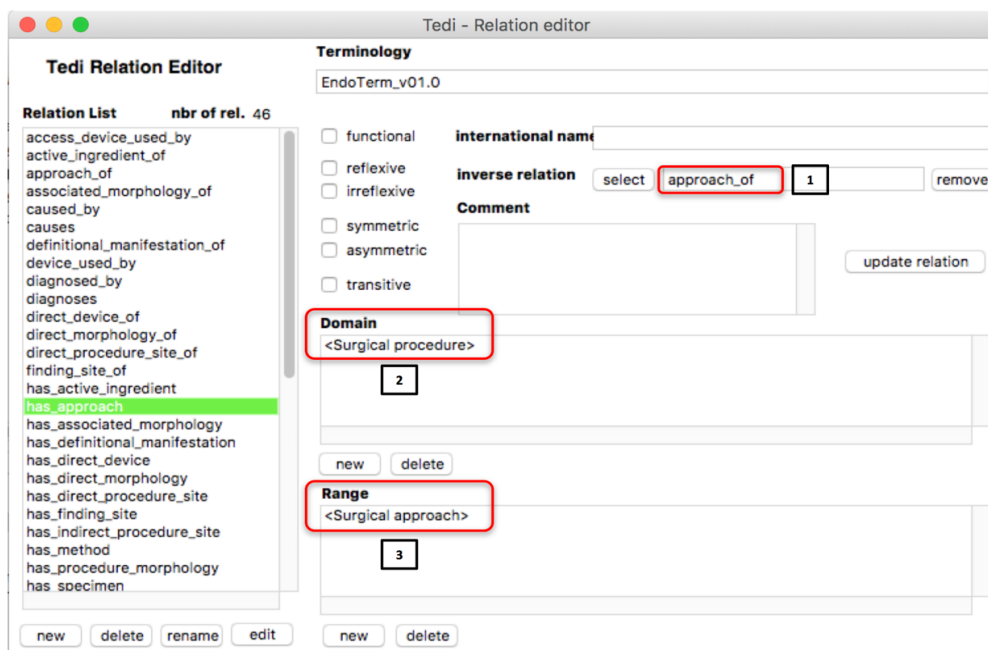


Figure 65 – An example of TeDI’s relation editor containing the extended EndoTerm’s non-hierarchical relations

As regards the sharing of EndoTerm’s conceptual content, TeDI allows the exporting of these formal concept definitions into W3C-compliant formats, namely RDF/OWL, which can pave the way to a potential integration into existing biomedical, concept-oriented terminological resources. An example is shown in **Figure 66** below.

```
<owl:Class rdf:about="Laparoendoscopic_single-site_total_hysterectomy">
  <rdfs:subClassOf rdf:resource="Laparoendoscopic_single-site_hysterectomy"/>
  <rdfs:subClassOf rdf:resource="Total_hysterectomy"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="has_direct_procedure_site"/>
      <owl:someValuesFrom rdf:resource="Uterus"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="has_direct_procedure_site"/>
      <owl:someValuesFrom rdf:resource="Cervix_of_uterus"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="has_method"/>
      <owl:someValuesFrom rdf:resource="Excision"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="has_approach"/>
      <owl:someValuesFrom rdf:resource="Minimally_invasive_surgical_approach_with_single_skin_incision"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="uses_access_device"/>
      <owl:someValuesFrom rdf:resource="Laparoscope"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```

Figure 66 – Formal definition of <Laparoendoscopic single-site total hysterectomy> in RDF/OWL

8.4. Terminological definitions in EndoTerm: two examples

Based on the conceptualization validated by TeDI's built-in reasoner, the following step consisted of verifying the initially stated hypothesis of whether or not a formal concept definition could effectively optimize the drafting of natural language definitions via a definitional template-like structure. Thus, and for each of the analyzed LESS procedures, a template was put forward with direct reference to both the specific difference approach and to the non-hierarchical relations supported by the categorial structure (Tables 30 and 31).

CONCEPT 1	
<Surgical procedure>	
has_method	<Surgical action>
has_direct_procedure_site	<Anatomical entity>
uses_access_device	<Device>
has_approach	<Surgical approach>

Table 30 – Definitional template for <Laparoendoscopic single-site total hysterectomy>

CONCEPT 2	
<Surgical procedure>	
has_method	<Surgical action>
has_direct_morphology	<Lesion>
has_indirect_procedure_site	<Anatomical entity>
uses_access_device	<Device>
has_approach	<Surgical approach>

Table 31 – Definitional template for <Laparoendoscopic single-site ovarian cystectomy>

Therefore, for the first concept, the proposed natural language definition in English is the following:

Minimally invasive surgical procedure which consists of the excision of the uterus and cervix using a laparoscope to access the site via a single skin incision.

As regards the second concept, the proposed natural language definition in English would read:

Minimally invasive surgical procedure which consists of the excision of a cyst located in the ovary using a laparoscope to access the site via a single skin incision.

Despite the existence of a term editor in TeDI, which aims to accommodate all the information that pertains to the linguistic dimension, it was decided not to use that module within the scope of this work, since its exporting functionalities into HTML were still under development in the version explored in this thesis (0.98). Hence, to illustrate the sharing of EndoTerm's linguistic content with human users, an example from the first natural language definition will be presented via Lexonomy.

Lexonomy is described as a “free, open-source, web-based dictionary writing and publishing system” where users can create a dictionary from start to finish, being able to plan the structure of the entries, edit them by resorting to a built-in XML editor, and finally publish their content online (Měchura, 2017). Since one of the objectives for future work consists of the development and dissemination of a multilingual dictionary of endometriosis (cf. Concluding remarks), exploring Lexonomy as a potential tool to use appeared to be pertinent within the scope of this part of the thesis. **Figure 67** shows the XML editor.

```
<entry>
  <headword>laparoendoscopic single-site total hysterectomy</headword>
  <partOfSpeech>n</partOfSpeech>
  <sense>
    <definition>Minimally invasive surgical procedure which consists of the excision of the uterus and cervix using a laparoscope to access the site via a single skin incision.</definition>
    <example>"The aim of this study was to compare the surgical outcomes of robotic single-site and laparoendoscopic single-site total hysterectomy and to evaluate the feasibility of the former in patients with benign gynecologic disease." [Source: Paek et al., 2016. URL: https://www.ncbi.nlm.nih.gov/pubmed/26092018]</example>
  </sense>
</entry>
```

Figure 67 – Natural language definition using Lexonomy: entry with the XML Editor

Figure 68, on the other hand, depicts the entry as it would be shown to the user. The author of the dictionary can then choose to make his/her lexicographic product public or not. If it goes public, a URL is created and made available under a Creative Commons license.

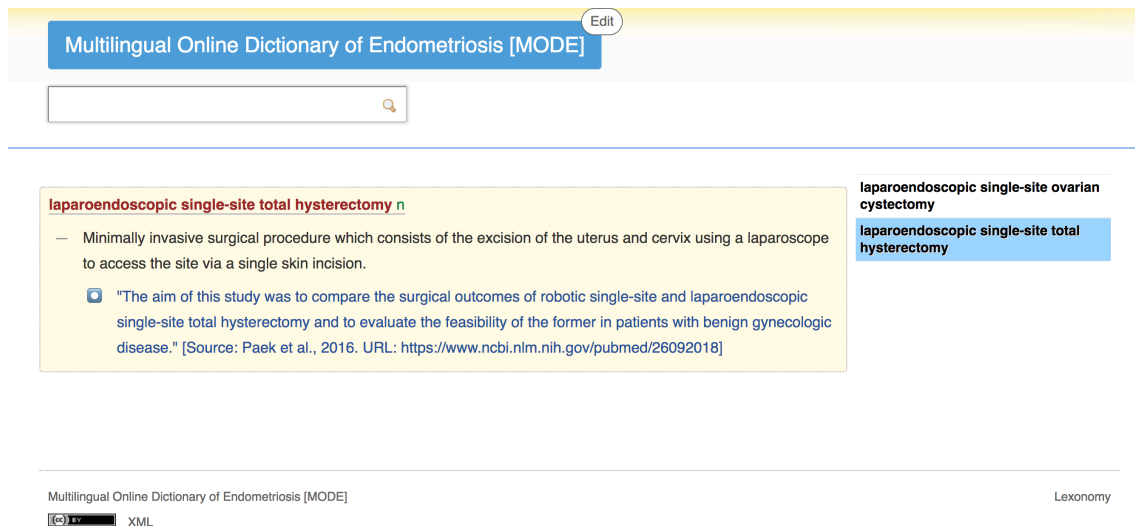


Figure 68 – Natural language definition using Lexonomy: entry as shown in the public dictionary

Although this entry depicts a basic structure, since it has been used merely to illustrate how the previously drafted natural language definitions could be shared among human end-users, Lexonomy's XML structure provides the necessary flexibility so that the author of the dictionary can customize the information that is to be included in that entry.

Overall, this section also aimed to demonstrate the added value that such a template-based methodology can have in contexts where natural language definitions need to be drafted, namely since it provides logical consistency, focuses on the essential and delimiting characteristics of the concept in question, and contributes to optimize the drafting process. These advantages also apply to multilingual settings.

8.5. SNOMED CT integration: a proposal

Taking into account all the linguistic and conceptual information gathered so far, the final section of this chapter aims to contribute to the enrichment of one of the most widespread biomedical terminological resources today: SNOMED CT. This resource has been selected as the basis of our proposal due to several reasons: **firstly**, the concept orientation that characterizes this resource and the fact that it aims to integrate the linguistic and the conceptual, while preserving their fundamental differences, appear to be consistent with Terminology's double dimensional approach and its core principles (cf. Chapter 2); **secondly**, SNOMED CT's structure allows post-coordination, i.e. more complex concepts may be created from a set of more primitive components. As shown in chapter 3, in this type of compositional system, there is no need to create all the elements in advance, but rather to ensure that all the basic building blocks exist. It is therefore possible to represent a given clinical content even when the precise concept is not present in SNOMED CT. Moreover, this compositional approach to concept representation requires the definition of a set of logical rules (constraints) that will govern the way concepts and relationships can be combined, in order to prevent nonsense representations; **thirdly**, and unlike other resources of its kind, SNOMED CT is not limited to hierarchical concept relations. Again, both this and the previous topic are in line with EndoTerm's complementary approach based on an Aristotelian-type hierarchy and on a categorial structure, composed by non-hierarchical relations, a set of categories, and their respective constraints; and **finally**, the concept that has been used as an example <Laparoendoscopic single-site total hysterectomy> currently does not exist in this terminological system, and given its increasing prevalence so it is believed that this proposal, supplemented by the methodology of drafting natural language definitions, could contribute to enrich SNOMED CT's content.

Therefore, the proposal would be as follows: as regards the descriptions¹⁶⁶ (the linguistic dimension), the choice of the Fully Specified Name and respective Synonyms

¹⁶⁶ For consistency purposes, SNOMED CT's terminology will be maintained in our proposal. Thus, instead of adopting the notion of **designation** (ISO 1087-1, 2000), the original SNOMED CT expression "description" is used.

respects the position issued by LESSCAR and supported by the aforementioned MEDLINE/PubMed® searches, as seen below (**Figure 69**):

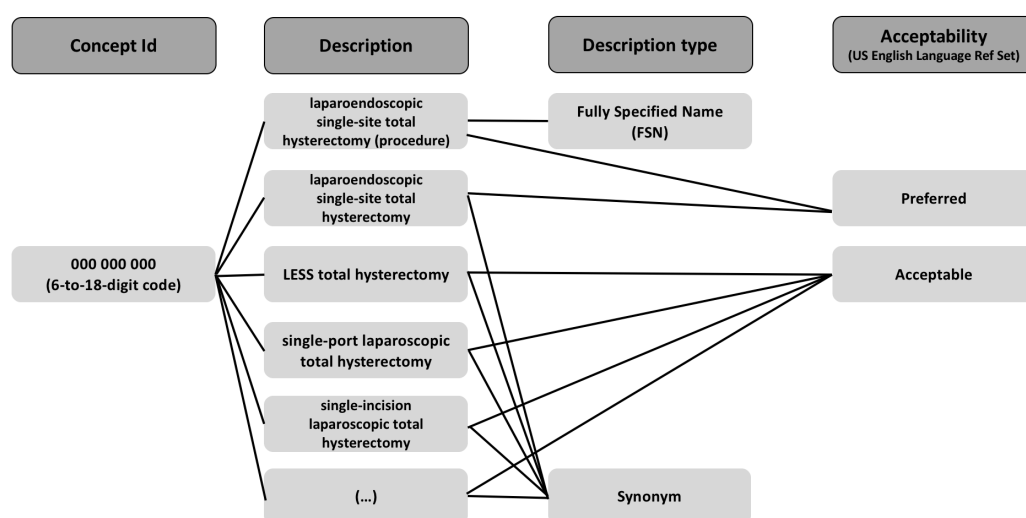


Figure 69 – Example of SNOMED CT descriptions for <Laparoendoscopic single-site total hysterectomy>

As for the conceptual dimension, which is the basis of concept definitions in SNOMED CT, the biggest challenge – and, to a certain extent, opportunity – lies in the absence, in this resource, of any [Surgical access value (qualifier value)] (i.e. EndoTerm's <Surgical approach>) for the attribute [Access] that exposes this differentiation between a conventional laparoscopic approach (with multiple skin incisions) and the one resorting to a single skin incision, which reflects the delimiting characteristic of the concept of <LESS>. In fact, in the current version of SNOMED CT there is no actual distinction, within the formal concept definition, between a [Single-port laparoscopic cholecystectomy (procedure)] and its supertype [Laparoscopic cholecystectomy (procedure)]. It is believed that such a distinction could be relevant and should be clearly stated in a person's EHR via SNOMED CT, since the differences in approach may have implications in terms of cosmetic results, postoperative pain, recovery/hospitalization time. It could also be pertinent in what concerns medical coding, insurance, and billing. Indeed, it is not the same, as regards the preceding topics, to undergo a surgery where three or four incisions are made, or one where there is only a single incision.

The concept in question could, thus, be defined through a combination of is_a and attribute relationships, represented via SNOMED CT's compositional grammar. As shown in chapter 3, its basis is a lightweight syntax that is both human-readable as well as machine-processable, thereby supporting interoperability. The proposal below (**Figure 70**) resorts, as much as possible, to existing concepts, descriptions, and relationships in SNOMED CT. The suggested content has been signaled in red.

```

000000000 |Laparoendoscopic single-site total hysterectomy (procedure)|

=== 236888001 |Laparoscopic total hysterectomy (procedure)|:
  {405813007 |Procedure site - Direct (attribute)| = 181452004 |Entire uterus (body structure)|,
  260686004 |Method (attribute)| = 129304002 |Excision – action (qualifier value)| ,
  425391005 |Using access device (attribute)| = 86174004 |Laparoscope, device (physical object)|,
  260507000 |Access (attribute)| = 000000000 |Minimally invasive surgical approach with single skin incision –
  access (qualifier value)|}

```

Figure 70 – Example of formal concept definition for <Laparoendoscopic single-site total hysterectomy> using SNOMED CT's compositional grammar

So, the first concept represents the is_a relationship and is followed by a refinement, which is introduced by a colon and consists of a sequence of four attribute-value pairs. The attribute is separated from the value by an equals sign. The different attribute-value pairs are separated by commas. The curly braces represent grouping of attributes within a refinement.

Despite the fact that SNOMED CT does not currently encompass any natural language definitions, nor any guidelines or drafting principles, as stated earlier, its solid concept orientation will undoubtedly constitute a valuable framework in that almost inevitable next step. And when that happens, perhaps EndoTerm can help enhance the yet rather marginal presence of certain <Surgical procedures> - and other endometriosis-related concepts – in this biomedical terminological resource.

Concluding remarks

Within the biomedical domain, changes have been substantial in the past few decades: on the one hand, health care provision has become more technology-based, with computerized examinations, procedures, prescriptions and health records. Furthermore, growing digital literacy has brought the patients into the driver's seat, where they have been playing a more active – and empowered – role. On the other hand, ageing population and the dramatic decline of the old-age support ratio have contributed to more pressure on public health expenditure, leading to the existing debate around the sustainability of social security systems and their role in healthcare. The ability to provide secure, reliable, efficient and cost-effective ways to process and exchange clinical information among all the stakeholders has emerged as the cornerstone of eHealth and mHealth initiatives worldwide.

As a scientific discipline in the confluence of several others (e.g. logic, information science, cognitive sciences, linguistics, etc.), Terminology brings an unquestionable added value as regards the representation, organization, dissemination and, ultimately, the stabilization of knowledge among a community of practice. Nowadays, however, that contribution can only be further enhanced if the results of that terminological work can be operationalized, i.e. represented in a computational format. Ontologies, in the sense of KE, represent a promising pathway that, nonetheless, must be based on collaborative work and on solid theoretical and methodological approaches. By acknowledging Terminology's linguistic and conceptual dimensions and by applying the resulting theoretical and methodological principles to the creation of terminological resources, multidisciplinary teams integrating both 'linguist-terminologists' and 'engineer-terminologists' will be able to respond more effectively to the growing needs of expert communities and, increasingly, of society as a whole (Carvalho et al., 2015).

By resorting to a case study approach, this research work aimed to reflect upon the fact that the conceptual systems produced directly from expert feedback or from concept-oriented biomedical terminological resources, on the one hand, and the semantic networks resulting from corpus analysis, on the other hand, may not always match, but both play a vital role in terminology work: through ontologies and the use of

W3C-compliant formats, conceptualization proposals open new possibilities in the Brave New World of healthcare, where interoperability, the Semantic Web, and Linked Data are playing an indispensable role; albeit with vagueness and inconsistencies, specialized corpora provide fundamental access to the expert community, contributing to stabilize knowledge in areas of expertise where technological innovation is constant and unprecedented, as is the case of surgery. When anchored in this double dimension, terminology work may contribute to further enhance that stability and, consequently, the quality of specialized communication. Based on the previous background, the EndoTerm project also attempted to illustrate the dynamic and iterative nature of the terminologist's workflow when devising a knowledge organization proposal.

Furthermore, this thesis aimed to demonstrate that conceptual representations, in this case an ontology supported by a combination of the specific difference approach and a categorial structure for procedure concepts, can make a valuable contribution to the drafting of natural language definitions, which can, in turn, be targeted at some of the current biomedical terminological resources, since most of them lack this feature. By providing a systematized and clear framework of interrelated concepts, relations, and domain constraints, these conceptualizations can become useful allies against the limitations of the so-called traditional terminological definitions. It is believed that such a template-based methodology as the one presented here can also play a key role in optimizing the creation of natural language definitions in multiple languages.

In any major undertaking, there is always room for improvement, and this case was no exception. Perhaps the most significant setback in this project was the fact that the validation of both candidate terms and natural language definitions by subject field experts could not take place within the timeframe of the thesis, as opposed to what had been initially planned. Therefore, one of our goals as regards future work is to reinforce that collaboration, particularly with Portuguese experts, aiming at the development of EndoTerm, namely via the expansion into other procedures related to endometriosis (not necessarily surgical), as well as into signs/symptoms, medical treatment, disease location, etc. Moreover, and through the support of MulherEndo, the Portuguese Association Supporting Women with Endometriosis, products can be devised that are mainly targeted at endometriosis patients and expert patients in this country. As a

matter of fact, although many significant steps have been taken in the last few years in Portugal concerning the awareness of the general population about the disease, a lot more still lies ahead. It is believed that the foundations that have been placed by EndoTerm can make a valid contribution in this respect, from the production of more comprehensive brochures/materials or press kits where, for instance, adapted versions of micro-concept systems could be included, to the creation of training materials for schools, where early awareness can trigger an early response and, as a result, lead to possibly less permanent damage or a better QoL for the countless endometriosis patients.

However, the objective is not only to support the production of patient-oriented content in European Portuguese, but to also build up on this thesis work in order to create a specialized, multilingual lexicographic resource that can help a larger number of patients. More specifically, the idea has arisen to develop the Multilingual Online Dictionary of Endometriosis (MODE), in collaboration with both subject field experts and with patient organizations from different countries, so as to include, at a first stage, English, European Portuguese, and French. For that, the EndoCorpus would also have to expand, in size and formats, given the enormous potential that images, but especially videos – namely medical video articles – could bring. Hence, the ‘M’ in MODE could also stand for Multimodal. In addition, one could go one step further and envision the publication of MODE content also as Linguistic Linked Open Data, resorting, for instance, to the LEMON-Ontolex model (cf. Chapter 1). Within the medical field, and given the current and ongoing technological progress, it is likely that some sort of multimedia knowledge base may become a more and more common resource in both patient and expert education, for instance. It is therefore believed that the study and inclusion of non-verbal and multimedia elements in such resources would represent an important qualitative leap in the joint research involving Terminology and Knowledge Engineering.

The ongoing challenges regarding the way medical information and knowledge are produced, used, stored and shared require increasingly efficient and reliable solutions, in a society that demands immediate and multi-platform access to all digital content. If one of the main postulates of terminology work is to provide tools and services that can respond to the concrete needs of a given target audience, at a certain

moment in time, concerning a specific domain, and under particular circumstances, then terminological projects developed within the subject field of healthcare, especially those focusing on knowledge representation, knowledge organization and knowledge sharing, must take this whole background into consideration.

Bibliography

- Acien, P., & Velasco, I. (2013). Endometriosis: a disease that remains enigmatic. *ISRN Obstetrics and Gynecology*, Article ID 242149.
- Acosta, A. A., Buttram, V. C., Besch, P. K., Malinak, L. R., Franklin, R. R., & Vanderheyden, J. D. (1973). A proposed classification of pelvic endometriosis. *Obstetrics and Gynecology*, 42(1), 19–25. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/4720202>
- Adamson, G. D. (2011). Endometriosis classification. *Current Opinion in Obstetrics and Gynecology*, 23(4), 213–220. <https://doi.org/10.1097/GCO.0b013e328348a3ba>
- Adamson, G. D. (2013). Endometriosis Fertility Index. *Current Opinion in Obstetrics and Gynecology*, 25(3), 186–192. <https://doi.org/10.1097/GCO.0b013e32836091da>
- Adamson, G. D., Kennedy, S., & Hummelshoj, L. (2010). Creating solutions in endometriosis: global collaboration through the World Endometriosis Research Foundation. *Journal of Endometriosis*, 2(1), 3–6.
- Adamson, G. D., & Pasta, D. J. (2010). Endometriosis fertility index: the new, validated endometriosis staging system. *Fertility and Sterility*, 94(5), 1609–1615. <https://doi.org/10.1016/j.fertnstert.2009.09.035>
- Ahn, S. H., Singh, V., & Tayade, C. (2017). Biomarkers in endometriosis: challenges and opportunities. *Fertility and Sterility*, 107(3), 523–532. <https://doi.org/10.1016/j.fertnstert.2017.01.009>
- Aluttis, C., Bishaw, T., & Frank, M. W. (2014). The workforce for health in a globalized context – global shortages and international migration. *Global Health Action*, 7(1), 23611. <https://doi.org/10.3402/gha.v7.23611>
- Anthony, L. (2014). AntConc Tutorial (ver. 3.4.0). Retrieved January 24, 2018, from <https://www.youtube.com/user/AntlabJPN>
- Anthony, L. (2018). AntConc (Version 3.5.7). Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/software/antconc/>
- Aristotle. (1989). *Metaphysics*, Book 4. Retrieved February 10, 2018, from <http://www.perseus.tufts.edu/hopper/text?doc=Perseus:abo:tlg,0086,025:4>
- Arruda, M. S., Petta, C. A., Abrão, M. S., & Benetti-Pinto, C. L. (2003). Time elapsed from onset of symptoms to diagnosis of endometriosis in a cohort study of Brazilian women. *Human Reproduction*, 18(4), 756–759.
- ASRM. (1997). Revised American Society for Reproductive Medicine classification of endometriosis. *Fertility and Sterility*, 67(5), 817–821.
- ATIH. (2017). Classification Commune des Actes Médicaux - Descriptive à Usage PMSI. Agence Technique de l'Information sur l'Hospitalisation - Ministère des Affaires Sociales et de la Santé.
- Balabanič, D., Rupnik, M., & Klemenčič, A. K. (2011). Negative impact of endocrine-disrupting compounds on human reproductive health. *Reproduction, Fertility, and Development*, 23(3), 403–416. <https://doi.org/10.1071/RD09300>
- Ballard, K., Lowton, K., & Wright, J. (2006). What's the delay? A qualitative study of women's experiences of reaching a diagnosis of endometriosis. *Fertility and Sterility*, 86(5), 1296–1301. <https://doi.org/10.1016/j.fertnstert.2006.04.054>

- Basterrechea, B. B. (2015). Written declaration, under Rule 136 of Parliament's Rules of Procedure, on research and prevention of endometriosis. Brussels: European Parliament.
- Batt, R. (2011). *A History of Endometriosis*. London: Springer London. <https://doi.org/10.1007/978-0-85729-585-9>
- Becker, C. M., Laufer, M. R., Stratton, P., Hummelshoj, L., Missmer, S. A., Zondervan, K. T., ... Zondervan, K. T. (2014). World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonisation Project: I. Surgical phenotype data collection in endometriosis research. *Fertility and Sterility*, 102(5), 1213–1222. <https://doi.org/10.1016/j.fertnstert.2014.07.709>
- Bedaiwy, M., & Falcone, T. (2007). Endometriosis and Adenomyosis. In A. I. Sokol & E. R. Sokol (Eds.), *General gynecology : the requisites in obstetrics and gynecology* (pp. 321–346). Philadelphia: Mosby.
- Benagiano, G., & Brosens, I. (2011). Who identified endometriosis? *Fertility and Sterility*, 95, 13–16. <https://doi.org/10.1016/j.fertnstert.2010.06.027>
- Benagiano, G., Brosens, I., & Habiba, M. (2015). Adenomyosis: a life-cycle approach. *Reproductive BioMedicine Online*, 30(3), 220–232. <https://doi.org/10.1016/j.rbmo.2014.11.005>
- Benagiano, G., Brosens, I., & Lippi, D. (2015). Endometriosis: Ancient or modern disease? *The Indian Journal of Medical Research*, 141(2), 236–238. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/25900960>
- Benson, T., & Grieve, G. (2016). Why Interoperability Is Hard. In T. Benson & G. Grieve (Eds.), *Principles of Health Interoperability* (pp. 19–35). Springer. https://doi.org/10.1007/978-3-319-30370-3_2
- Bergenholtz, H., & Kaufmann, U. (1997). Terminography and Lexicography. A Critical Survey of Dictionaries from a Single Specialised Field. *Hermes, Journal of Linguistics*, 18, 91–125.
- Berners-Lee, T. (2006). Linked Data - Design Issues. Retrieved February 4, 2016, from <https://www.w3.org/DesignIssues/LinkedData.html>
- Berners-Lee, T. (2009). The next web. Retrieved July 23, 2015, from https://www.ted.com/talks/tim_berners_lee_on_the_next_web
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. Retrieved March 14, 2014, from <http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html>
- BFO. (2017). Basic Formal Ontology (BFO). Retrieved January 11, 2018, from <http://basic-formal-ontology.org/>
- Bianconi, L., Hummelshoj, L., Coccia, M. E., Vigano, P., Vittori, G., Veit, J., ... D'Hooghe, T. (2007). Recognizing endometriosis as a social disease: the European Union-encouraged Italian Senate approach. *Fertility and Sterility*, 88(5), 1285–1287. <https://doi.org/10.1016/j.fertnstert.2007.07.1324>
- Birnbaum, L. S., & Cummings, A. M. (2002). Dioxins and endometriosis: a plausible hypothesis. *Environmental Health Perspectives*, 110(1), 15–21. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11781160>
- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, (Special Issue on Linked Data).
- Bodenreider, O. (2004). The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Research*, 32(Database issue), D267–70. <https://doi.org/10.1093/nar/gkh061>

- Boehr, D. L., & Bushman, B. (2018). Preparing for the Future: National Library of Medicine's® Project to Add MeSH® RDF URIs to its Bibliographic and Authority Records. *Cataloging & Classification Quarterly*, 56(2–3), 262–272. <https://doi.org/10.1080/01639374.2017.1382642>
- Booger, E. A., Arts, T., Engelen, L. J., & van de Belt, T. H. (2015). "What Is eHealth": Time for An Update? *JMIR Research Protocols*, 4(1), e29. <https://doi.org/10.2196/resprot.4065>
- Borst, W. N. (1997). *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*. University of Twente.
- Bourigault, D., & Slodzian, M. (1999). Pour une terminologie textuelle. *Terminologies Nouvelles*, 19, 29–32.
- Bowker, L. (1997). Multidimensional Classification of Concepts and Terms. In G. Budin & S. E. Wright (Eds.), *Handbook of Terminology Management: Volume 1: Basic Aspects of Terminology Management* (pp. 133–143). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Bowker, L., & Pearson, J. (2002). *Working with specialized language : a practical guide to using corpora*. Routledge.
- Bresnick, J. (2015). How Can Healthcare Big Data Analytics Bust Data Silos? Retrieved May 13, 2018, from <https://healthitanalytics.com/news/how-can-healthcare-big-data-analytics-bust-data-silos>
- Brosens, I., & Benagiano, G. (2011). Endometriosis, a modern syndrome. *The Indian Journal of Medical Research*, 133, 581–593. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/21727656>
- Bulletti, C., Coccia, M. E., Battistoni, S., & Borini, A. (2010). Endometriosis and infertility. *Journal of Assisted Reproduction and Genetics*, 27(8), 441–447. <https://doi.org/10.1007/s10815-010-9436-1>
- Bulun, S. E. (2009). Endometriosis. *New England Journal of Medicine*, 360(3), 268–279. <https://doi.org/10.1056/NEJMr0804690>
- Butler, R., Mullin, R., Grant, T., Averill, R., & Steinbeck, B. (2012). ICD-10-PCS Reference Manual. 3M Health Information Systems.
- Buttram, V. C. (1987). Classification of Endometriosis. In *Endometriosis International Symposium, Clemont-Ferrand, October 1986* (Vol. 16, pp. 73–83). Basel: Karger Publishers. <https://doi.org/10.1159/000414851>
- Cabré Castellví, M. T. (2003). Theories of terminology: Their description, prescription and explanation. *Terminology*, 9(2), 163–199. <https://doi.org/10.1075/term.9.2.03cab>
- Cabré, M. T. (1999). *Terminology: Theory, Methods, and Applications*. (J. Sager, Ed.). Amsterdam: John Benjamins Publishing Company.
- Canis, M., Bouquet De Jolinières, J., Wattiez, A., Pouly, J. L., Mage, G., Manhes, H., & Bruhat, M. A. (1993). Classification of endometriosis. *Bailliere's Clinical Obstetrics and Gynaecology*, 7(4), 759–774. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8131314>
- Canis, M., Pouly, J. L., Wattiez, A., Manhes, H., Mage, G., & Bruhat, M. A. (1992). Incidence of bilateral adnexal disease in severe endometriosis (revised American Fertility Society [AFS], stage IV): should a stage V be included in the AFS classification? *Fertility and Sterility*, 57(3), 691–692.
- Carvalho, S., Costa, R., & Roche, C. (2016a). LESS Can Indeed Be More: Linguistic and Conceptual Challenges in the Age of Interoperability. In B. Erdman Thomsen, Hanne; Pareja-Lora, Antonio; Nistrup Madsen (Ed.), *Proceedings of the 12th International conference on Terminology and Knowledge Engineering* (pp. 157–167). Copenhagen: Copenhagen Business School. CBS; Department of International Business Communication and Politics; IBC; Department of International Business Communication and Politics.

- Carvalho, S., Costa, R., & Roche, C. (2016b). Ontoterminology meets Lexicography: the Multimodal Online Dictionary of Endometriosis (MODE). In Ilan Kernerman, Iztok Kosem, Simon Krek, & Lars Trap-Jensen (Eds.), *Globalex 2016: Lexicographic Resources for Human Language Technology* (pp. 8–15). Portorož, Slovenia: LREC 2016 Workshop.
- Carvalho, S., Roche, C., & Costa, R. (2015). Ontologies for terminological purposes: the EndoTerm project. In Thierry Poibeau & Pamela Faber (Eds.), *Proceedings of the 11th International Conference on Terminology and Artificial Intelligence* (pp. 17–27). Granada, Spain: Universidad de Granada. Retrieved from http://ceur-ws.org/Vol-1495/paper_30.pdf
- Celsus. (1971). De Medicina, Book III. *Perseus Catalog*. Boston: Harvard University Press.
- CEN. (2005). CEN EN 12264:2005 - Health informatics. Categorial structures for systems of concepts. Brussels: European Committee for Standardization.
- Chern, B. S. M., Lakhota, S., Khoo, C. K., & Siow, A. Y. M. (2012). Single incision laparoscopic surgery in gynecology: Evolution, current trends, and future perspectives. *Gynecology and Minimally Invasive Therapy*, 1(1), 9–18. <https://doi.org/10.1016/J.GMIT.2012.08.005>
- Chiarcos, C., Nordhoff, S., & Hellmann, S. (2012). *Linked Data in Linguistics: Representing and Connecting Language Data and Language Metadata*. Berlin: Springer.
- Chute, C. (1998). The Copernican era of healthcare terminology: a re-centering of health information systems. *Proceedings. AMIA Symposium*, 68–73.
- Chute, C. (2005). Medical Concept Representation. In *Medical Informatics* (pp. 163–182). Boston: Kluwer Academic Publishers. https://doi.org/10.1007/0-387-25739-X_6
- Chute, C., Cohn, S. P., & Campbell, J. R. (1998). A framework for comprehensive health terminology systems in the United States: development guidelines, criteria for selection, and public policy implications. ANSI Healthcare Informatics Standards Board Vocabulary Working Group and the Computer-Based Pati. *Journal of the American Medical Informatics Association : JAMIA*, 5(6), 503–510.
- Cimino, J. J. (1998). Desiderata for Controlled Medical Vocabularies in the Twenty-First Century. *Methods of Information in Medicine*, 37(4–5), 394–403.
- Cimino, J. J. (2000). From data to knowledge through concept-oriented terminologies: experience with the Medical Entities Dictionary. *Journal of the American Medical Informatics Association : JAMIA*, 7(3), 288–297. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10833166>
- Cimino, J. J. (2001). Terminology tools: state of the art and practical lessons. *Methods of Information in Medicine*, 40(4), 298–306.
- Cimino, J. J. (2006). In defense of the Desiderata. *Journal of Biomedical Informatics*, 39(3), 299–306.
- Cochrane. (2018). About us. Retrieved May 5, 2018, from <http://www.cochrane.org/about-us>
- Coiera, E. (2015). *Guide to health informatics* (3rd editio). Boca Raton: CRC Press - Taylor and Francis Group.
- Condamines, A., Josselin-Leray, A., Fabre, C., Lefevre, L., Picton, A., & Rebeyrolle, J. (2013). Using Comparable Corpora to Characterize Knowledge-Rich Contexts for Various Kinds of Users: Preliminary Steps. *Procedia - Social and Behavioral Sciences*, 95, 581–586. <https://doi.org/10.1016/j.sbspro.2013.10.685>
- Condamines, A., & Rebeyrolle, J. (2001). Searching for and identifying conceptual relationships via a corpus-based approach to a Terminological Knowledge Base (CTKB). In D. Bourigault, M. C. L'Homme, & C. Jacquemin (Eds.), *Recent Advances in Computational Terminology*, (pp. 127–148). Amsterdam/New York: John Benjamins.

- Costa, R. (2006). Plurality of Theoretical Approaches to Terminology. In Heribert Picht (Ed.), *Modern Approaches to Terminological Theories and Applications* (pp. 77–89). Bern: Peter Lang.
- Costa, R. (2013). Terminology and Specialised Lexicography: two complementary domains. *Lexicographica*, 29(1), 29–42.
- Costa, R. (2017). Les normes en terminologie - Que faire des synonymes ? *Cahiers de Lexicologie*, 2017 – 1(n° 110), 45–57. <https://doi.org/10.15122/isbn.978-2-406-07057-3.p.0045>
- Costa, R., & Silva, R. (2008). De la typologie à l'ontologie de textes. In C. Roche (Ed.), *TOTh 2008. Actes de la deuxième conférence TOTH - Annecy - 5-6 juin 2008* (pp. 3–16). Annecy: Institut Porphyre - Savoir et Connaissance.
- Costa, R., Silva, R., & Almeida, Z. S. de. (2011). L'organisation et la diffusion des connaissances terminologiques et textuelles au sein du Parlement portugais – le projet BDTT-AR. *Arena Romanistica. Journal of Romance Studies. Professional Communication and Terminology*, (7), 32–52.
- Costa, R., Silva, R., & Almeida, Z. S. de. (2013). Methodology Design for Terminology in the Portuguese Parliament. In V. Jesenšek (Ed.), *Specialised Lexicography: Print and Digital, Specialised Dictionaries, Databases* (pp. 113–126). Berlin: Walter de Gruyter.
- Costa, R., Silva, R., Barros, S., & Soares, A. L. (2012). Mediation strategies between terminologists and experts. In *Proceedings of GLAT 2012 - Terminologies: textes, discours et accès aux savoirs spécialisés* (pp. 297–308). Genova.
- Costas-Chavarri, A., & Meara, J. G. (2016). Need for a standardised procedure classification system in global surgery. *BMJ Global Health*, 1(2), e000034. <https://doi.org/10.1136/bmjgh-2016-000034>
- Cullen, T. (1908). *Adenomyoma of the uterus*. Philadelphia: W.B.Saunders Company.
- Cullen, W. (1800). *Nosology: Or, a Systematic Arrangement of Diseases, by Classes, Orders, Genera and Species*. Edinburgh: C. Stewart and Co.
- Culley, L., Law, C., Hudson, N., Denny, E., Mitchell, H., Baumgarten, M., & Raine-Fenning, N. (2013). The social and psychological impact of endometriosis on women's lives: a critical narrative review. *Human Reproduction Update*, 19(6), 625–639. <https://doi.org/10.1093/humupd/dmt027>
- de Bessé, B. (1997). Terminological Definitions. In G. Budin & S. E. Wright (Eds.), *Handbook of Terminology Management: Volume 1: Basic Aspects of Terminology Management* (pp. 63–74). Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1016/j.recli.2015.12.002>
- De Graaff, A. A., D'Hooghe, T. M., Dunselman, G. A. J., Dirksen, C. D., Hummelshoj, L., Simoens, S., ... Wulschleger, M. (2013). The significant effect of endometriosis on physical, mental and social wellbeing: results from an international cross-sectional survey. *Human Reproduction*, 28(10), 2677–2685. <https://doi.org/10.1093/humrep/det284>
- de Keizer, N. F., Abu-Hanna, A., & Zwetsloot-Schonk, J. H. (2000). Understanding terminological systems. I: Terminology and typology. *Methods of Information in Medicine*, 39(1), 16–21. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10786065>
- Desai, V. B., Guo, X. M., Fan, L., Wright, J. D., & Xu, X. (2017). Inpatient Laparoscopic Hysterectomy in the United States: Trends and Factors Associated With Approach Selection. *Journal of Minimally Invasive Gynecology*, 24(1), 151–158.e1. <https://doi.org/10.1016/j.jmig.2016.08.830>
- Donnez, J. (2012). Endometriosis: enigmatic in the pathogenesis and controversial in its therapy. *Fertility and Sterility*, 98(3), 509–510. <https://doi.org/10.1016/j.fertnstert.2012.07.1125>

- Donnez, J., Donnez, O., Lousse, J.-C., & Squifflet, J. (2012). Peritoneal, Ovarian, and Rectovaginal Endometriosis are Three Different Entities. In *Endometriosis: Science and Practice* (pp. 92–107). Oxford, UK: Wiley-Blackwell. <https://doi.org/10.1002/9781444398519.ch9>
- Donnez, J., Nisolle, M., Smoes, P., Gillet, N., Beguin, S., & Casanas-Roux, F. (1996). Peritoneal endometriosis and “endometriotic” nodules of the rectovaginal septum are two different entities. *Fertility and Sterility*, 66(3), 362–368. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8751730>
- Drouin, P. (2003). Term extraction using non-technical corpora as a point of leverage. *Terminology*, 9(1), 99–117.
- Drouin, P. (2010). TermoStat Web 3.0. Retrieved April 8, 2018, from <http://termostat.ling.umontreal.ca/>
- Duclos, C., Burgun, A., Lamy, J. B., Landais, P., Rodrigues, J. M., Soualmia, L., & Zweigenbaum, P. (2014). Medical Vocabulary, Terminological Resources and Information Coding in the Health Domain. In A. Venot, A. Burgun, & C. Quantin (Eds.), *Medical Informatics, e-Health* (pp. 11–41). Springer. https://doi.org/10.1007/978-2-8178-0478-1_2
- Dunselman, G., Vermeulen, N., Becker, C., Calhaz-Jorge, C., & D’Hooghe, T. (2014). ESHRE guideline: management of women with endometriosis. *Human Reproduction*, 29(3), 400–412.
- Durán-Munoz, I. (2016). Producing frame-based definitions. *Terminology*, 22(2), 223–249. <https://doi.org/10.1075/term.22.2.04mun>
- EC. (2012). *eHealth Action Plan 2012-2020 - Innovative healthcare for the 21st century*. Brussels: European Commission.
- EC. (2014). Green Paper on mobile health (“mHealth”). Brussels: European Commission.
- Eidam, S., Redenz, A., Sonius, D., & Stein, N. vom. (2017). Ubiquitous Healthcare — Do the Health and Information Technology Sectors Converge? *International Journal of Innovation and Technology Management*, 14(06), 1750039. <https://doi.org/10.1142/S0219877017500390>
- Elgazzar, K., Aboelfotoh, M., Martin, P., & Hassanein, H. S. (2012). Ubiquitous Health Monitoring Using Mobile Web Services. *Procedia Computer Science*, 10, 332–339. <https://doi.org/10.1016/J.PROCS.2012.06.044>
- Elhanan, G., Perl, Y., & Geller, J. (2011). A survey of SNOMED CT direct users, 2010: impressions and preferences regarding content and quality. *Journal of the American Medical Informatics Association*, 18(Supplement 1), i36–i44. <https://doi.org/10.1136/amiajnl-2011-000341>
- Elkin, P., & Tuttle, M. (2012). History of Terminology and Terminological Logics. In Peter Elkin (Ed.), *Terminology and Terminological Systems* (pp. 5–10). London: Springer-Verlag.
- Evans, D. A., Cimino, J. J., Hersh, W. R., Huff, S. M., & Bell, D. S. (1994). Toward a medical-concept representation language. The Canon Group. *Journal of the American Medical Informatics Association : JAMIA*, 1(3), 207–217.
- Eysenbach, G. (2001). What is e-health? *Journal of Medical Internet Research*, 3(2), 1–5. <https://doi.org/10.2196/jmir.3.2.e20>
- Faber Benítez, P. (2009). The cognitive shift in terminology and specialized translation. *MonTi: Monografías de Traducción e Interpretación*, 1, 107–134. <https://doi.org/10.6035/MonTI.2009.1.5>
- Faber, P. (2002). Terminographic definition and concept representation. In Belinda Maia, Johann Haller, & Margherita Ulyrich (Eds.), *Training the Language Services Provider for the New Millennium* (pp. 343–354). Porto: Universidade do Porto. <https://doi.org/10.1075/term.12.2.03fab>

- Faber, P. (Ed.). (2012). *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin, Boston: De Gruyter. <https://doi.org/10.1515/9783110277203>
- Faber, P. (2015). Frames as a framework for terminology. In Hendrik J. Kockaert & Frieda Steurs (Eds.), *Handbook of Terminology: Volume 1* (pp. 14–33). Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/hot.1.fra1>
- Faber, P., León-Araúz, P., & Reimerink, A. (2014). Representing Environmental Knowledge in EcoLexicon. In E. Bárcena, T. Read, & J. Arús (Eds.), *Languages for Specific Purposes in the Digital Era* (pp. 267–301). Springer, Cham. https://doi.org/10.1007/978-3-319-02222-2_13
- Faber, P., León-Araúz, P., & Reimerink, A. (2016). EcoLexicon: New Features and Challenges. In L. Kernerman, I. Kosem Trojina, I. Krek, S. & Trap-Jensen (Ed.), *GLOBALEX 2016: Lexicographic Resources for Human Language Technology in conjunction with the 10th edition of the Language Resources and Evaluation Conference* (pp. 73–80). Portorož.
- Facchin, F., Barbara, G., Saita, E., Mosconi, P., Roberto, A., Fedele, L., & Vercellini, P. (2015). Impact of endometriosis on quality of life and mental health: pelvic pain makes the difference. *Journal of Psychosomatic Obstetrics & Gynecology*, 36(4), 135–141. <https://doi.org/10.3109/0167482X.2015.1074173>
- Falcone, T., & Setúbal, A. (2013). JMIG video articles. *Journal of Minimally Invasive Gynecology*, 20(3), 267. <https://doi.org/10.1016/j.jmig.2013.04.002>
- Fassbender, A., Rahmioglu, N., Vitonis, A. F., Viganò, P., Giudice, L. C., D'Hooghe, T. M., ... Zondervan, K. T. (2014). World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonisation Project: IV. Tissue collection, processing, and storage in endometriosis research. *Fertility and Sterility*, 102(5), 1244–1253. <https://doi.org/10.1016/j.fertnstert.2014.07.1209>
- Felber, H. (1984). *Terminology Manual*. (U. (Program) Unesco. General Information Programme, Ed.). Unesco.
- Felber, H., & Budin, G. (1989). *Terminologie in Theorie und Praxis*. Tübingen: Gunter Narr Verlag. R
- Fiormonte, D., & Priego, E. (2016). Knowledge Monopolies and Global Academic Publishing. *The Winnower*. <https://doi.org/10.15200/WINN.147220.00404>
- Flores, M., Glusman, G., Brogaard, K., Price, N. D., & Hood, L. (2013). P4 medicine: how systems medicine will transform the healthcare sector and society. *Personalized Medicine*, 10(6), 565–576. <https://doi.org/10.2217/pme.13.57>
- Frankl, O. (1925). Adenomyosis uteri. *American Journal of Obstetrics and Gynecology*, 10(5), 680–684. [https://doi.org/10.1016/S0002-9378\(25\)90632-1](https://doi.org/10.1016/S0002-9378(25)90632-1)
- Gante, I., Medeiros-Borges, C., & Águas, F. (2017). Hysterectomies in Portugal (2000–2014): What has changed? *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 208, 97–102. <https://doi.org/10.1016/j.ejogrb.2016.11.021>
- Gao, X., Outley, J., Botteman, M., Spalding, J., Simon, J. A., & Pashos, C. L. (2006). Economic burden of endometriosis. *Fertility and Sterility*, 86(6), 1561–1572. <https://doi.org/10.1016/j.fertnstert.2006.06.015>
- García-Velasco, J. A., & Rizk, B. (2010). *Endometriosis : current management and future trends*. Jaypee Brothers Medical Publishers.
- García de Quesada, M. (2001). Estructura definicional terminográfica en el subdominio de la oncología clínica. *Estudios de Lingüística Española (ELiES)*, 14.
- Gaudin, F. (2005). La socioterminologie. *Langages*, 39(157), 80–92. <https://doi.org/10.3406/lgge.2005.976>

- Gaudin, F. (2007). *Socioterminologie : une approche sociolinguistique de la terminologie*, , Bruxelles, 2003. Bruxelles: Duculot.
- Gene Ontology. (2018). Gene Ontology Consortium. Retrieved April 11, 2018, from <http://www.geneontology.org/>
- Gill, I. S., Advincula, A. P., Aron, M., Cadeddu, J., Canes, D., Curcillo, P. G., ... Teixeira, J. (2010). Consensus statement of the consortium for laparoendoscopic single-site surgery. *Surgical Endoscopy*, 24(4), 762–768. <https://doi.org/10.1007/s00464-009-0688-8>
- Gilmour, J. A., Huntington, A., & Wilson, H. V. (2008). The impact of endometriosis on work and social participation. *International Journal of Nursing Practice*, 14(6), 443–448. <https://doi.org/10.1111/j.1440-172X.2008.00718.x>
- Giudice, L. C. (2010). Endometriosis. *New England Journal of Medicine*, 362(25), 2389–2398. <https://doi.org/10.1056/NEJMc1000274>
- Giudice, L., Evers, J. L. H., & Healy, D. L. (David L. (2012). *Endometriosis : Science and Practice*. Wiley-Blackwell. Retrieved from https://books.google.pt/books/about/Endometriosis.html?id=kxzaBqmgIrMC&redir_esc=y
- Golbreich, C., Grosjean, J., & Darmoni, S. J. (2013). The Foundational Model of Anatomy in OWL 2 and its use. *Artificial Intelligence in Medicine*, 57(2), 119–132. <https://doi.org/10.1016/j.artmed.2012.11.002>
- Grabar, N., Hamon, T., & Bodenreider, O. (2012). Ontologies and terminologies: Continuum or dichotomy? *Applied Ontology*, 7(4), 375–386. <https://doi.org/10.3233/AO-2012-0119>
- Grabar, N., & Zweigenbaum, P. (2004). Lexically-based terminology structuring. *Terminology*, 10(1), 23–53. <https://doi.org/10.1075/term.10.1.03gra>
- Graunt, J. (1977). Natural and Political Observations Mentioned in a Following Index, and Made Upon the Bills of Mortality (pp. 11–20). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-81046-6_2
- Gruber, T. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2), 199–220.
- Gruber, T. R. (1993). Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal Human-Computer Studies*, 43(5–6), 907–928.
- Guarino, N. (1998). Formal Ontology and Information Systems. In *Proceedings of FOIS '98* (pp. 3–15). Trento: IOS Press.
- Guarino, N., & Giarretta, P. (1995). Ontologies and knowledge bases: towards a terminological clarification. In N. Mars (Ed.), *Towards Very Large Knowledge Bases* (pp. 25–32). Amsterdam: IOS Press.
- Guarino, N., Oberle, D., & Staab, S. (2009). What Is an Ontology? In *Handbook on Ontologies* (pp. 1–17). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Haas, D., Shebl, O., Shamiyeh, A., & Oppelt, P. (2013). The rASRM score and the Enzian classification for endometriosis: their strengths and weaknesses. *Acta Obstetrica et Gynecologica Scandinavica*, 92(1), 3–7. <https://doi.org/10.1111/aogs.12026>
- Hadfield, R., Mardon, H., Barlow, D., & Kennedy, S. (1996). Delay in the diagnosis of endometriosis: a survey of women from the USA and the UK. *Human Reproduction*, 11(4), 878–880.
- Heath, T., & Bizer, C. (2011). *Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web: Theory and Technology* (Vol. 1). Morgan & Claypool.

- Hippocrates of Cos. (1988). Affections. https://doi.org/10.4159/DLCL.hippocrates_cos-affections.1988
- Hood, L., & Flores, M. (2012). A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory. *New Biotechnology*, 29(6), 613–624. <https://doi.org/10.1016/j.nbt.2012.03.004>
- Hudelist, G., Fritzer, N., Thomas, A., Niehues, C., Oppelt, P., Haas, D., ... De Aveiro User, U. (2012). Diagnostic delay for endometriosis in Austria and Germany: causes and possible consequences. *Human Reproduction*, 27(12), 3412–3416. <https://doi.org/10.1093/humrep/des316>
- Humphreys, B. L., & Lindberg, D. A. (1993). The UMLS project: making the conceptual connection between users and the information they need. *Bulletin of the Medical Library Association*, 81(2), 170–177. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8472002>
- Husby, G. K., Haugen, R. S., & Moen, M. H. (2003). Diagnostic delay in women with pain and endometriosis. *Acta Obstetrica et Gynecologica Scandinavica*, 82(7), 649–653.
- IEEE. (1990). IEEE Standard Glossary of Software Engineering Terminology. *IEEE Std 610.12-1990*. IEEE. <https://doi.org/doi:10.1109/IEEESTD.1990.101064>
- IHTSDO. (2018a). Compositional Grammar - Specification and Guide. International Health Terminology Standards Development Organisation. Retrieved from <https://confluence.ihtsdotools.org/display/DOCSCG>
- IHTSDO. (2018b). Diagramming Guideline. International Health Terminology Standards Development Organisation. Retrieved from <https://confluence.ihtsdotools.org/display/DOCDIAG/Diagramming+Guideline>
- IHTSDO. (2018c). SNOMED CT Editorial Guide. International Health Terminology Standards Development Organisation. Retrieved from <https://confluence.ihtsdotools.org/display/DOCEG/SNOMED+CT+Editorial+Guide>
- IHTSDO. (2018d). SNOMED CT Starter Guide. International Health Terminology Standards Development Organisation. Retrieved from <https://confluence.ihtsdotools.org/display/DOCSTART/SNOMED+CT+Starter+Guide>
- ISO. (2000). ISO 1087-1:2000 - Terminology work - Vocabulary - Part 1: Theory and application. Geneva: International Standardization Organization.
- ISO. (2007). ISO 17115: Health informatics -- Vocabulary for terminological systems. Geneva: International Organization for Standardization.
- ISO. (2009). ISO 704:2009 - Terminology work - Principles and methods. Geneva: International Organization for Standardization.
- ISO. (2011). ISO 25964-1: 2011: Information and documentation - Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval. Geneva: International Organization for Standardization.
- ISO. (2012). ISO 1828:2012 - Health informatics - Categorical structure for terminological systems of surgical procedures. Geneva: International Organization for Standardization.
- Johnson, N. P., & Hummelshoj, L. (2013). Consensus on current management of endometriosis. *Human Reproduction*, 28(25), 1552–1568. <https://doi.org/10.1093/humrep/det050>
- Johnson, N. P., Hummelshoj, L., Adamson, G. D., Keckstein, J., Taylor, H. S., Abrao, M. S., ... Giudice, L. C. (2017). World Endometriosis Society consensus on the classification of endometriosis. *Human Reproduction*, 32(2), 315–324. <https://doi.org/10.1093/humrep/dew293>
- Johnston, I. (2006). *Galen: On Diseases and Symptoms*. Cambridge: Cambridge University Press.

- Kageura, K. (1997). Multifaceted/Multidimensional Concept Systems. In G. Budin & S. E. Wright (Eds.), *Handbook of Terminology Management: Volume 1: Basic Aspects of Terminology Management* (pp. 119–132). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Kennedy, S., Bergqvist, A., Chapron, C., D’Hooghe, T., Dunselman, G., Greb, R., ... Saridogan, E. (2005). ESHRE guideline for the diagnosis and treatment of endometriosis. *Human Reproduction*, 20(10), 2698–2704.
- Kent, W. (2012). *Data and Reality: A Timeless Perspective on Perceiving and Managing Information in our Imprecise World* (3rd edition). Westfield: Technics Publications.
- Kerremans, K., De Baer, P., & Temmerman, R. (2010). Competency-based job descriptions and termontography: the case of terminological variation. In M. Thelen & F. Steurs (Eds.), *Terminology in Everyday Life* (pp. 181–193). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Kerremans, K., & Temmerman, R. (2004). Towards multilingual, termontological support in ontology engineering. In *Proceedings of Termino 2004. Université de Lyon* (p. n.a.). Lyon.
- Kerremans, K., Temmerman, R., & Tummers, J. (2004). Discussion on the Requirements for a Workbench supporting Termontography. In *Proceedings of the XI Euralex International Congress* (pp. 6–10).
- Khazali, S. (2016). Endometriosis Classification-The Quest for the Holy Grail? *Journal of Reproduction & Infertility*, 17(2), 67.
- Kibbe, W. A., Arze, C., Felix, V., Mitraka, E., Bolton, E., Fu, G., ... Schriml, L. M. (2015). Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Research*, 43(D1), D1071–D1078. <https://doi.org/10.1093/nar/gku1011>
- Kim, K. M., Park, K. S., Lee, H. J., Lee, Y. H., Bae, J. S., Lee, Y. J., ... Lim, S. (2016). Efficacy of a New Medical Information system, Ubiquitous Healthcare Service with Voice Inception Technique in Elderly Diabetic Patients. *Scientific Reports*, 5(1), 18214. <https://doi.org/10.1038/srep18214>
- Kleisiaris, C. F., Sfakianakis, C., & Papathanasiou, I. V. (2014). Health care practices in ancient Greece: The Hippocratic ideal. *Journal of Medical Ethics and History of Medicine*, 7, 6.
- Knapp, V. J. (1999). How old is endometriosis? Late 17th- and 18th-century European descriptions of the disease. *Fertility and Sterility*, 72(1), 10–14. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10428141>
- Koltermann, K. C., Dornquast, C., Ebert, A. D., & Reinhold, T. (2017). Economic Burden of Endometriosis: A Systematic Review. *Annals of Reproductive Medicine and Treatment*, 2(2), 1015. Retrieved from <https://pdfs.semanticscholar.org/e5a9/e0134a372beb0620e742b7ec12f85751dc2b.pdf>
- Koninckx, P. R., Ussia, A., Adamyan, L., & Wattiez, A. (2011). An endometriosis classification, designed to be validated. *Gynecological Surgery*, 8(1), 1–6. <https://doi.org/10.1007/s10397-010-0626-8>
- Kraft, D. (2017). Quantified self to quantified health: how tech helps doctors fill gaps in patient records. *Wired Magazine*. Retrieved from <http://www.wired.co.uk/article/hospital-prescribing-tech?>
- L’Homme, M.-C. (2004). *La terminologie : principes et techniques*. Montréal: Presses de l’Université de Montréal.
- L’Homme, M.-C., & Marshman, E. (2006). Terminological Relationships and Corpus-based Methods for Discovering them: An Assessment for Terminographers. In L. Bowker (Ed.), *Lexicography, Terminology, and Translation. Text-based studies in honour of Ingrid Meyer* (pp. 67–80). Ottawa: University of Ottawa Press.

- Levy, A. R., Osenenko, K. M., Lozano-Ortega, G., Sambrook, R., Jeddi, M., Bélisle, S., & Reid, R. L. (2011). Economic Burden of Surgically Confirmed Endometriosis in Canada. *Journal of Obstetrics and Gynaecology Canada*, 33(8), 830–837. [https://doi.org/10.1016/S1701-2163\(16\)34986-6](https://doi.org/10.1016/S1701-2163(16)34986-6)
- Levy, G., Dehaene, A., Laurent, N., Lernout, M., Collinet, P., Lucot, J.-P., ... Poncelet, E. (2013). An update on adenomyosis. *Diagnostic and Interventional Imaging*, 94(1), 3–25. <https://doi.org/10.1016/j.diii.2012.10.012>
- Leyendecker, G. (2000). Redefining endometriosis: endometriosis is an entity with extreme pleiomorphism. *Human Reproduction (Oxford, England)*, 15(1), 4–7. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10611178>
- Liaropoulos, L., & Goranitis, I. (2015). Health care financing and the sustainability of health systems. *International Journal for Equity in Health*, 14, 80. <https://doi.org/10.1186/s12939-015-0208-5>
- Liliana, M., Alessandro, P., Giada, C., & Luca, M. (2011). Single-port access laparoscopic hysterectomy: a new dimension of minimally invasive surgery. *Journal of Gynecological Endoscopy and Surgery*, 2(1), 11–17. <https://doi.org/10.4103/0974-1216.85273>
- Lilienfeld, D. (2007). Celebration: William Farr (1807-1883) an appreciation on the 200th anniversary of his birth. *International Journal of Epidemiology*, 36(5), 985–987. <https://doi.org/10.1093/ije/dym132>
- Lindberg, D. A., Humphreys, B. L., & McCray, A. T. (1993). The Unified Medical Language System. *Methods of Information in Medicine*, 32(4), 281–291. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8412823>
- Linnaeus, C. (1763). *Genera Morborum*. Upsalla: The Linnean Collections. Retrieved from <http://linnean-online.org/120052/>
- Lipscomb, C. E. (2000). Medical Subject Headings (MeSH). *Bulletin of the Medical Library Association*, 88(3), 265–266. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10928714>
- Löckinger, G., Kockaert, H. J., & Budin, G. (2015). Intensional definitions. In Hendrik J. Kockaert & Frida Steurs (Eds.), *Handbook of Terminology: Volume 1* (pp. 60–81). Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/hot.1.int1>
- Lorig, K. (2014). Chronic disease self-management program: insights from the eye of the storm. *Frontiers in Public Health*, 2, 253. <https://doi.org/10.3389/fpubh.2014.00253>
- Maciocia, G. (2015). *The foundations of Chinese medicine : a comprehensive text* (Third edit). Churchill Livingstone.
- Maks, I. (2006). Frame-based definitions in a Learners' Dictionary for Dutch Business Language. In P. Ten Hacken (Ed.), *Terminology, Computing and Translation* (pp. 191–206). Tübingen: Gunter Narr Verlag.
- Malinak, L. R., Buttram, V. C., Elias, S., & Simpson, J. L. (1980). Heritage aspects of endometriosis. II. Clinical characteristics of familial endometriosis. *American Journal of Obstetrics and Gynecology*, 137(3), 332–337. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7377253>
- Mann, W. (2017). Radical hysterectomy. Retrieved May 13, 2018, from <https://www.uptodate.com/contents/radical-hysterectomy>
- Martin, D. C., Redwine, D. B., Reich, H., Foreword, A. J. K., & Rock, J. A. (2017). *Laparoscopic Appearance of Endometriosis - Color Atlas* (2nd edition). Memphis: The Resurge Press. Retrieved from <http://www.danmartinmd.com/files/coloratlas1990.pdf>

- Martin, W. (1998). Frames as definition models for terms. In A. Munteanu (Ed.), *Proceedings of the International Conference on Professional Communication and Knowledge Transfer*, (pp. 189–221). Vienna: Termnet.
- Maruyama, T. (2014). Endometrial stem/progenitor cells. *Journal of Obstetrics and Gynaecology Research*, 40(9), 2015–2022. <https://doi.org/10.1111/jog.12501>
- Mccrae, J. P., Bosque-Gil, J., Gracia, J., Buitelaar, P., & Cimiano, P. (2017). The OntoLex-Lemon Model: development and applications. In *Electronic Lexicography in the 21st Century: Lexicography from Scratch. Proceedings from the eLex 2017 Conference, 19-21 September 2017*. Leiden, The Netherlands.
- McCray, A. (2003). An Upper Level Ontology for the Biomedical Domain. *Comp Funct Genom*, 4, 80–84.
- McCray, A. T., Burgun, A., & Bodenreider, O. (2001). Aggregating UMLS semantic types for reducing conceptual complexity. *Studies in Health Technology and Informatics*, 84(Pt 1), 216–220. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11604736>
- McEnery, T., & Wilson, A. (2001). *Corpus linguistics : an introduction*. Edinburgh: Edinburgh University Press.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies : an advanced resource book*. London: Routledge.
- Mea, V. Della, Vuattolo, O., Ustun, B. T., Rector, A., Schulz, S., Rodrigues, J.-M., ... Persson, K. B. (2014). A Web-Based Tool for Development of a Common Ontology between ICD11 and SNOMED-CT. In *2014 IEEE International Conference on Healthcare Informatics* (pp. 144–148). IEEE. <https://doi.org/10.1109/ICHI.2014.26>
- Meara, J. G., Leather, A. J. M., Hagander, L., Alkire, B. C., Alonso, N., Ameh, E. A., ... Yip, W. (2015). Global Surgery 2030: evidence and solutions for achieving health, welfare, and economic development. *The Lancet*, 386(9993), 569–624. [https://doi.org/10.1016/S0140-6736\(15\)60160-X](https://doi.org/10.1016/S0140-6736(15)60160-X)
- Měchura, M. B. (2017). Introducing Lexonomy: an open-source dictionary writing and publishing system. In *Electronic Lexicography in the 21st Century: Lexicography from Scratch. Proceedings of the eLex 2017 conference, 19-21 September 2017*. Leiden, The Netherlands.
- Menvielle, L., Audrain, A.-F., & Menvielle, W. (Eds.). (2017). *The digitization of healthcare : new challenges and opportunities*. London: Palgrave Macmillan.
- Meuleman, C., Vandenabeele, B., Fieuws, S., Spiessens, C., Timmerman, D., & D’Hooghe, T. (2009). High prevalence of endometriosis in infertile women with normal ovulation and normospermic partners. *Fertility and Sterility*, 92(1), 68–74. <https://doi.org/10.1016/j.fertnstert.2008.04.056>
- Meyer, I. (2001). Extracting knowledge-rich contexts for terminography: a conceptual and methodological framework. In Didier Bourigault, Christian Jacquemin, & Marie-Claude L’Homme (Eds.), *Recent Advances in Computational Terminology* (pp. 279–302). Amsterdam/New York: John Benjamins Publishing Company. <https://doi.org/10.1075/nlp.2.15mey>
- Meyer, I., Eck, K., & Skuce, D. (1997). 1.4.2 Systematic Concept Analysis within a Knowledge-Based Approach to Terminology. In Sue Ellen Wright & Gerhard Budin (Eds.), *Handbook of Terminology Management - Vol. 1: Basic Aspects of Terminology Management* (pp. 98–118). Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/z.htm1.14mey>
- Meyer, I., & Mackintosh, K. (1996). The Corpus from a Terminographer’s Viewpoint. *International Journal of Corpus Linguistics*, 1(2), 257–285. <https://doi.org/10.1075/ijcl.1.2.05mey>

- Meyer, I., Skuce, D., Bowker, L., & Eck, K. (1992). Towards a new generation of terminological resources: an experiment in building a terminological knowledge base. In *Proceedings of the 14th conference on Computational linguistics (COLING)* (Vol. 3, pp. 956–960). Association for Computational Linguistics. <https://doi.org/10.3115/993079.993106>
- Millar, J. (2016). The Need for a Global Language - SNOMED CT Introduction. *Studies in Health Technology and Informatics*, 225, 683–685.
- Minsky, M. (1977). Frame theory. In P.N. Johnson-Laird & P.C. Wason (Eds.), *Thinking: Readings in Cognitive Science* (pp. 355–376). Cambridge: Cambridge University Press. Retrieved from <https://biolawgy.wordpress.com/2010/02/11/minsky-m-1977-frame-theory-in-p-n-johnson-laird-and-p-c-wason-thinking-reasings-in-cognitive-science-cambridge-cambridge-university-press-pp-355-376/>
- Moradi, M., Parker, M., Sneddon, A., Lopez, V., & Ellwood, D. (2014). Impact of endometriosis on women's lives: a qualitative study. *BMC Women's Health*, 14(1), 123. <https://doi.org/10.1186/1472-6874-14-123>
- Nap, A. (2012). Theories on the Pathogenesis of Endometriosis. In D. H. Linda Giudice, Johannes Evers (Ed.), *Endometriosis: Science and Practice* (pp. 42–53). Oxford: Wiley-Blackwell. <https://doi.org/10.1002/9781444398519.ch5>
- Nelson, S. J., Johnston, W. D., & Humphreys, B. L. (2001). Relationships in Medical Subject Headings (MeSH). In C. A. Bean & R. Green (Eds.), *Relationships in the Organization of Knowledge* (pp. 171–184). Springer, Dordrecht. https://doi.org/10.1007/978-94-015-9696-1_11
- Nezhat, C., Nezhat, F., & Nezhat, C. (2012). Endometriosis: ancient disease, ancient treatments. *Fertility and Sterility*, 98(6), S1–S62. <https://doi.org/10.1016/j.fertnstert.2012.08.001>
- NICE. (2017). Endometriosis: diagnosis and management - NICE Guideline [NG73]. National Institute for Health and Care Excellence.
- Nisolle, M. (2002). Ovarian endometriosis and peritoneal endometriosis: are they different entities from a fertility perspective? *Current Opinion in Obstetrics & Gynecology*, 14(3), 283–288. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12032383>
- Nisolle, M., & Donnez, J. (1997). Peritoneal endometriosis, ovarian endometriosis, and adenomyotic nodules of the rectovaginal septum are three different entities. *Fertility and Sterility*, 68(4), 585–596. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9341595>
- NLM. (2018). Publication Characteristics (Publication Types) with Scope Notes. Retrieved April 2, 2018, from <https://www.nlm.nih.gov/mesh/pubtypes.html>
- NML. (2009). *UMLS Reference Manual*. Bethesda: National Library of Medicine. <https://doi.org/10.1093/nar/gku1011>
- NML. (2016). The Unified Medical Language System (UMLS). Retrieved May 18, 2017, from https://www.nlm.nih.gov/research/umls/new_users/online_learning/OVR_001.html
- Nnoaham, K. E., Hummelshoj, L., Webster, P., d'Hooghe, T., de Cicco Nardone, F., de Cicco Nardone, C., ... World Endometriosis Research Foundation Global Study of Women's Health consortium. (2011). Impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries. *Fertility and Sterility*, 96(2), 366–373.e8. <https://doi.org/10.1016/j.fertnstert.2011.05.090>
- Nordenfelt, L. (2007). The concepts of health and illness revisited. *Medicine, Health Care and Philosophy*, 10(1), 5–10. <https://doi.org/10.1007/s11019-006-9017-3>

- Norman, C. D., & Skinner, H. A. (2006). eHealth Literacy: Essential Skills for Consumer Health in a Networked World. *Journal of Medical Internet Research*, 8(2), e9. <https://doi.org/10.2196/jmir.8.2.e9>
- Noy, N. F., Mejino, J. L., Rosse, C., & Musen, M. A. (2018). Foundational Model of Anatomy. Retrieved February 21, 2018, from <http://sig.biostr.washington.edu/projects/fm/>
- Nuopponen, A. (2011). Methods of concept analysis – tools for systematic concept analysis. *LSP Journal*, 2(1), 4–15.
- Nuopponen, A. (2014). Tangled Web of Concept Relations. Concept relations for ISO 1087-1 and ISO 704. In *Proceedings of Terminology and Knowledge Engineering 2014* (p. 10 p.). Berlin.
- Nutbeam, D. (2000). Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. *Health Promotion International*, 15(3), 259–267. <https://doi.org/10.1093/heapro/15.3.259>
- Ogden, C. K., & Richards, I. A. (1923). *The Meaning Of Meaning: A Study of the influence of Language upon Thought and of the Science of Symbolism*. London: Routledge & Kegan Paul Ltd.
- Oh, H., Rizo, C., Enkin, M., & Jadad, A. (2005). What Is eHealth (3): A Systematic Review of Published Definitions. *Journal of Medical Internet Research*, 7(1), e1.
- Pagliari, C., Sloan, D., Gregor, P., Sullivan, F., Detmer, D., Kahan, J. P., ... MacGillivray, S. (2005). What is eHealth (4): A Scoping Exercise to Map the Field. *Journal of Medical Internet Research*, 7(1), e9.
- Patwardhan, B., Warude, D., Pushpangadan, P., & Bhatt, N. (2005). Ayurveda and traditional Chinese medicine: a comparative overview. *Evidence-Based Complementary and Alternative Medicine : ECAM*, 2(4), 465–473. <https://doi.org/10.1093/ecam/neh140>
- Pazienza, M. T., Pennacchiotti, M., & Zanzotto, F. M. (2005). Terminology Extraction: An Analysis of Linguistic and Statistical Approaches. In *Knowledge Mining* (pp. 255–279). Berlin/Heidelberg: Springer-Verlag. https://doi.org/10.1007/3-540-32394-5_20
- Pearson, J. (1998). *Terms in context*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Picht, H., & Draskau, J. (1985). *Terminology: an introduction*. Surrey: University of Surrey, Department of Linguistic and International Studies.
- Practice Committee of the American Society for Reproductive Medicine. (2012). Endometriosis and infertility: a committee opinion. *Fertility and Sterility*, 98(3), 591–598. <https://doi.org/10.1016/j.fertnstert.2012.05.031>
- Prast, J., Oppelt, P., Shamiyeh, A., Shebl, O., Brandes, I., & Haas, D. (2013). Costs of endometriosis in Austria: a survey of direct and indirect costs. *Archives of Gynecology and Obstetrics*, 288(3), 569–576. <https://doi.org/10.1007/s00404-013-2793-0>
- Prieto-Velasco, J. A., & López Rodríguez, C. I. (2009). Managing graphic information in terminological knowledge bases. *Terminology*, 15(2), 179–213. <https://doi.org/10.1075/term.15.2.02pri>
- Prieto Velasco, J. A. (2016). Depicting specialized concepts: strategies for the visualization of terminological knowledge. In Susanne Lervad, Peder Flemestad, & Lotte Weilgaard Christensen (Eds.), *Verbal and nonverbal representation in Terminology: Proceedings of the TOTh 2013 Workshop* (pp. 37–50). Copenhagen: DNRf's Centre for Textile Research; Institut Porphyre, Savoir et Connaissance. Retrieved from <http://ctr.hum.ku.dk/conferences/2013/abstracts.pdf>
- Prieto Velasco, J. A., & Tercedor Sánchez, M. (2014). The embodied nature of medical concepts: image schemas and language for pain. *Cognitive Processing*, 15(3), 283–296. <https://doi.org/10.1007/s10339-013-0594-9>

- Przywara, B. (2010). *Projecting future health care expenditure at European level: drivers, methodology and main results*. Brussels. <https://doi.org/10.2765/42844>
- Quazi, H. (2013). *Presenteeism : the invisible cost to organizations*. London: Palgrave Macmillan.
- Rahmioglu, N., Fassbender, A., Vitonis, A. F., Tworoger, S. S., Hummelshoj, L., D'Hooghe, T. M., ... Zondervan, K. T. (2014). World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonization Project: III. Fluid biospecimen collection, processing, and storage in endometriosis research. *Fertility and Sterility*, 102(5), 1233–1243. <https://doi.org/10.1016/j.fertnstert.2014.07.1208>
- Rao, G., Mansard, M., Ravula, P., Rebala, P., Dama, R., & Reddy, D. (2009). Single-port surgery: Current applications and limitations. *Asian Journal of Endoscopic Surgery*, 2(3), 56–64. <https://doi.org/10.1111/j.1758-5910.2009.00023.x>
- Rao, P. P., Rao, P. P., & Bhagwat, S. (2011). Single-incision laparoscopic surgery - current status and controversies. *Journal of Minimal Access Surgery*, 7(1), 6–16.
- Rector, A. (1999). Clinical terminology: Why is it so hard? *Methods of Information in Medicine*, 38, 239–252.
- Rector, A., Bechhofer, S., Goble, C., Horrocks, I., Nowlan, W. A., & Solomon, W. D. (1997). The GRAIL concept modelling language for medical terminology. *Artificial Intelligence in Medicine*, 9(2), 139–171. Retrieved from https://scholar.google.co.uk/citations?user=qolPoFgAAAAJ&hl=en#d=gs_md_cita-d&p=&u=%2Fcitations%3Fview_op%3Dview_citation%26hl%3Den%26user%3DqolPoFgAAAAJ%26citation_for_view%3DqolPoFgAAAAJ%3Ad1gkVwhDpl0C%26tzm%3D-60
- Rector, A., Gangemi, A., Galeazzi, E., Glowinski, A., & Rossi-Mori, A. (1994). The GALEN CORE model schemata for anatomy: Towards a re-usable application-independent model of medical concepts. *Proceedings MIE*, 94, 229.
- Rector, A., Nowlan, W. A., & Consortium, G. (1994). The GALEN project. *Computer Methods and Programs in Biomedicine*, 45(1–2), 75–78.
- Rector, A., Nowlan, W. A., & Glowinski, A. (1993). Goals for concept representation in the GALEN project. *Proceedings of the Annual Symposium on Computer Application in Medical Care*. American Medical Informatics Association.
- Rector, A., & Rogers, J. (2006). Ontological and Practical Issues in Using a Description Logic to Represent Medical Concept Systems: Experience from GALEN. In *Reasoning Web International Summer School* (pp. 197–231). Berlin, Heidelberg: Springer. https://doi.org/10.1007/11837787_9
- Redwine, D. B. (2012). *Googling Endometriosis: The Lost Centuries*. Createspace Independent Publishing Platform. Retrieved from https://books.google.pt/books/about/Googling_Endometriosis.html?id=8NR7MAEACAAJ&redir_esc=y
- Rey, A. (1995). *Essays on Terminology*. Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/btl.9>
- Ridley, J. (1968). The Histogenesis of Endometriosis: A Review of Facts and Fancies. *Obstetrical & Gynecological Survey*, 23(1), 1–35. Retrieved from https://journals.lww.com/obgynsurvey/citation/1968/01000/the_histogenesis_of_endometriosis_a_review_of.1.aspx
- Rier, S., & Foster, W. G. (2002). Environmental Dioxins and Endometriosis. *Toxicological Sciences*, 70(2), 161–170. <https://doi.org/10.1093/toxsci/70.2.161>

- Roche, C. (2007). Saying Is Not Modelling. In *Proceedings of NLPCS 2007 (Natural Language Processing and Cognitive Science)*, Funchal, June 2007 (pp. 47–56). Funchal.
- Roche, C. (2012). Ontoterminology: How to unify terminology and ontology into a single paradigm. In N. Calzolari, K. Choukri, T. Declerck, M. U. Dogan, B. Maegaard, J. Mariani, ... S. Piperidis (Eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, Istanbul, Turkey, May 23-25, 2012 (pp. 2626–2630). Istanbul: European Language Resources Association (ELRA).
- Roche, C. (2015). Ontological definition. In H. J. Kockaert & F. Steurs (Eds.), *Handbook of Terminology - vol. 1* (pp. 128–152). Amsterdam: John Benjamins Publishing Company.
- Roche, C. (2017). TeDI - Ontoterminology Editor. Retrieved April 21, 2018, from <http://ontoterminology.com/tedi>
- Roche, C., Calberg-Challot, M., Damas, L., & Rouard, P. (2009). Ontoterminology: A new paradigm for terminology. In *International Conference on Knowledge Engineering and Ontology Development, Oct 2009* (pp. 321–326). Funchal.
- Rodrigues, J.-M., Kim, S., Aljunid, S., Lee, J. J., Ten Napel, H., & Trombert, B. (2017). From a Categorical Structure to a Concept Model: The International Classification of Health Interventions (ICHI). *Studies in Health Technology and Informatics*, 245, 1334.
- Rodrigues, J.-M., Robinson, D., Della Mea, V., Campbell, J., Rector, A., Schulz, S., ... Brand Persson, K. (2015). Semantic Alignment between ICD-11 and SNOMED CT. *Studies in Health Technology and Informatics*, 216, 790–794. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/26262160>
- Rodrigues, J.-M., Schulz, S., Rector, A., Spackman, K., Millar, J., Campbell, J., ... Persson, K. B. (2014). ICD-11 and SNOMED CT Common Ontology: circulatory system. *Studies in Health Technology and Informatics*, 205, 1043–1047. <https://doi.org/10.3233/978-1-61499-432-9-1043>
- Rodrigues, J.-M., Schulz, S., Rector, A., Spackman, K., Üstün, B., Chute, C. G., ... Persson, K. B. (2013). Sharing ontology between ICD 11 and SNOMED CT will enable seamless re-use and semantic interoperability. *Studies in Health Technology and Informatics*, 192, 343–346. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/23920573>
- Rodrigues, J. M., Casey, A., Bousquet, C., Kumar, A., Lewalle, P., & Trombert Paviot, B. (2011). Comparing existing national and international classification systems of surgical procedures with the CEN/ISO 1828 ontology framework standard. *Studies in Health Technology and Informatics*, 169, 844–848.
- Rodrigues, J. M., Kim, S., Trombert Paviot, B., Lee, J. J., & Aljunid, S. (2017). How to Link SNOMED CT Procedure and WHO International Classification of Health Interventions (ICHI). *Studies in Health Technology and Informatics*, 236, 40–47.
- Rogers, J., Roberts, A., Solomon, D., van der Haring, E., Wroe, C., Zanstra, P., & Rector, A. (2001). GALEN ten years on: tasks and supporting tools. *Studies in Health Technology and Informatics*, 84(Pt 1), 256–260. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11604744>
- Rogers, M., & Wright, S. E. (2006). Approaches to Terminological Theories: The Anglo-Saxon Approach. In H. Picht (Ed.), *Modern Approaches to Terminological Theories and Applications* (pp. 107–134). Bern: Peter Lang.
- Rogers, P. A. W., Adamson, G. D., Al-Jefout, M., Becker, C. M., D’Hooghe, T. M., Dunselman, G. A. J., ... Zondervan, K. T. (2017). Research Priorities for Endometriosis. *Reproductive Sciences*, 24(2), 202–226. <https://doi.org/10.1177/1933719116654991>
- Rogers, P. A. W., D’Hooghe, T. M., Fazleabas, A., Gargett, C. E., Giudice, L. C., Montgomery, G. W., ... Zondervan, K. T. (2009). Priorities for Endometriosis Research: Recommendations From an International Consensus Workshop. *Reproductive Sciences*, 16(4), 335–346.

- Rose, J., Weiser, T. G., Hider, P., Wilson, L., Gruen, R. L., & Bickler, S. W. (2015). Estimated need for surgery worldwide based on prevalence of diseases: a modelling strategy for the WHO Global Health Estimate. *The Lancet Global Health*, 3, S13–S20.
- Rosse, C., & Mejino, J. L. V. (2003). A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *Journal of Biomedical Informatics*, 36(6), 478–500.
- Sager, J. (1990). *A practical course in terminology processing*. Amsterdam/New York: John Benjamins Publishing Company.
- Sager, J. C. (2000). *Essays on Definition*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Salvadores, M., Alexander, P. R., Musen, M. A., & Noy, N. F. (2013). BioPortal as a Dataset of Linked Biomedical Ontologies and Terminologies in RDF. *Semantic Web*, 4(3), 277–284.
- Sampson, J. A. (1927). Peritoneal endometriosis due to the menstrual dissemination of endometrial tissue into the peritoneal cavity. *American Journal of Obstetrics and Gynecology*, 14(4), 422–469.
- Santos, C., & Costa, R. (2015). Domain specificity: Semasiological and onomasiological knowledge representation. In H. J. Kockaert & F. Steurs (Eds.), *Handbook of Terminology - vol. 1* (pp. 153–179). Amsterdam: John Benjamins Publishing Company.
- Sauvages, F. B. de. (1772). *Nosologie méthodique ou distribution des maladies en classes, en genres et en especes suivant l'esprit de Sydenham, et la méthode des botanistes*. Lyon: Jean-Marie Bruyset, Imprimeur Libraire. Retrieved from <http://gallica.bnf.fr/ark:/12148/bpt6k1041803k.image>
- Scheuermann, R., Ceusters, W., & Smith, B. (2009). Toward an Ontological Treatment of Disease and Diagnosis. In *AMIA Summit on Translational Bioinformatics* (pp. 116–120). San Francisco, CA.
- Schlich, T., & Crenner, C. (Eds.). (2017). *Technological change in modern surgery : historical perspectives on innovation*. Rochester: University of Rochester Press / Boydell & Brewer.
- Schmid, H. (n.d.). TreeTagger. Retrieved March 1, 2018, from <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>
- Schriml, L. M., Arze, C., Nadendla, S., Chang, Y.-W. W., Mazaitis, M., Felix, V., ... Kibbe, W. A. (2012). Disease Ontology: a backbone for disease semantic integration. *Nucleic Acids Research*, 40(D1),
- Schriml, L. M., & Mittra, E. (2015). The Disease Ontology: fostering interoperability between biological and clinical human disease-related data. *Mammalian Genome : Official Journal of the International Mammalian Genome Society*, 26(9–10), 584–589. <https://doi.org/10.1007/s00335-015-9576-9>
- SEF. (2012). Enzian 2012 - Klassifikation der tief infiltrierenden Endometriose. Retrieved May 19, 2018, from http://www.endometriose-sef.de/dateien/ENZIAN_2013_web.pdf
- Seppälä, A., Nykänen, P., & Ruotsalainen, P. (2014). Privacy-related context information for ubiquitous health. *JMIR MHealth and UHealth*, 2(1), e12. <https://doi.org/10.2196/mhealth.3123>
- Seppälä, S. (2007). La définition en terminologie : typologies et critères définitoires. In *Terminologie & Ontologies : Théories et Applications - Actes de la première conférence TOTH* (pp. 23–43). Annecy: Porphyre: Savoir et Connaissance.
- Seppälä, S. (2015). An ontological framework for modeling the contents of definitions. *Terminology*, 21(1), 23–50. <https://doi.org/10.1075/term.21.1.02sep>
- Shadbolt, N., Hall, W., & Berners-Lee, T. (2006). The Semantic Web Revisited. *IEEE Intelligent Systems*, (May/June), 96–101.
- Sicilia, M. Á., & Balazote, P. (2013). *Interoperability in Healthcare Information Systems: Standards, Management, and Technology*. (M. Á. Sicilia & P. Balazote, Eds.). IGI Global.

- Silva, R. (2014). *Gestão de terminologia pela qualidade*. Faculdade de Ciências Sociais e Humanas - Universidade Nova de Lisboa.
- Simoens, S., Dunselman, G., Dirksen, C., Hummelshoj, L., Bokor, A., Brandes, I., ... D'Hooghe, T. (2012). The burden of endometriosis: costs and quality of life of women with endometriosis and treated in referral centres. *Human Reproduction*, 27(5), 1292–1299.
- Simpson, J. L., Elias, S., Malinak, L. R., & Buttram, V. C. (1980). Heritable aspects of endometriosis. I. Genetic studies. *American Journal of Obstetrics and Gynecology*, 137(3), 327–331. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7377252>
- Sinclair, J. (2005). Corpus and Text - Basic Principles. In M. Wynne (Ed.), *Developing Linguistic Corpora: a Guide to Good Practice* (pp. 1–16). Oxford: Oxbow Books.
- Smith, R. (2018). Aristotle's Logic. In *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition). Stanford University. Retrieved from <https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=aristotle-logic>
- Soliman, A. M., Yang, H., Du, E. X., Kelley, C., & Winkel, C. (2016). The direct and indirect costs associated with endometriosis: a systematic literature review. *Human Reproduction*, 31(4), 712–722.
- Sørensen, K., Van den Broucke, S., Fullam, J., Doyle, G., Pelikan, J., Slonska, Z., & Brand, H. (2012). Health literacy and public health: A systematic review and integration of definitions and models. *BMC Public Health*, 12(1), 80. <https://doi.org/10.1186/1471-2458-12-80>
- Sourial, S., Tempest, N., & Hapangama, D. K. (2014). Theories on the pathogenesis of endometriosis. *International Journal of Reproductive Medicine*, 2014, 179515.
- Sowa, J. F. (1992). Conceptual graphs as a universal knowledge representation. *Computers & Mathematics with Applications*, 23(2–5), 75–93. [https://doi.org/10.1016/0898-1221\(92\)90137-7](https://doi.org/10.1016/0898-1221(92)90137-7)
- Sowa, J. F. (2000). *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove, CA: Brooks/Cole. Retrieved from <https://www.researchgate.net/publication/225070439>
- Sowa, J. F. (2007). Fads and Fallacies about Logic. *IEEE Intelligent Systems*, 22(2), 84–87.
- Statista. (2018). Development of the world population until 2030. Retrieved May 12, 2018, from <https://www.statista.com/statistics/262875/development-of-the-world-population/>
- Steindel, S. J. (2010). International classification of diseases, 10th edition, clinical modification and procedure coding system: descriptive overview of the next generation HIPAA code sets. *Journal of the American Medical Informatics Association*, 17(3), 274–282.
- Studer, R., Benjamins, R., & Fensel, D. (1998). Knowledge Engineering: Principles and Methods. *Data and Knowledge Engineering*, 25(1–2), 161–197.
- Suonuuti, H. (2001). *Guide to terminology* (2nd edition). Finnish Centre for Technical Terminology / Nordterm.
- Swan, M. (2013). The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery. *Big Data*, 1(2), 85–99. <https://doi.org/10.1089/big.2012.0002>
- Swanepoel, P. (2011). Improving the Functionality of Dictionary Definitions for Lexical Sets: The Role of Definitional Templates, Definitional Consistency, Definitional Coherence and the Incorporation of Lexical Conceptual Models. *Lexikos*, 20, 425–449. <https://doi.org/10.5788/20-0-151>
- Sydenham, T., Greenhill, W., & Latham, R. (1850). *The works of Thomas Sydenham, M.D.* London: Printed for the Sydenham Society. Retrieved from <https://archive.org/details/worksofthomassyd02sydeuoft>

- Temmerman, R. (2000). *Towards new ways of terminology description : the sociocognitive-approach*. Amsterdam/New York: John Benjamins Publishing Company. Retrieved from https://books.google.pt/books/about/Towards_New_Ways_of_Terminology_Descript.html?id=q9gUJ_ykkVAC&redir_esc=y
- Temmerman, R., & Kerremans, K. (2003). Termontography: Ontology Building and the Sociocognitive Approach to Terminology Description. In *Proceedings of CIL17* (p. n.a.). Matfyzpress.
- Temmerman, R., & Kerremans, K. (2005). What will be considered a term? From static terminology description to dynamic terminology processing. In *Proceedings of the LSP conference* (pp. 265–277). Bergamo.
- Thagard, P. (1997). The Concept of Disease: Structure and Change. Retrieved May 15, 2017, from <http://cogsci.uwaterloo.ca/~pthagard/Articles/Pages/Concept.html>
- The World Bank. (2018). Population estimates and projections. Retrieved May 12, 2018, from <http://databank.worldbank.org/data/reports.aspx?source=Health-Nutrition-and-Population-Statistics:-Population-estimates-and-projections>
- Tinneberg, H.-R., Tuttlies, F., Berkes, E., Krombach, G. A., Hackethal, A., Konrad, L., & Oehmke, F. (2014). Classification of Endometriosis. In T. Harada (Ed.), *Endometriosis* (pp. 341–362). Tokyo: Springer Japan. https://doi.org/10.1007/978-4-431-54421-0_21
- Tuttlies, F., Keckstein, J., Ulrich, U., Possover, M., Schweppe, K., Wustlich, M., ... Tinneberg, H. (2005). ENZIAN-Score, eine Klassifikation der tief infiltrierenden Endometriose. *Zentralblatt Für Gynäkologie*, 127(05), 275–281. <https://doi.org/10.1055/s-2005-836904>
- United Nations, Department of Economic and Social Affairs, P. D. (2015). *World Population Ageing 2015*. New York.
- Vayena, E., Dzenowagis, J., Brownstein, J. S., & Sheikh, A. (2018). Policy implications of big data in the health sector. *Bulletin of the World Health Organization*, 96(1), 66–68.
- Vitonis, A. F., Vincent, K., Rahmioglu, N., Fassbender, A., Buck Louis, G. M., Hummelshoj, L., ... Zondervan, K. T. (2014). World Endometriosis Research Foundation Endometriosis Phenome and biobanking harmonization project: II. Clinical and covariate phenotype data collection in endometriosis research. *Fertility and Sterility*, 102(5), 1223–1232. <https://doi.org/10.1016/j.fertnstert.2014.07.1244>
- W3C Ontology Lexicon Community Group. (2016). LEMON - Final Model Specification. <https://doi.org/10.1007/s10579-012-9182-3>
- WERF. (n.d.). World Endometriosis Research Foundation. Retrieved February 19, 2017, from <https://endometriosisfoundation.org/>
- WES. (n.d.). World Endometriosis Society. Retrieved February 19, 2017, from <http://endometriosis.ca/>
- WHA. (2013). WHA 66.24 eHealth standardization and interoperability. Geneva: World Health Assembly.
- WHA. (2015). WHA 68/15 - Strengthening Emergency and Essential Surgical Care and Anaesthesia as a Component of Universal Health Coverage. Geneva: World Health Assembly.
- WHO; ITU. (2012). *National eHealth Strategy Toolkit*. Geneva.
- WHO. (2009). *WHO Guidelines on Hand Hygiene in Health Care. Patient Safety*. Geneva.
- WHO. (2011a). Content Model Reference Guide - ICD-11 alpha. Geneva: World Health Organization.
- WHO. (2011b). *mHealth: new horizons for health through mobile technologies*. Geneva.

- WHO. (2016a). *From Innovation to Implementation: eHealth in the WHO European Region*. Copenhagen.
- WHO. (2016b). *Global strategy on human resources for health: Workforce 2030*. Geneva.
- WHO. (2016c). International statistical classification of diseases and related health problems - 10th revision. Volume 2 - Instruction manual. Geneva: World Health Organization.
- WHO. (2018). The International Classification of Health Interventions (Beta 2018). Retrieved March 21, 2018, from <https://mitel.dimi.uniud.it/ichi/>
- Wolf, G. (2010). The quantified self. Retrieved May 12, 2017, from https://www.ted.com/talks/gary_wolf_the_quantified_self?language=en
- Wood, D., Zaidman, M., Ruth, L., & Hausenblas, M. (2014). *Linked data : structured data on the Web*. Manning Publications.
- Wright, J. D., Herzog, T. J., Tsui, J., Ananth, C. V., Lewin, S. N., Lu, Y.-S., ... Hershman, D. L. (2013). Nationwide Trends in the Performance of Inpatient Hysterectomy in the United States. *Obstetrics & Gynecology*, 122(2, PART 1), 233–241. <https://doi.org/10.1097/AOG.0b013e318299a6cf>
- Wüster, E. (1968). *The machine tool : an interlingual dictionary of basic concepts [comprising an alphabetical dictionary and a classified vocabulary with definitions and illustrations]*. (United Nations. Economic Commission for Europe., Ed.). London: Technical Press.
- Wüster, E. (1974). Die allgemeine Terminologielehre - Ein Grenzgebiet zwischen Sprachwissenschaft, Logik, Ontologie, Informatik und den Sachwissenschaften. *Linguistics: An Interdisciplinary Journal of the Language Sciences*, 12(119), 61–106. <https://doi.org/10.1515/ling.1974.12.119.61>
- Wüster, E. (1991). *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*. Ergon Verlag. Retrieved from <http://www.ergon-verlag.de/bibliotheks--informationswissenschaft/monografien/wuester-eugen.php>
- Yao, M., Hu, T., Wang, Y., Du, Y., Hu, C., & Wu, R. (2017). Polychlorinated biphenyls and its potential role in endometriosis. *Environmental Pollution*, 229, 837–845. <https://doi.org/10.1016/j.envpol.2017.06.088>
- Zegers-Hochschild, F., Adamson, G. D., de Mouzon, J., Ishihara, O., Mansour, R., Nygren, K., ... World Health Organization. (2009). International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) revised glossary of ART terminology, 2009*. *Fertility and Sterility*, 92(5), 1520–1524. <https://doi.org/10.1016/j.fertnstert.2009.09.009>
- Zeyl, D., & Sattler, B. (2017). Plato's Timaeus. In *The Stanford Encyclopedia of Philosophy* (Winter 201). Stanford University. Retrieved from <https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=plato-timaeus>
- Zillner, S., & Neururer, S. (2016). Big Data in the Health Sector. In *New Horizons for a Data-Driven Economy* (pp. 179–194). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-21569-3_10

LIST OF FIGURES

Figure 1 – Word cloud with some of the key notions of healthcare today	9
Figure 2 – Current stakeholders within the eHealth framework	16
Figure 3 – Types of interoperability layers: terminology used within the scope of this work....	24
Figure 4 – The Semantic Web technology stack	26
Figure 5 – The Linked Open Data cloud	28
Figure 6 – The double semiotic triangle.....	40
Figure 7 – Tree of Porphyry as drawn by Peter of Spain (1239) (Sowa, 2007)	42
Figure 8 – Examples of knowledge representation within the scope of endometriosis.....	49
Figure 9 – Examples of medical illustrations.....	50
Figure 10 – Hippocrates’s proposal concerning the classification of diseases	52
Figure 11 – ICD-10 results for <Endometriosis>	67
Figure 12 – MeSH tree structure for <Endometriosis>	71
Figure 13 – Historical overview of SNOMED CT.....	73
Figure 14 – SNOMED CT’s Logical Model.....	74
Figure 15 – SNOMED CT’s Top-Level Hierarchy	76
Figure 16 – <Endometriosis> in SNOMED CT.....	78
Figure 17 – Definition of <Endometriosis> using SNOMED CT’s compositional grammar.....	80
Figure 18 – Definition of <Laparoscopic radical hysterectomy> via pre- and post-coordinated expressions using SNOMED CT’s compositional grammar.....	81
Figure 19 – The Disease Ontology’s core hierarchy	83
Figure 20 – The Disease Ontology’s web browser	83
Figure 21 – The Disease Ontology’s metadata for <Endometriosis>	84
Figure 22 – The Disease Ontology’s visualization canvas for <Endometriosis>.....	85
Figure 23 – The UMLS Metathesaurus tree (partial view)	88
Figure 24 – The UMLS Metathesaurus: results for <Endometriosis>	89
Figure 25 – The UMLS Metathesaurus: partial results for <Endometriosis> within SNOMED CT	91
Figure 26 – A portion of the UMLS Semantic Network (with relations) (NML, 2009).....	92
Figure 27 – Systematization of the UMLS Semantic Relations	92
Figure 28 – Example of a data graph	97
Figure 29 – Example of an RDF graph	99
Figure 30 – Example of RDF/XML.....	99
Figure 31 – ‘2014-2017’, by Georgie Wileman	106
Figure 32 – Endometriosis lesions identified by Sampson (1924)	107
Figure 33 – Possible locations of endometriosis.....	108
Figure 34 – The rASRM classification of endometriosis (1997)	124
Figure 35 – Examples of endometriosis implants (ASRM, 1997)	124
Figure 36 – The Enzian classification of Deep Infiltrating Endometriosis (SEF, 2012)	125
Figure 37 – The Endometriosis Fertility Index (Adamson & Pasta, 2010).....	128
Figure 38 – Patient-oriented resources on endometriosis	135
Figure 39 – Expert-oriented resources on endometriosis	135
Figure 40 – Proposal of a domain tree for Endometriosis	136
Figure 41 – The three axes of a <Surgical procedure>.....	143

Figure 42 – The example of <Hysterectomy> and its sub-hierarchies (ICHI and SNOMED CT)	144
Figure 43 – UML diagram of the ISO 1828 Categorial Structure.....	147
Figure 44 – Types of hysterectomy.....	150
Figure 45 – EndoTerm’s conceptualization of <Surgical procedures> by surgical approach....	153
Figure 46 – Types of <Hysterectomy> according to the <Surgical approach>.....	156
Figure 47 – Types of <Hysterectomy> according to the <Anatomical entity> removed via this procedure	156
Figure 48 – EndoTerm’s hierarchy for <Anatomical entity>	157
Figure 49 – Example of micro-concept system for <Total laparoscopic hysterectomy>	159
Figure 50 – The EndoCorpus and AntConc’s N-Gram function	168
Figure 51 – Example of how TermoStat was used to process the EndoCorpus.....	169
Figure 52 – Snippet of the list of candidate terms with its annotations.....	170
Figure 53 – Candidate terms containing “hysterectomy”	171
Figure 54 – Concordance Plot for “hysterectom*”	172
Figure 55 – Concordance results for “hysterectom*” in KWIC format, sorted by forms on the left of the Search Term	172
Figure 56 – Example of AntConc’s File View functionality.....	178
Figure 57 – Lexical network 1: types of “hysterectomy” by targeted body structure	179
Figure 58 – Lexical network 2: types of “hysterectomy” by surgical approach	180
Figure 59 – Lexical network for “laparoendoscopic single-site surgery”	189
Figure 60 – New iteration of the micro-concept system on <Surgical approach>.....	192
Figure 61 – Micro-concept system for <Laparoendoscopic single-site total hysterectomy> ...	193
Figure 62 – Micro-concept system for <Laparoendoscopic single-site ovarian cystectomy>...	194
Figure 63 – <Laparoendoscopic single-site total hysterectomy> in TeDI’s concept editor	196
Figure 64 – <Minimally invasive surgical approach with single skin incision> and its specific differences	197
Figure 65 – An example of TeDI’s relation editor containing the extended EndoTerm’s non-hierarchical relations	199
Figure 66 – Formal definition of <Laparoendoscopic single-site total hysterectomy> in RDF/OWL	199
Figure 67 – Natural language definition using Lexonomy: entry with the XML Editor	201
Figure 68 – Natural language definition using Lexonomy: entry as shown in the public dictionary	202
Figure 69 – Example of SNOMED CT descriptions for <Laparoendoscopic single-site total hysterectomy>	204
Figure 70 – Example of formal concept definition for <Laparoendoscopic single-site total hysterectomy> using SNOMED CT’s compositional grammar	205

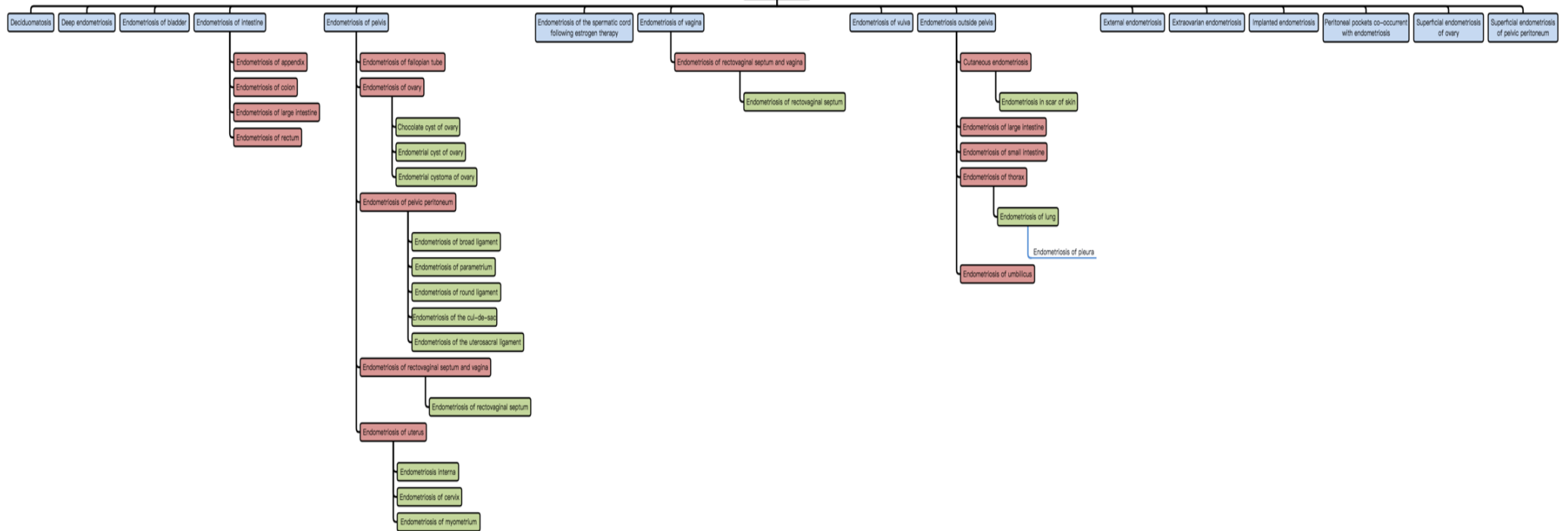
LIST OF TABLES

Table 1 – Categories within eHealth and respective examples	14
Table 2 – Types of interoperability layers (overview).....	23
Table 3 – Overview of Cimino (1998), Chute et al. (1998) and Rector’s (1999) approaches	59
Table 4 – Typologies of biomedical terminological systems.....	63
Table 5 – Core chapters of the ICD-10	66
Table 6 – Main hierarchy in MeSH.....	70
Table 7 – MeSH results for <Endometriosis>: metadata	72
Table 8 – Overview of SNOMED CT’s Top-Level Concepts.....	77
Table 9 – SNOMED-CT’s information on the descriptions associated to <Endometriosis>	79
Table 10 – The Disease Ontology’s relationships.....	85
Table 11 – AUI, SUI, LUI and CUI: the building blocks of the Metathesaurus	90
Table 12 – Example of RDF statements	97
Table 13 – Studies on the diagnostic delay of endometriosis	114
Table 14 – Studies concerning the economic burden of endometriosis.....	115
Table 15 – EndoTerm’s potential core categories	138
Table 16 – The example of <Total abdominal hysterectomy> in current procedure TS.....	142
Table 17 – The ISO 1828’s Categorial Structure.....	145
Table 18 – Specific-difference matrix supporting EndoTerm’s conceptualization by <Surgical approach>	152
Table 19 – EndoTerm’s categorial structure for <Surgical procedure> concepts	154
Table 20 – Overview of ISO 1828 CAST, SNOMED CT’s and EndoTerm’s categorial structures.....	155
Table 21 – Overview of the files that constitute the EndoCorpus	167
Table 22 – Morphosyntactic patterns for “hysterectomy”	175
Table 23 – Examples of forms displaying potential cases of ellipsis	176
Table 24 – KPs used to express the types of surgical approaches to “hysterectomy”	177
Table 25 – KPs used to express the types of “hysterectomy” by targeted body structure	178
Table 26 – KPs used to express synonymy.....	179
Table 27 – Examples from the corpus regarding “robotic hysterectomy” and other related terms.....	181
Table 28 – Terms referring to LESS surgery	187
Table 29 – EndoTerm’s full list of non-hierarchical conceptual relations.....	198
Table 30 – Definitional template for <Laparoendoscopic single-site total hysterectomy>	200
Table 31 – Definitional template for <Laparoendoscopic single-site ovarian cystectomy>	200

ANNEXES

ANNEX 1

Subtypes of |Endometriosis (disorder)| in SNOMED CT, represented via the
mind-mapping software XMind



ANNEX 2

Overview of surgical approaches in procedure terminological systems

ICHI [Means - Approach]		ICD-10-PCS [Approach]		CCI [Approach]		CCAM [Mode d'accès]	
Open approach [AA]	"Exposing the actual site of the intervention by incision of the skin or mucous membrane and any other underlying tissue."	Open [0]	"Cutting through the skin or mucous membrane and any body layers necessary to expose the site of the procedure."	Open [LA]	"Usually implies that an incision is made to gain access to the site."	Ouvert [A]	"Accès exposant le site opératoire, par incision des téguments (peau, muqueuses) et de tout autre tissu sousjacent, sans introduction d'un instrument d'optique."
Percutaneous endoscopic [AB]	"Endoscopic access, by puncture or minor incision, through the skin and any other body layers necessary to access the site of the intervention."	Percutaneous endoscopic [4]	"Entry, by puncture or minor incision, of instrumentation through the skin or mucous membrane and/or any other body layers necessary to reach and visualize the site of the procedure."	Endoscopic [DA]	"Includes interventions done via the laparoscope, thoracoscope, hysteroscope, etc. Implies a scope is used to view the intervention. Incisions are minimal in size and are often referred to as ports or as minimally invasive technique."	Endoscopique transpariétal [C]	"Accès au site opératoire, avec introduction d'un instrument d'optique, par ponction ou incision minime des téguments (peau, muqueuses) et de tout autre tissu sousjacent."
Per orifice [AC]	"Access by a natural or artificial orifice."	Via natural or artificial opening [7]	"Entry of instrumentation through a natural or artificial external opening to reach the site of the procedure."	Per orifice [CA]	"Includes the interventions that are done through an existing orifice but without a scope or incision."	Transorificiel [D]	"Accès à la lumière d'une cavité ou d'un conduit anatomique en passant par un orifice externe naturel ou artificiel (stomie cutanée), sans introduction d'un instrument d'optique."
Per orifice endoscopic [AD]	"Endoscopic access by a natural or artificial orifice."	Via natural or artificial opening endoscopic [8]	"Entry of instrumentation through a natural or artificial external opening to reach and visualize the site of the procedure."	Endoscopic per orifice [BA]	"Includes interventions done via the cystoscope, bronchoscope, etc. The scope is inserted via an orifice (natural or one surgically created) and no incision is required."	Endoscopique transorificiel [E]	"Accès à la lumière d'une cavité ou d'un conduit anatomique en passant par un orifice externe naturel ou artificiel (stomie cutanée), avec introduction d'un instrument d'optique."
Percutaneous [AE]	"Entry, by puncture or minor incision, with instrumentation through the skin and any other body layers necessary to access the site of the intervention."	Percutaneous [3]	"Entry, by puncture or minor incision, of instrumentation through the skin or mucous membrane and/or any other body layers necessary to reach the site of the procedure."	Percutaneous [HA]	"Includes the interventions done through a needle, large bore needle or catheter. Diagnostic imaging may accompany these interventions."	Transpariétal [B]	"Accès au site opératoire par ponction ou incision ponctuelle des téguments (peau, muqueuses) et de tout autre tissu sousjacent, sans introduction d'un instrument d'optique."
Percutaneous transluminal [AF]	"Entry, by puncture or minor incision, of instrumentation through the skin and any other body layers necessary to reach the site of the intervention in the lumen of a vessel or duct."	-----	-----	-----	-----	-----	-----

External [AE]	"The intervention is either performed directly on the skin or mucous membrane or indirectly by application of external force through the skin or mucous membrane."	External [X]	"Procedures performed directly on the skin or mucous membrane or indirectly by the application of external force through the skin or mucous membrane."	External [JA]	"Includes the interventions done on the outside of the body that do not require an incision, scope or needle to gain access to the site."	-----	-----
Combined approach, percutaneous endoscopic and per orifice [AI]	-----	Via natural or artificial opening with percutaneous endoscopic assistance [F]	"Entry of instrumentation through a natural or artificial external opening, and entry, by puncture or minor incision, of instrumentation through the skin or mucous membrane and any other body layers necessary to aid in the performance of the procedure."	-----	-----	-----	-----
Combined approach, percutaneous and per orifice endoscopic [AJ]	-----	-----	-----	-----	-----	-----	-----
Combined approach, not otherwise specified [AK]	-----	-----	-----	-----	-----	-----	-----
Unspecified approach [AZ]	-----	-----	-----	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	Intraluminal transpariétal [F]	"Accès à la lumière d'une cavité ou d'un conduit anatomique par ponction ou incision minime des téguments, sans introduction d'un instrument d'optique."
-----	-----	-----	-----	-----	-----	Endoscopique intraluminal transpariétal [G]	"Accès à la lumière d'une cavité ou d'un conduit anatomique par ponction ou incision minime des téguments, avec introduction d'un instrument d'optique."

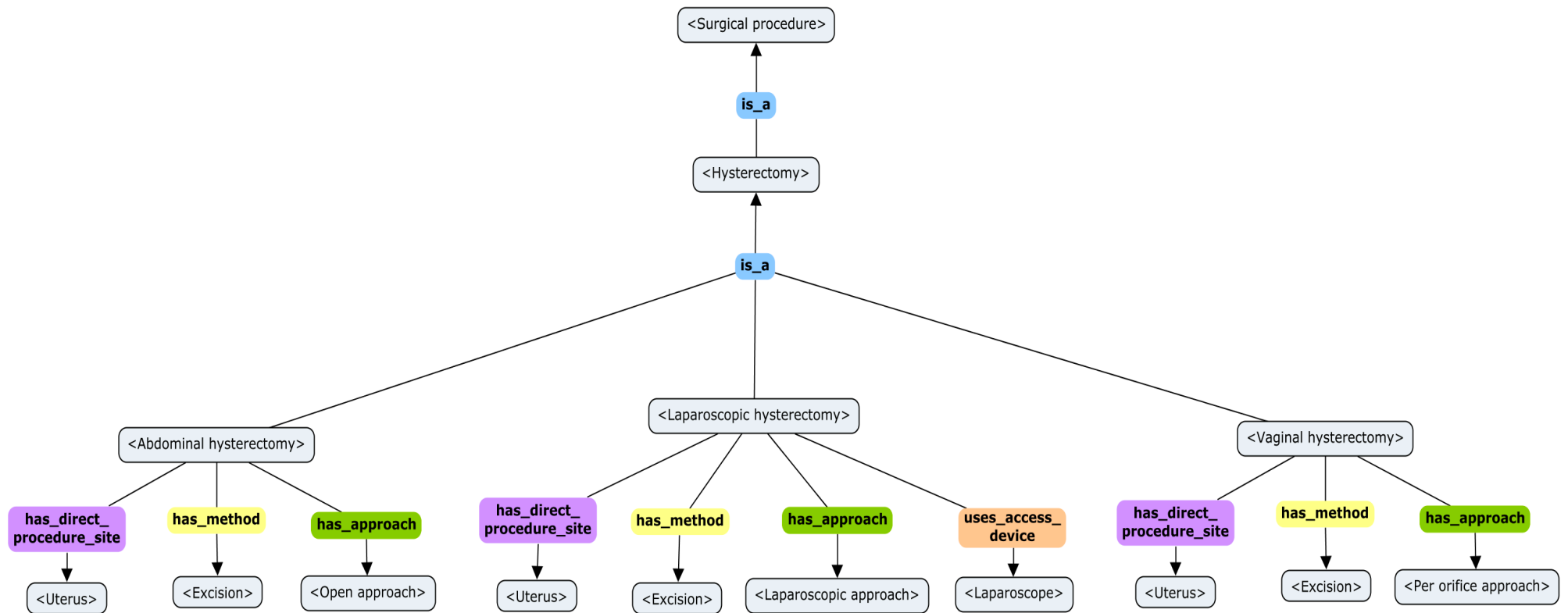
ANNEX 3

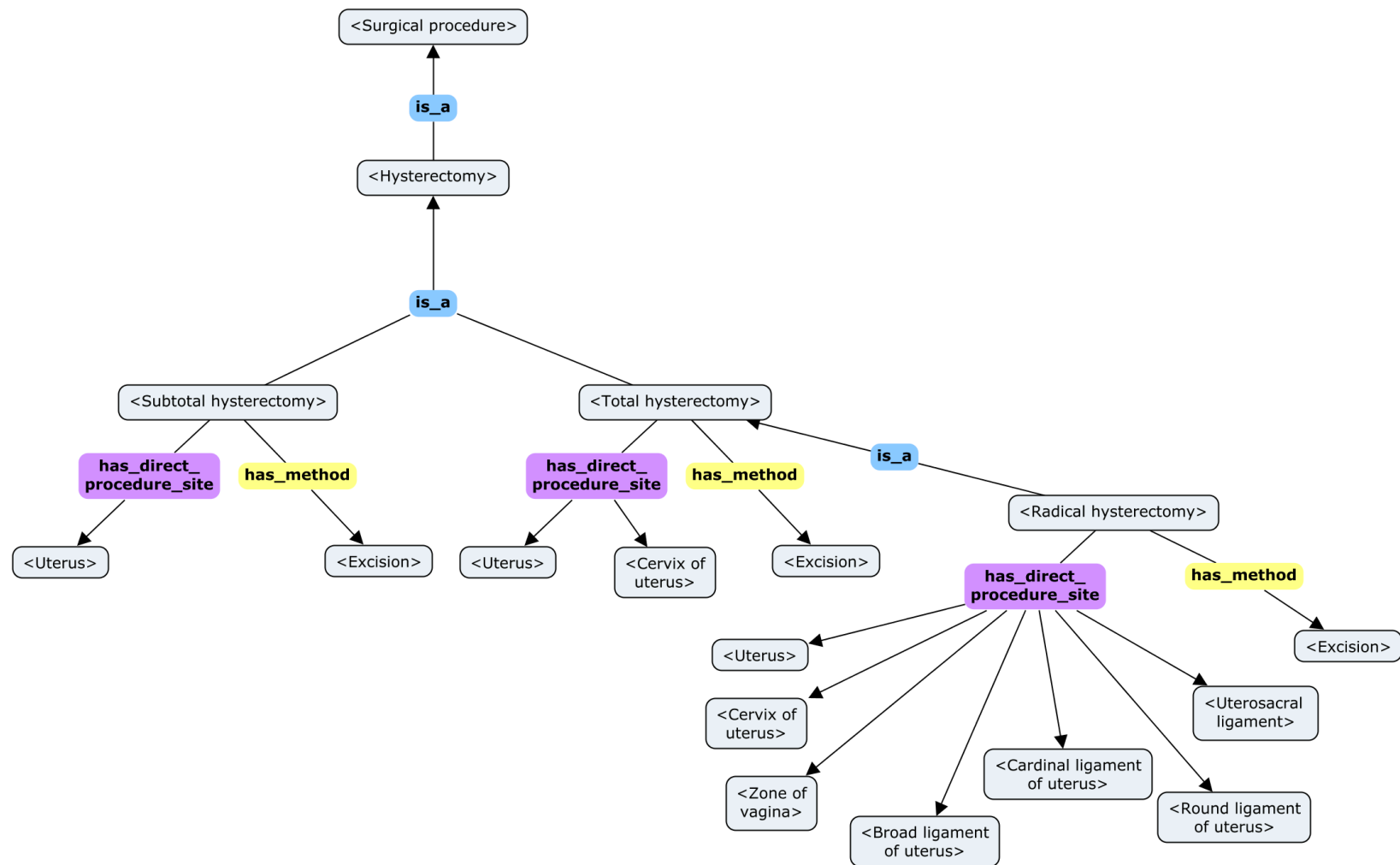
SNOMED CT |Procedure| concepts (attribute relationships and
respective values)

ATTRIBUTES	DESCRIPTION	SUBHIERARCHY	ALLOWABLE RANGE
Access	describes the route used to access the site of the procedure		Surgical access values
Direct substance	describes the Substance or Pharmaceutical/biologic product on which the procedure's method directly acts		Substance Pharmaceutical/biologic product
Has focus	specifies the Clinical finding or Procedure which is the focus of a procedure		Clinical finding Procedure
Has intent	specifies the intent of a procedure		Intents (nature of procedure values)
Method	specifies the action being performed to accomplish the procedure. It does not include the surgical approach, equipment or physical forces		Action
Priority	refers to the priority assigned to a procedure		Priorities
Procedure device	describes the devices associated with a procedure	Direct device Indirect device Using device Using access device	Device
Procedure morphology	specifies the morphology or abnormal structure involved in a procedure	Direct morphology Indirect morphology	Morphologically abnormal structure
Procedure site	describes the body site acted on or affected by a procedure	Procedure site - Direct Procedure site - Indirect	Anatomical or acquired body structure
Recipient category	specifies the type of individual or group upon which the action of the procedure is performed		Person Family Community Donor for medical or surgical procedure Group
Revision status	specifies whether a procedure is primary or a revision		Primary operation Revision - value Part of multistage procedure
Route of administration	represents the route by which a procedure introduces a given substance into the body		Route of administration value
Surgical approach	specifies the directional, relational or spatial access to the site of a surgical procedure		Procedural approach
Using energy	describes the energy used to execute an action		Physical force
Using substance	describes the substance used to execute the action of a procedure, but it is not the substance on which the procedure's method directly acts		Substance

ANNEX 4

Micro-concept systems for types of <Hysterectomy>
by <Surgical approach> and by <Procedure site>





ANNEX 5

Lexical network for “hysterectomy”
according to the HOW (EndoCorpus data)

