

**Nouvelle méthodologie de construction  
d'ontologies médicales  
Cas d'étude : diagnostic de la pneumonie**

Par  
Sabrina Azzi

Thèse présentée en vue de l'obtention  
du diplôme de Philosophiae Doctor (Ph.D.)  
(Sciences et Technologies de l'Information)

Département d'informatique et d'ingénierie  
Université du Québec en Outaouais  
Mars 2021

© Sabrina Azzi, 2021

©Tous droits réservés

Cette licence signifie qu'il est interdit de reproduire, d'enregistrer ou de diffuser en tout ou en partie, le présent document. Le lecteur qui désire imprimer ou conserver sur un autre média une partie importante de ce document, doit obligatoirement en demander l'autorisation à l'auteur.

# PRÉSENTATION DU JURY

Cette thèse a été évaluée par un jury composé de :

Prof. Kamel Adi  
Président du jury  
Université du Québec en Outaouais

Prof. Karim El Guemhioui  
Examineur interne  
Université du Québec en Outaouais

Prof. Wojtek Michalowski  
Examineur externe  
Université d'Ottawa

Prof. Véronique Nabelsi  
Directrice de thèse  
Université du Québec en Outaouais

Prof. Michal Iglewski  
Co-directeur de thèse  
Université du Québec en Outaouais

# REMERCIEMENTS

Ma thèse qui s'étend de février 2016 à août 2020 a été rendue possible grâce aux efforts et à la collaboration de plusieurs personnes.

Je remercie vivement Prof. Véronique Nabelsi, ma directrice de thèse et Prof. Michal Iglewski, mon co-directeur de thèse, pour leur soutien indéfectible et leur implication tout au long du processus. Leurs nombreux conseils, encouragements, disponibilités, patiences et valeurs humaines m'ont été fort utiles lors de la réalisation de la thèse. Un autre merci très spécial à Michal Iglewski pour son support inestimable malgré toutes les difficultés rencontrées. Ces moments de travail, durant cette période de COVID-19, seront à jamais gravés dans ma tête.

Mes remerciements sont adressés aux médecins qui ont participé à la définition du domaine et la portée de l'ontologie. Leur évaluation et leur contribution furent d'une pertinence et d'une utilité sans équivoque pour la finalisation de ce projet de recherche.

Je tiens également à remercier Prof. Karim El Guemhioui de l'Université du Québec en Outaouais et Prof. Wojtek Michalowski de l'Université d'Ottawa qui ont aimablement évalué cette thèse.

Je remercie du fond du cœur les professeurs Stéphane Gagnon, Hamed Motaghi et Larbi Talbi de l'Université du Québec en Outaouais pour leur accompagnement très enrichissant durant toutes ces années, ainsi qu'Alan Bernardi du CILEX pour son support, notamment en 2016.

J'adresse une mention très particulière au Prof. Djamel Rebaïne de l'Université du Québec à Chicoutimi qui a eu un impact considérable sur mon parcours universitaire. Merci Djamel At Ammar !

Je tiens à souligner que sans l'apport financier du Centre intégré de santé et des services sociaux de l'Outaouais (CISSSO), ce travail n'aurait pas vu le jour.

Merci à ma famille pour sa patience infinie, ses encouragements et son amour inconditionnel.

Un clin d'œil à mes ami(e)s pour leur soutien et d'avoir toujours cru en moi. Ils se reconnaîtront.

*Pour toi, douce maman.  
I kem, a yemma taħnint.*

# RÉSUMÉ

L'application des ontologies dans le domaine médical a connu une grande popularité cette dernière décennie. Ces ontologies sont développées pour répondre à des besoins de compréhension, de partage et d'intégration des connaissances. Le développement d'une nouvelle ontologie médicale représente un sérieux défi, car les méthodologies existantes ne fournissent pas beaucoup d'indications sur la manière dont les différentes étapes doivent être réalisées. Ce défi devient plus difficile lorsqu'il existe une multitude de sources de connaissances médicales et d'ontologies couvrant le domaine et ayant chacune une représentation différente du même concept. De plus, le développement d'ontologies est un travail multidisciplinaire impliquant des experts du domaine et des ontologues. Il n'est pas facile de réunir une telle équipe et, par conséquent, les ontologies sont souvent créées avec une expertise limitée soit dans le domaine médical ou dans le domaine de l'ingénierie des ontologies.

Dans ce contexte, notre réflexion a porté d'une part, sur la construction d'une nouvelle méthodologie détaillée pour la création d'une ontologie médicale avec une participation limitée des experts du domaine. D'autre part, sur une preuve de concept qui consiste en l'application de la méthodologie sur le développement d'une ontologie couvrant le diagnostic de la pneumonie.

La méthodologie utilise les guides de bonnes pratiques cliniques (GBPC) pour construire une ontologie préliminaire, et ensuite, elle réutilise des ontologies existantes qui sont pertinentes pour le domaine. La méthodologie guide l'ontologue à résoudre les incohérences et les conflits possibles entre des concepts ayant des représentations différentes dans plusieurs ontologies. Ce n'est que lors des phases de définition du domaine et d'évaluation de l'ontologie que les experts du domaine sont consultés. Pour des fins d'interopérabilité et de réutilisabilité avec d'autres ontologies biomédicales, la méthodologie suit les principes d'Open Biomedical Ontologies (OBO) Foundry<sup>1</sup>.

---

<sup>1</sup> <http://www.obofoundry.org/principles/fp-000-summary.html>, dernier accès le 17/03/2021

La méthodologie a été expérimentée pour la construction d'une ontologie pour le diagnostic de la pneumonie (PNADO). La définition du domaine de PNADO et son évaluation ont impliqué plusieurs médecins. Les résultats de l'évaluation ont permis de valider notre méthodologie et de démontrer son efficacité.

PNADO est la première ontologie de diagnostic de la pneumonie déclarée, développée pour représenter les différents aspects du diagnostic dans un format logique et formel. Elle couvre les éléments suivants : les symptômes, les signes cliniques, les types de pneumonie, les antécédents, les pathogènes, les diagnostics différentiels, les complications, les tests de laboratoire, l'imagerie et les résultats. Cette dernière sera utilisée par un système d'aide à la décision clinique qui sera intégré dans la plateforme SEKMED (Software for the Evolution of Knowledge in MEDicine).

Bien que notre méthodologie ait été destinée à la construction d'ontologies médicales, elle peut très bien être appliquée à la construction d'ontologies couvrant d'autres domaines et disciplines, à condition d'avoir des documents qui synthétisent les connaissances du domaine à l'instar des GBPC.

**Mots clés :** Ingénierie des ontologies médicales, Réutilisation d'ontologies, Guides de bonnes pratiques cliniques, OBO Foundry, Questions de compétence, Diagnostic de la pneumonie.

# ABSTRACT

The application of ontologies in the medical field has become very popular in the last decade. These ontologies are developed to meet the needs of understanding, sharing and integrating knowledge. The development of a new medical ontology is a serious challenge since existing methodologies do not provide much guidance on how the steps should be carried out. This challenge becomes more difficult when there is a multitude of sources of medical knowledge and ontologies covering the field, each having a different representation of the same concept. Moreover, ontology development is a multidisciplinary work involving domain experts and knowledge engineers. Bringing together such a team is not easy, and therefore often, ontologies are created with limited expertise either in the medical domain or in ontology engineering domain.

In this context, our reflection focused, on the one hand, on the construction of a new detailed methodology for creating a medical ontology with limited involvement of domain experts. On the other hand, on a proof of concept that consists of applying the methodology on the development of an ontology covering the pneumonia diagnosis.

The methodology uses clinical practice guidelines (CPG) to build a preliminary ontology and then, reuse existing ontologies that are relevant to the domain of interest. The methodology guides the ontologist to resolve inconsistencies and conflicts between concepts having different representations in different ontologies. It is only in the domain definition and the evaluation phases when the domain experts are consulted. To enhance interoperability and reusability with other biomedical ontologies, the methodology follows the OBO Foundry principles.

The methodology has been tested for building a pneumonia diagnosis ontology (PNADO). The definition of the PNADO domain as well as its evaluation involved several physicians. The results of the evaluation validated our methodology and showed its effectiveness.

PNADO is the first reported pneumonia diagnosis ontology developed to represent different diagnosis aspects in a formal logical format. It covers symptoms, clinical signs, types of pneumonia, history, pathogens, differential diagnoses, complications, laboratory tests, imaging, and results. It will be used by a clinical decision support system that will be

integrated into the SEKMED (Software for the Evolution of Knowledge in MEDicine) platform.

Although our methodology was intended for medical ontology building, it can be very well applied to the building of ontologies covering the domains of other fields, by using documents that synthesize the knowledge of the domain, such as CPG.

**Keywords:** Medical ontology engineering, Ontology reuse, Clinical practice guidelines, OBO Foundry, Competency questions, Pneumonia diagnosis.

# AGZUL

Deg teecrett (mraw) iseggasen ineggura asnas n tentulujiyin deg tayult n tsujjiyt yuyal mucae, yettwassen s waṭas ger medden. Tintulujiyin-a ttwabnant i lmendad n yisiriyeen n tigzi, n uzizzer d usekcem n tmussniwin. Asnerni n tentulujit tasujjayt tamaynut d acqirrew n tidett, acku tsnarrayin yellan yakan ur ttakkent ara atas n yiwellihen yef wamek ilaq ad ttwabnunt tšeddarin yemxallafaxen. Acqirrew-a yettiweir ugar mara ilin waṭas n yiwbula n tmussniwin tisnujjayin akked tentulujiyin yerzan tayult, yerna tal yiwet s ugenses-ines yemgaraden i yiwet n tmiḍrant. Rnu yer-s, asnerni n tentulujiyin d axeddim agetneḍfaṛ yeenan atas n yimassanen n tyulin nniḍen. Asemblili n ugraw am wa mačči d ayen fessusen, s wakka tintulujiyin zgant ttwasebeddayent s tmussni mezziyet ama deg tayult n tesnujjit ney n tjenyurt n tentulujiyin.

Deg usatal-a (tagnitt-a), axemmem-nney yewwi-d seg tama yef usali n tesnarrayt tiseddit tamaynut i usnulfu n tentulujit tasnujjit s uttekki amectuḥ n yimassanen n tayult. Seg tama nniḍen, yef utebbet n tmiḍrant yerzan asnas n tesnarrayt yef usnerni n tentulujin ara d-iledhun d usennefru n tergayt n turin.

Tasnarrayt tessexdam tidlisin n uwelleh yef usemres akliniki i lebni n Tuntulujit tamezwarut, syina ad teiwed (tales) asexdem n tentulujiyin yellan yakan yerna d tisefkin i tayult-a. Tasnarrayt tettwehhi amussnaw n tentulujit akken ad yefru wer akerez d tenmegliwin izemren ad ilint gar tmiḍranin yesean tigensas yemgaraden deg waṭas n tentulujiyin. Imussnawen (imazzagen) n tayult ad tuciwren kan deg nnuba n usbadu d uktazal n tayult. Gef tgermhelt d tulsaseqdec akked tentulijiyin tibiyusnujjiyin-nniḍen, tasnarrayt nney tettafar imenzayen n OBO Foundry.

Tasnarrayt tettwajerreb i lebni n tentulujit n usennefru n tergayt n turin (PNADO). Asbadu n tayult n PNADO d uktazal-is glan-d s tikkin n waṭas yimejjayen. S igemmaḍ n uktazal nwekked-d tanfa n tesnarrayt, yerna nbeyyen-d tamlilt-is.

PNADO d tuntulujit n usennefru tamezwarut n tergayt n turin i d-yettwaxelqen, tettwasbedd-d akken ad tesgenses timezra n usennefru deg umasal amezlan alyan. Nettat terza iferdisen-agi : iskanen, isyilen iklinikiyen, tisekkiwin n tergayt, inezwar, igatṭanen, ikayaden n tnaremt, n tissugent, akked yigemmaḍ. Tuntulujit adtettwasexdem deg unagraw n uwiwen n uytas

akliniki ara yettwagren deg t̄er̄yert SEKMED (Software for the Evolution of Knowledge in MEDicine).

Ulamma tasnarrayt-nney twehha ̄er usali n ten̄ulujiyin tisnujjiyin, yezmer ad tettwasnes akken ilaq deg usali n ten̄ulujiyin yerzan tayulin d tneḏfaḗ-nniḏen.

**Awalen-tisura:** Tajenyurt n ten̄ulujiyin tisnujjayin, Tulsaseqdec n ten̄ulujiyin, Tidlisin n usnas akliniki yelhan, OBO Foundry, Timsal n tmussni, Asennefru n tergayt n turin.

# TABLES DES MATIÈRES

<b>PRÉSENTATION DU JURY</b> .....	<b>I</b>
<b>REMERCIEMENTS</b> .....	<b>II</b>
<b>RÉSUMÉ</b> .....	<b>V</b>
<b>ABSTRACT</b> .....	<b>VII</b>
<b>AGZUL</b> .....	<b>IX</b>
<b>TABLES DES MATIÈRES</b> .....	<b>XI</b>
<b>LISTE DES TABLEAUX</b> .....	<b>XV</b>
<b>LISTE DES FIGURES</b> .....	<b>XVI</b>
<b>LISTE DES ACRONYMES, SIGLES ET ABRÉVIATIONS</b> .....	<b>XVII</b>
<b>INTRODUCTION GÉNÉRALE</b> .....	<b>1</b>
<b>CHAPITRE 1 PROBLÉMATIQUE</b> .....	<b>3</b>
1.1 DIFFICULTÉS FACE À LA GESTION DES CONNAISSANCES ET DE LA PRISE DE DÉCISION MÉDICALES .....	3
<i>1.1.1 Processus décisionnels relatifs à la pratique médicale</i> .....	3
1.1.1.1 Processus d'évaluation médicale .....	5
1.1.1.2 Processus de diagnostic.....	5
1.1.1.3 Processus d'intervention .....	5
1.1.1.4 Processus d'évaluation continue .....	6
1.2 CONSTRUCTION D'ONTOLOGIES MÉDICALES .....	6
1.3 PNEUMONIE ET ERREURS DANS LE PROCESSUS DE DIAGNOSTIC .....	7
1.4 CONCLUSION.....	9
<b>CHAPITRE 2 REVUE DE LITTÉRATURE</b> .....	<b>10</b>
2.1 ONTOLOGIES .....	10
<i>2.1.1 Définitions</i> .....	10
2.1.1.1 Conceptualisation.....	11
2.1.1.2 Spécification formelle et explicite .....	12
2.1.1.3 Importance du terme 'partagé' .....	12
<i>2.1.2 Ontologie et terminologie, quelle différence ?</i> .....	13
<i>2.1.3 Mot, terme, concept, notion</i> .....	13
<i>2.1.4 Composantes d'une ontologie</i> .....	15

2.1.5	<i>Ontologies et logiques de description</i> .....	16
2.1.6	<i>Raisonnement</i> .....	17
2.1.7	<i>Quelques ressources terminologiques et ontologiques en médecine</i> .....	18
2.1.8	<i>Conclusion</i> .....	22
2.2	CONSTRUCTION DES ONTOLOGIES .....	23
2.2.1	<i>Questions de compétence et leur rôle dans l'ingénierie des connaissances</i>	23
2.2.2	<i>Cycle de vie d'une ontologie</i> .....	23
2.2.3	<i>Aperçu des méthodologies de construction manuelle</i> .....	26
2.2.3.1	Critères d'analyse.....	26
2.2.3.2	Analyse des méthodologies.....	27
2.2.4	<i>Aperçu des méthodologies de construction automatique des ontologies</i> ...	33
2.2.5	<i>Failles des méthodologies manuelles de construction des ontologies</i> .....	36
2.2.6	<i>Conclusion</i> .....	36
2.3	RÉUTILISATION DES ONTOLOGIES - ENTRE THÉORIE ET APPLICATION .....	37
2.3.1	<i>Définition de la réutilisation</i> .....	37
2.3.2	<i>Défis de la réutilisation</i> .....	38
2.3.3	<i>Les ontologies génériques ou de niveau supérieur</i> .....	39
2.3.4	<i>Qu'est-ce qu'on réutilise ?</i> .....	40
2.3.5	<i>Pourquoi une ontologie de niveau supérieur ?</i> .....	41
2.3.6	<i>Basic Formal Ontology (BFO)</i> .....	42
2.3.6.1	Quand et comment réutiliser BFO ? .....	45
2.3.7	<i>Ontology for Biomedical Science (OGMS)</i> .....	45
2.3.8	<i>Conclusion</i> .....	46
2.4	REPERTOIRES D'ONTOLOGIES : PROLIFERATION ET CHEVAUCHEMENT .....	46
2.4.1	<i>Des répertoires de données aux répertoires d'ontologies</i> .....	46
2.4.2	<i>Répertoires d'ontologies biomédicales</i> .....	48
2.4.2.1	OBO Foundry.....	48
2.4.2.2	BioPortal .....	51
2.4.2.3	UMLS .....	52
2.4.2.4	AberOWL .....	54
2.4.3	<i>Prolifération des ontologies, chevauchements et conflits</i> .....	54
2.4.4	<i>Conclusion</i> .....	55
2.5	GUIDES DE BONNES PRATIQUES CLINIQUES .....	56

2.5.1	<i>Définition</i>	56
2.5.2	<i>GBPC dans la pyramide des preuves</i>	57
2.5.3	<i>Élaboration des GBPC</i>	58
2.5.4	<i>Obstacles à l'utilisation des GBPC</i>	61
2.5.5	<i>Formalisation des GBPC</i>	62
2.5.5.1	<i>Différents niveaux de formalisation</i>	62
2.5.5.2	<i>Langages de représentation des GBPC</i>	64
2.5.6	<i>Conclusion</i>	65
2.6	CONCLUSION DE LA REVUE DE LITTÉRATURE	66
	<b>CHAPITRE 3 OBJECTIFS DE LA RECHERCHE</b>	<b>67</b>
	<b>CHAPITRE 4 MÉTHODOLOGIE DE LA RECHERCHE</b>	<b>69</b>
4.1	DÉFINITION DU DOMAINE ET DE LA PORTÉE DE L'ONTOLOGIE	71
4.1.1	<i>Cas de PNADO</i>	71
4.2	CONSTITUTION D'UN CORPUS DE TEXTES ET EXTRACTION DE TERMES	72
4.2.1	<i>Élaboration du corpus des connaissances</i>	72
4.2.1.1	<i>Cas de PNADO</i>	73
4.2.2	<i>Extraction des termes</i>	75
4.2.2.1	<i>Cas de PNADO</i>	76
4.3	CONSTRUCTION DE L'ONTOLOGIE PRÉLIMINAIRE	77
4.3.1	<i>Normalisation sémantique</i>	78
4.3.1.1	<i>Cas de PNADO</i>	79
4.3.2	<i>Formalisation des connaissances</i>	80
4.3.2.1	<i>Cas de PNADO</i>	80
4.3.3	<i>Opérationnalisation</i>	80
4.3.3.1	<i>Cas de PNADO</i>	81
4.4	RÉUTILISATION DES ONTOLOGIES	82
4.4.1	<i>Recherche des ontologies</i>	82
4.4.1.2	<i>Cas de PNADO</i>	83
4.4.2	<i>Choix des ontologies</i>	83
4.4.2.1	<i>Cas de PNADO</i>	84
4.4.3	<i>Résolution de conflits</i>	86
4.4.3.1	<i>Cas de PNADO</i>	87
4.5	ÉVALUATION DE L'ONTOLOGIE	90

4.2.1	<i>Approches et critères d'évaluation</i> .....	90
4.5.2	<i>Cohérence interne et respect des pratiques standard de l'ontologie</i> .....	90
4.5.3	<i>Évaluation avec des sources dépendantes du domaine</i> .....	91
4.5.4	<i>Évaluation avec des experts du domaine</i> .....	91
4.5.5	<i>Cas de PNADO</i> .....	91
4.6	DOCUMENTATION .....	95
4.6.1	<i>Cas de PNADO</i> .....	96
4.7	MAINTENANCE ET EVOLUTION .....	96
4.7.1	<i>Cas de PNADO</i> .....	96
4.8	CONCLUSION.....	97
	<b>CHAPITRE 5 RÉSULTATS ET DISCUSSION.....</b>	<b>98</b>
5.1	CARACTERISTIQUES DE PNADO.....	101
5.2	EXEMPLES DE PROBLÈMES LIÉS À L'INTÉGRATION AVEC D'AUTRES ONTOLOGIES MÉDICALES .....	105
5.3	ASPECT TEMPOREL DANS PNADO .....	108
	<b>CONCLUSIONS ET PERSPECTIVES.....</b>	<b>109</b>
	<b>ANNEXES .....</b>	<b>112</b>
	ANNEXE A : MATÉRIEL POUR L'ÉVALUATION DE PNADO PAR LES MÉDECINS .....	112
	<i>Partie I : Généralités sur PNADO</i> .....	112
	<i>Partie II : Évaluation en utilisant la technique de l'échelle (laddering technique)</i> .....	118
	<i>Partie III : Évaluation en utilisant les questions de compétence</i> .....	119
	ANNEXE B : DOCUMENTATION DE PNADO .....	121
	<b>BIBLIOGRAPHIE.....</b>	<b>124</b>

# LISTE DES TABLEAUX

TABLEAU 1 : ANALYSE DES METHODOLOGIES DE CONSTRUCTION D'ONTOLOGIES A PARTIR DE TEXTE. ....	32
TABLEAU 2 : RESUME DE LA NATURE ET REALISATEURS DES PHASES DE LA METHODOLOGIE. ..	70
TABLEAU 3 : TABLEAU ILLUSTRANT DES REPERTOIRES NATIONAUX ET INTERNATIONAUX CONTENANT DES GBPC SUR LA PNEUMONIE. ....	74
TABLEAU 4 : EXEMPLES DE SYNTAGMES EXTRAITS PAR TEXT2ONTO. ....	77
TABLEAU 5 : EXEMPLES DE CONFLITS RÉSOLUS AVEC LES QRC. ....	89
TABLEAU 6 : EXEMPLE D'ANNOTATION D'UN PARAGRAPHE ISSU D'UN GBPC AVEC PNADO...	93
TABLEAU 7 : RÉSUMÉ DES AMÉLIORATIONS APPORTÉES À METHONTOLOGY. ....	99
TABLEAU 8 : REUTILISATION D'AUTRES ONTOLOGIES DANS PNADO. ....	100
TABLEAU 9 : QUELQUES RELATIONS DE PNADO. ....	102

# LISTE DES FIGURES

FIGURE 1 : CYCLE DE VIE D'UNE ONTOLOGIE [71-73].	24
FIGURE 2 : PROCESSUS DE DEVELOPPEMENT D'ONTOLOGIE AVEC METHONTOLOGY [80].	29
FIGURE 3 : LES TROIS NIVEAUX D'ONTOLOGIES (SCHÉMA ADAPTÉ DE [107]).	40
FIGURE 4 : HIÉRARCHIE DE BFO CONTINUANT [118].	43
FIGURE 5 : HIÉRARCHIE DE BFO OCCURRENT [118].	44
FIGURE 6 : LES DIFFERENTS SOUS-DOMAINES INTEGRES DANS UMLS [124].	54
FIGURE 7 : PYRAMIDE DES PREUVES.	58
FIGURE 8 : PROCESSUS DE DEVELOPPEMENT DE GBPC SELON L'OMS, EXTRAIT DU HANDBOOK FOR GUIDELINE DEVELOPMENT [133].	60
FIGURE 9 : NIVEAUX DE FORMALISATION DES GBPC.	63
FIGURE 10 : EXTRAIT D'UN NICE PATHWAY POUR LA DESCRIPTION DES SYMPTOMES ET DES SIGNES CLINIQUES POUR LES ENFANTS.	64
FIGURE 11 : LES PHASES DE CONSTRUCTION D'UNE ONTOLOGIE MÉDICALE.	70
FIGURE 12 : LES TROIS ETAPES D'ARCHONTE [150].	78
FIGURE 13 : LA RELATION ENTRE LES MODÈLES VOULUS D'UNE ONTOLOGIE ET LES MODÈLES DES AXIOMES DE L'ONTOLOGIE (BASÉE SUR [155]).	84
FIGURE 14 : APPROCHE POUR LA SÉLECTION DES CONCEPTS.	87
FIGURE 15 : CAS DE PNEUMONIE DANS MIMIC-III.	94
FIGURE 16 : LA CLASSE <i>INFECTIVE PNEUMONIA</i> AVEC <i>VIRAL PNEUMONIA</i> COMME EXEMPLE DE SOUS-CLASSE.	102
FIGURE 17 : LA CLASSE <i>SYMPTOM</i> AVEC <i>RESPIRATORY SYSTEM AND CHEST SYMPTOM</i> COMME EXEMPLE DE SOUS-CLASSE.	103
FIGURE 18 : REPRÉSENTATION DE LA CLASSE <i>DIAGNOSTIC PROCESS</i> DANS PROTÉGÉ.	104
FIGURE 19 : LES ENTITÉS FONDATIONNELLES DE PNADO.	104

# LISTE DES ACRONYMES, SIGLES ET ABRÉVIATIONS

AARC	American Association for Respiratory Care
AMMI	Association of Medical Microbiology and Infectious Disease
ATS	American Thoracic Society
BFO	Basic Formal Ontology
BTS	British Thoracic Society
CAP	Collège des Pathologistes Américains
CIDO	Coronavirus Infectious Disease Ontology
CIM	Classification internationale des maladies
CMA	Canadian Medical Association
CMQ	Collège des médecins du Québec
CPG	Clinical Practice Guideline
CPRO	Computer Based Record Ontology
CSSO	Clinical Signs and Symptoms Ontology
CSRT	Canadian Society of Respiratory Therapists
CUI	Concept Unique Identifier
DOID	Human Disease Ontology
DOLCE	Descriptive Ontology for Linguistic and Cognitive Engineering
DIAGONT	Diagnostic Ontology
EBM	Evidence-based medicine
ESCMID	European Society of Clinical Microbiology and Infectious Diseases
HPO	Human Phenotype Ontology
GBPC	Guide de bonnes pratiques cliniques
GFO	General Formal Ontology
ICD-10	International classification of diseases
ICIS	Institut Canadien d'Information sur la Santé
IDO	Infectious Disease Ontology
IDSA	Infectious Diseases Society of America
INESSS	Institut National d'Excellence en Santé et en Services Sociaux
Inserm	Institut national de la santé et de la recherche médicale
IOM	Institute of Medicine
LODE	Live OWL Documentation Development Environment
LOINC	Logical Observation Identifiers Names and Codes
MeSH	Medical Subject Controlled Vocabulary
MO	Modèle Omis
NCBO	National Center for Biomedical Ontology
NGC	National Guideline Clearinghouse
NICE	National Institute for Health and Clinical Excellence

NLM	National Library of Medicine
OBO	Open Biomedical Ontologies
OGMS	Ontology for General Medical Science
OMS	Organisation mondiale de la santé
OWL	Web Ontology Language
PNADO	Pneumonia Diagnostic Ontology
QC	Question de compétence
QRC	Question de résolution de conflit
RadLex	Radiological Lexicon
RDF	Resource Description Framework
RO	Relation Ontology
RPC	Recommandations pour la pratique clinique
RTF	Relative Term Frequency
SADC	Système d'aide à la décision clinique
SEKMED	Software for the Evolution of Knowledge in MEDicine
SKOS	Simple Knowledge Organization System
SNOMED-CT	Systematized Nomenclature of Medicine-Clinical Terms
SNOMED-RT	Systematized Nomenclature of Medicine-Reference Terminology
SUMO	Suggested Upper Merged Ontology
SUP	Modèle superflu
SYMP	Symptom Ontology
TAL	Traitement Automatique de la langue
TF-IDF	Term Frequency – Inverse Document Frequency
UMLS	Unified Medical Language System
URI	Uniform Resource Identifier

# INTRODUCTION GÉNÉRALE

Quoique le terme ontologie ait été défini pour la première fois par Aristote, au 4<sup>e</sup> siècle avant Jésus-Christ, son utilisation n'a connu un grand essor que durant ces trois dernières décennies. Son utilité comme moyen de compréhension, de partage, de réutilisation et d'intégration de connaissances est reconnue dans plusieurs domaines y compris la médecine. En effet, nous constatons que le développement d'ontologies médicales est prolifique, et elles sont généralement prédestinées, entre autres, à aider les professionnels de la santé dans le processus décisionnel durant leurs pratiques cliniques. Elles apparaissent aujourd'hui comme des composants logiciels apportant une dimension sémantique et un raisonnement aux systèmes informatiques dans lesquels elles s'insèrent.

L'ingénierie ontologique qui, comme la réingénierie des connaissances, est née pour répondre aux besoins de représentation des connaissances. Elle réfère à l'ensemble des activités qui concernent le processus de développement des ontologies, leur cycle de vie, les méthodologies, les outils et les langages de leur construction [1]. Cette ingénierie représente un domaine multidisciplinaire qui demande à la fois une expertise du domaine pour lequel l'ontologie serait construite et une expertise en informatique.

Les méthodologies de construction des ontologies qui varient entre la construction à partir du texte ou en réutilisant les ontologies et les ressources terminologiques existantes ne guident pas assez l'ontologue tout au long du processus de construction. De surcroît, elles nécessitent une forte implication des experts du domaine et il est très difficile de réussir à constituer une telle équipe.

Ce travail de recherche s'inscrit dans le domaine de l'ingénierie ontologique et contribue avec une nouvelle méthodologie de construction d'une ontologie médicale qui s'adresse à un ontologue ou un ingénieur des connaissances. Cette méthodologie détaillée s'inspire des méthodologies existantes, guide l'ontologue dans toutes les étapes et réduit l'implication des experts du domaine. Ainsi, nous construisons une ontologie qui couvre le diagnostic de la pneumonie que nous avons appelé PNADO. PNA est l'abréviation médicale de pneumonia, DO est utilisé pour Diagnostic Ontology, en suivant la méthodologie pour la valider.

Ce présent document se structure en cinq chapitres. Le premier chapitre présente la problématique de la thèse ; celle-ci est subdivisée en quatre sous-sections qui abordent dans l'ordre : les difficultés face à la gestion des connaissances et de la prise de décision médicales, les méthodologies de construction d'ontologies, et enfin, la pneumonie et les erreurs dans le processus de diagnostic. Le deuxième chapitre fait une revue de littérature et couvre les ontologies, la réutilisation des ontologies entre théorie et pratique, les répertoires d'ontologies : prolifération et chevauchement, questions de compétence et guides de bonnes pratiques cliniques. Le troisième chapitre présente, quant à lui, nos objectifs de recherche. Le quatrième chapitre explique de manière détaillée notre méthodologie de construction d'une ontologie médicale, ainsi que l'application de la méthodologie sur le développement d'une ontologie couvrant le diagnostic de la pneumonie. Enfin, le cinquième chapitre présente les résultats et leur discussion. La fin de la thèse est consacrée aux conclusions qui rappellent la problématique, les principales contributions et réalisations ainsi que les perspectives de recherche.

**Remarque :** Les concepts issus des ontologies seront écrits en italique pour les distinguer du reste des concepts dans cette thèse.

# CHAPITRE 1 PROBLÉMATIQUE

Ce chapitre présente notre problématique générale de recherche. La section 1.1 permet de mieux comprendre les difficultés face à la gestion des connaissances et la prise de décision ainsi que les processus décisionnels relatifs à la pratique médicale. La section 1.2 porte sur les méthodologies de construction des ontologies médicales existantes et leurs déficiences. La section 1.3 s'attarde aux erreurs de diagnostic de la maladie de la pneumonie. Enfin, la section 1.4 conclut le présent chapitre.

## 1.1 Difficultés face à la gestion des connaissances et de la prise de décision médicales

L'accroissement continu du volume des connaissances pratiques et théoriques dans le domaine médical font en sorte qu'il peut être difficile pour les médecins généralistes ou spécialistes de maintenir à jour leurs connaissances et de les appliquer dans leurs pratiques cliniques. Pour ce faire, ils sont contraints de consulter entre autres plusieurs sources d'information externes, traditionnellement les pairs, les livres classiques d'enseignement, les publications scientifiques relevant de la médecine fondée sur des preuves et des ressources médicales reconnues par les membres de la communauté de praticiens ainsi que par les organismes accréditeurs et normatifs. Néanmoins, il existe des lacunes quant à l'application de ces sources d'information en pratique clinique. En dépit de la diffusion en ligne des connaissances médicales qui permettent aux médecins de faire des choix plus judicieux et d'assurer des soins de qualité, la recherche de solutions au problème clinique spécifique du patient est une tâche difficile [2, 3].

### 1.1.1 Processus décisionnels relatifs à la pratique médicale

Les connaissances professionnelles et expérientielles des médecins sont des facteurs importants pouvant guider et améliorer le processus de prise de décision selon la situation clinique précise des patients. La prise de décision clinique est à la fois multidimensionnelle et multifactorielle où les meilleures données probantes, les pratiques exemplaires promues par organisations savantes ne sont pas toujours utilisées, et certains patients ne reçoivent pas les traitements les plus appropriés [4].

La recherche de preuves scientifiques est une tâche qui nécessite un effort et un travail de revue et de synthèse importants. Ces dernières années, nous assistons de plus en plus à une production prolifique de publications scientifiques et cette augmentation n'est plus compatible avec la sphère de la vie quotidienne des médecins. En effet, une recherche d'un concept clinique sur PubMed<sup>2</sup> peut rassembler des milliers de citations et de résumés d'articles scientifiques. Bien que des moyens sont mis en place pour mieux cibler les thématiques, le volume d'articles scientifiques reste abondant. Un travail sérieux de tri et de synthèse s'impose d'une part, pour mieux servir les médecins et assurer leur formation continue, et d'autre part, pour améliorer la qualité des soins. C'est dans cet objectif que sont développés les guides de bonnes pratiques cliniques (GBPC). Néanmoins, l'état des lieux concernant l'implémentation des GBPC en milieu hospitalier est relativement faible [5-9]. Dès lors, la question qui s'impose est de savoir pourquoi les médecins ont des difficultés à respecter les GBPC ? La multitude des GBPC qui peuvent se retrouver dans plusieurs répertoires et sont développés par plusieurs organismes accréditeurs et normatifs peut constituer une barrière. En effet, il existe de multiples répertoires dans différentes langues telles que Cochrane<sup>3</sup>, NICE<sup>4</sup>, INESS<sup>5</sup>, IDSA<sup>6</sup>, ATS<sup>7</sup>, etc. D'autres barrières liées à l'implémentation des GBPC et le manque d'adhérence des médecins seront abordées dans la section 2.6.

Le raisonnement clinique est défini comme étant un processus cognitif hautement complexe [10]. Il permet au médecin d'évaluer et de gérer les problèmes de santé des patients. Ce processus de résolution de problèmes « est une activité intellectuelle selon laquelle un clinicien synthétise l'information obtenue dans une situation clinique, l'intègre à ses connaissances et à ses expériences antérieures dans le but de les utiliser pour prendre une décision diagnostique et de prise en charge » [10], p.236.

Dans les processus décisionnels relatifs à la pratique médicale, nous retrouvons dans la littérature académique les processus reconnus suivants : évaluation médicale, diagnostic,

---

<sup>2</sup> <https://www.ncbi.nlm.nih.gov/pubmed>, dernier accès le 17/03/2021.

<sup>3</sup> <https://www.cochranelibrary.com/>, dernier accès le 17/03/2021.

<sup>4</sup> <https://www.nice.org.uk/guidance>, dernier accès le 17/03/2021.

<sup>5</sup> <https://www.inesss.qc.ca/>, dernier accès le 17/03/2021.

<sup>6</sup> <https://www.idsociety.org/>, dernier accès le 17/03/2021.

<sup>7</sup> <https://www.thoracic.org/>, dernier accès le 17/03/2021.

intervention et évaluation continue. Ceux-ci seront brièvement abordés dans les prochaines sous-sections.

### **1.1.1.1 Processus d'évaluation médicale**

Ce processus concerne l'évaluation initiale de l'état du patient par le clinicien. Il consiste à consulter son historique clinique, l'interviewer dans le but de recenser le maximum d'informations sur ses symptômes et antécédents familiaux, et d'effectuer l'examen physique complet [11].

### **1.1.1.2 Processus de diagnostic**

Lors de ce processus, le médecin recommande un certain nombre de tests de laboratoire ainsi que des examens pour mieux appréhender le problème du patient et confirmer ou infirmer ses hypothèses. Le médecin peut se référer à ses collègues dans le but d'accumuler de l'information pouvant être pertinente pour le problème de santé du patient [12].

### **1.1.1.3 Processus d'intervention**

Dans ce processus, le médecin doit établir un plan d'intervention médicale pour chaque problème de santé. Selon le guide du collège des médecins du Québec (CMQ) sur la pratique médicale en soins de longue durée, ce plan désigne les investigations et les traitements nécessaires en lien avec le niveau d'intervention et les souhaits du patient. Un niveau d'intervention désigne la nature des soins médicaux à dispenser et il existe quatre niveaux : 1) correction d'un état altéré par tout moyen disponible ; 2) correction de toute détérioration possiblement réversible ; 3) correction des pathologies réversibles et contrôle des symptômes par des moyens diagnostiques et thérapeutiques ne causant pas d'inconfort ; et 4) soins palliatifs [11].

Également, dans le processus d'intervention, nous retrouvons le processus de prescription rationnelle qui met l'accent sur la façon structurée de choisir un médicament. Selon l'Organisation mondiale de la santé (OMS), il y a six étapes dans ce processus : 1) définir le problème du patient ; 2) spécifier l'objectif thérapeutique ; 3) vérifier l'adéquation du médicament ; 4) commencer le traitement ; 5) donner des renseignements, des instructions et avertissements ; et 6) superviser le traitement [13].

### 1.1.1.4 Processus d'évaluation continue

Enfin, ce processus consiste à faire le suivi du patient selon son état de santé en demandant de nouveaux tests de laboratoire et examens [11].

## 1.2 Construction d'ontologies médicales

Pour mieux exploiter les connaissances d'un domaine, elles sont organisées, modélisées et représentées dans un format. Celui-ci peut être un thésaurus, un réseau sémantique, une terminologie, une classification, une taxonomie, une ontologie, etc. Dans notre travail, nous nous intéressons à l'ontologie.

Le processus de construction d'une ontologie se compose de plusieurs étapes. Chacune nécessite un ensemble d'instructions et de techniques détaillées sur la manière de la réaliser, et ensemble, elles forment une méthodologie de construction d'ontologies. Les ontologies peuvent être construites *ex nihilo*, c'est-à-dire à partir de textes et/ou en réutilisant des éléments d'ontologies existantes. Comme nous allons le voir à la section 2.2 du chapitre 2, plusieurs méthodologies de construction *ex nihilo* ont été proposées. Leur principal problème est qu'elles ne fournissent pas assez d'indications sur la manière dont les différentes étapes doivent être réalisées. Elles n'indiquent pas les problèmes qui peuvent être rencontrés à chaque étape et comment ces problèmes peuvent être résolus. Cela rend les méthodologies citées très difficiles à utiliser, longues, sujettes aux erreurs et coûteuses.

La construction d'ontologies en réutilisant celles qui existent a démontré une efficacité en matière de gain de temps et d'efforts, car l'ontologue se focalisera uniquement sur le nouveau contenu spécifique à son domaine d'ontologie plutôt que de créer et de gérer du contenu déjà créé et validé. De plus, cette construction empêche la prolifération aléatoire des ontologies. Néanmoins, le grand nombre d'ontologies existantes rend le choix difficile. En outre, la réutilisation d'une ontologie présente l'inconvénient de reproduire et de dupliquer les erreurs de l'ontologie parente dans celle qui est résultante. Les méthodologies actuelles présentent un manque de principes directeurs pour réutiliser les ontologies et résoudre les problèmes qui sont susceptibles d'être rencontrés pendant tout le processus [14, 15]. En effet, bien que la réutilisation soit fortement préconisée lors de la construction de nouvelles ontologies [16], il n'existe que peu d'outils qui soutiennent ce processus.

Nous constatons que les ontologies médicales sont en hausse, mais leur qualité est très discutable. Certaines d'entre elles sont appuyées par des articles scientifiques qui fournissent quelques détails concernant leur construction. Néanmoins, très peu de méthodologies ont été suivies. Voici quelques exemples de travaux récents qui discutent la construction d'ontologies médicales publiées [17-23].

La construction d'ontologies constitue un enjeu important aussi bien pour les ontologues que pour les experts du domaine. Quelle que soit la méthodologie adoptée, le processus de construction doit faire l'objet d'une collaboration qui réunit non seulement les experts du domaine concerné par la modélisation, mais aussi, les ontologues [24]. L'absence d'un tel effort interdisciplinaire se traduit par la faible qualité de l'ontologie qui en résulte. Comme dans le génie logiciel, une ontologie est conçue pour répondre à des besoins exprimés par des experts du domaine qui peuvent être des utilisateurs finaux.

Une ontologie de haute qualité nécessite donc une approche d'ingénierie rigoureuse et systématique qui guidera l'ontologue durant tout au long du processus.

### **1.3 Pneumonie et erreurs dans le processus de diagnostic**

La pneumonie est parmi les maladies concernées par les erreurs de diagnostic [12]. Parmi les 190 cas identifiés, 7% sont liées à la pneumonie, 7 % à l'insuffisance cardiaque congestive décompensée, 5 % à l'insuffisance rénale aiguë, 5% au cancer, et 5% à l'infection des voies urinaires [25]. Brendish et ses collègues (2019) affirment que parmi les 720 patients atteints des maladies respiratoires aiguës, 28,2% d'entre eux qui ont été diagnostiqués de la pneumonie n'avaient aucune preuve radiologique de pneumonie, et que 34,9% des patients présentant des preuves clinico-radiologiques de pneumonie n'avaient pas eu de diagnostic de pneumonie [26]. D'autres résultats d'une étude démontrent que plusieurs patients ont reçu de mauvais diagnostics de pneumonie acquise en communauté jusqu'à la troisième ou à la quatrième consultation [27]. Dans une autre étude, au total, 669 cas ont été rapportés par 310 cliniciens de 22 institutions aux États-Unis et parmi ces cas, 583 ont été retenus pour des erreurs de diagnostic. Plus précisément, 18 cas (1,7 %) d'erreurs de diagnostic étaient liés à la pneumonie, 26 cas (4,5%) à l'embolie pulmonaire, 26 cas (4,5%) aux réactions médicamenteuses ou surdosage, 23 cas (3,9%) au cancer des poumons, 19 cas (3,3%) au cancer colorectal, 18 cas (3,1%) au syndrome coronarien aigu et 18 cas (3,1%) au cancer du

sein. D'après les auteurs, ces erreurs sont survenues lors de la phase de tests (défaut de commande, rapport et suivi des résultats de laboratoire, 44%), les évaluations des cliniciens (erreur de diagnostic, 32%), la prise d'histoire (10%), l'examen physique (10%) et les retards de référence ou de consultation (3%) [28]. Selon certains auteurs, pour éviter que des erreurs semblables se reproduisent, il serait utile de créer des approches systématiques et de supports décisionnels automatisés pour faciliter le rappel, l'apprentissage et le partage de tels cas [29, 30].

Kubilay et ses collègues (2016) affirment qu'il est difficile de distinguer entre la pneumonie, la maladie pulmonaire sous-jacente ou les infections avec des complications pulmonaires. En effet, la faible sensibilité<sup>8</sup> et spécificité<sup>9</sup> des critères cliniques, des résultats radiologiques et de la culture microbiologique représentent un défi majeur pour le diagnostic de la pneumonie associée aux soins de santé [31]. Une autre étude montre que 27 % de toutes les infections nosocomiales aux États-Unis sont dues à la pneumonie, et que 86% d'entre elles sont associées à la ventilation. Selon Kubilay et ses collègues (2016), bien que les radiographies thoraciques et la perspicacité clinique soient raisonnablement sensibles à la détection de la pathologie pulmonaire anormale sous-jacente, elles ne peuvent pas différencier de façon fiable la pneumonie des autres pathologies pulmonaires telles que l'insuffisance cardiaque chronique, la maladie pulmonaire interstitielle ou l'atélectasie. Les auteurs affirment aussi que le diagnostic clinique de la pneumonie chez les patients hospitalisés est fréquemment erroné [31-33].

Les membres du comité sur l'erreur diagnostique dans les soins de santé (2016) ont émis plusieurs recommandations pour améliorer l'utilité de la technologie de l'information dans le processus de diagnostic. Le rapport recommande le développement de nouvelles approches pour contrôler ce processus ainsi que pour identifier, apprendre et réduire les erreurs [34]. Ces approches incluent entre autres les systèmes d'aide à la décision clinique (SADC). L'une des approches utilisées par les développeurs pour améliorer ces SADC consiste à modéliser le raisonnement clinique à l'aide des ontologies pour simuler les processus décisionnels des

---

<sup>8</sup> La sensibilité indique la probabilité que le test diagnostiquera correctement un cas, ou la probabilité qu'un cas donné sera identifié par le test.

<sup>9</sup> La spécificité indique la probabilité qu'un test réalisé sur une personne saine se révèle négatif.

cliniciens [35]. D'ailleurs, plusieurs ontologies de diagnostic de maladies ont été intégrées dans ces systèmes pour aider les cliniciens dans leurs pratiques cliniques [36-38].

Pour autant que nous sachions, aucune ontologie ne couvre le diagnostic de la pneumonie. Certaines des ontologies médicales existantes, telles que DOID<sup>10</sup> (Human Disease Ontology) et HPO<sup>11</sup> (Human Phenotype Ontology), contiennent quelques concepts liés au diagnostic de la pneumonie. Cependant, elles ne couvrent pas entièrement les connaissances du domaine.

## 1.4 Conclusion

Les ontologies médicales se sont imposées ces deux dernières décennies pour diverses utilités. Elles sont considérées comme un excellent moyen de compréhension, de partage, de réutilisation et d'intégration des connaissances. Elles représentent aussi un outil intéressant pour supporter les systèmes à base de connaissances tels que les SADC. Les méthodologies de construction sont nombreuses, mais présentent des lacunes quant à l'orientation (guidance en anglais) de l'ontologue durant tout le processus, notamment lorsque la collaboration avec les experts du domaine fait défaut. Ce manque d'orientation se reflète dans la qualité de l'ontologie résultante.

Nous proposons une nouvelle méthodologie de construction d'ontologies médicales en fournissant le plus possible de directives pour mieux guider l'ontologue. Comme preuve de concept, le contexte d'application privilégié se retrouve dans le domaine de connaissances lié à la pneumonie. En l'occurrence, nous appliquons la méthodologie pour le développement d'une ontologie pour le diagnostic de la pneumonie (PNADO).

---

<sup>10</sup> <http://www.obofoundry.org/ontology/doid.html>, dernier accès le 17/03/2021.

<sup>11</sup> <http://www.obofoundry.org/ontology/hp.html>, dernier accès le 17/03/2021.

## CHAPITRE 2 REVUE DE LITTÉRATURE

Afin de répondre à nos questions soulevées au premier chapitre, nous présentons à la section 2.1 une revue de littérature sur les ontologies. À la section 2.2, nous exposons les différentes méthodologies de construction des ontologies. La section 2.3 est consacrée au concept de la réutilisation des ontologies et aux ontologies génériques. La section 2.4 présente un état des lieux concernant les répertoires d'ontologies, et enfin, la section 2.5, s'intéresse aux les guides de bonnes pratiques cliniques (GBPC).

### 2.1 Ontologies

Les ontologies ont une place importante dans la modélisation et la représentation des connaissances. Elles servent à les formaliser dans un domaine spécifique et ajoutent une couche sémantique aux systèmes et aux applications informatiques. Leur utilisation est largement répandue dans de nombreux domaines applicatifs dont la médecine. Cette section est consacrée aux ontologies. Plus précisément, leurs définitions sont présentées dans la section 2.1.1, la différence entre une ontologie et une terminologie dans la section 2.1.2, la distinction entre un mot, un terme, un concept et une notion dans la section 2.1.3, les composantes d'une ontologie dans la section 2.1.4, les ontologies et les logiques de description dans la section 2.1.5, le raisonnement dans la section 2.1.6, et les ressources terminologiques et ontologiques en médecine dans la section 2.1.7.

#### 2.1.1 Définitions

Le terme ontologie a été défini pour la première fois par Aristote, au 4<sup>e</sup> siècle avant Jésus-Christ, dans son livre de philosophie intitulé « Métaphysique » comme étant la science de l'Être en tant qu'Être, et les attributs qui lui appartiennent de par lui-même. Autrement dit, l'ontologie est la science de l'être ou l'étude des êtres en tant qu'ils existent [39]. Selon le philosophe, la signification fondamentale de l'être apparaîtrait dans la réponse à la question : « cette chose, *qu'est-ce ?* ». Néanmoins, le terme ontologie n'a fait sa réapparition qu'à l'orée des années 1600 (17<sup>e</sup> siècle) pour n'être utilisé que durant ces trois dernières décennies. En effet, le terme ontologie est utilisé dans l'ingénierie des connaissances dans le début des années 90 par des chercheurs comme Christopher P.Menzel, Richard J.Mayer, John F.Sowa et bien d'autres. Ce terme a été ensuite repris par Gruber [40] en 1993 qui lui donna

la première définition en introduisant la conceptualisation, « une ontologie est une spécification explicite d'une conceptualisation ». La définition de Gruber permet de nombreuses interprétations et demande à être clarifiée et précisée. Guarino et son collègue définissent dans leur travail la conceptualisation [41]. Il s'agit, selon eux, de 1) préciser la notion du domaine d'application en distinguant parmi les entités du domaine, les individus, les propriétés et les relations ; et 2) définir les propriétés selon les critères de rigidité, d'unicité, d'identité et de dépendance. La définition de Gruber a aussi été complétée par Borst qui définit une ontologie comme « une spécification formelle et explicite d'une conceptualisation partagée » [42]. Une définition qui a été, une année plus tard, appuyée et reprise par Studer et ses collègues [43].

L'utilisation des ontologies est devenue incontournable pour représenter et exploiter les données et les connaissances d'un domaine. Enfin, elle permet de représenter les concepts d'un domaine particulier et leurs relations ou leurs rôles ainsi que d'un sous-ensemble d'objets appartenant à ces concepts et reliés par des liens.

Une approche algébrique plus formelle [44] identifie une ontologie comme une paire  $(S, A)$ , où :

- S est le vocabulaire (ou la signature) de l'ontologie ;
- A est l'ensemble des axiomes ontologiques, qui spécifient l'interprétation voulue du vocabulaire dans un domaine discursif donné.

Une définition similaire est donnée [45] où la signature S est décomposée en quatre ensembles (pas nécessairement disjoints) : 1) l'ensemble des concepts ; 2) l'ensemble des relations (R) ; 3) l'ensemble des instances (I) ; et 4) l'ensemble des axiomes (A). Ainsi, une ontologie est définie comme un quadruple  $(C, R, I, A)$ .

### **2.1.1.1 Conceptualisation**

La notion de conceptualisation chez Gruber [40, 46] est reprise dans les travaux de Genesereth et Nilsson qui affirment [47] :

*« Un ensemble de connaissances formellement représentées est basé sur une conceptualisation : les objets, les concepts et autres entités qui sont supposés exister dans un domaine d'intérêt donné et les relations qui existent entre elles. Une conceptualisation est une vue abstraite et simplifiée du monde que nous souhaitons représenter dans un but quelconque », page 9.*

La conceptualisation selon Genesereth et Nilson est un tuple  $(D, R)$  où :

- D est un ensemble appelé l'univers du discours ;
- R est un ensemble de relations sur D.

### **2.1.1.2 Spécification formelle et explicite**

Comme mentionnée ci-haut, l'ontologie est une « spécification explicite d'une conceptualisation ». Une conceptualisation est explicitement spécifiée de deux manières : par extension et par intention. L'intention désigne la sémantique, c'est-à-dire l'ensemble des attributs et propriétés définissant un concept. L'ensemble des objets qu'englobe un concept est son extension.

Pour préciser une conceptualisation, il faudrait énumérer les extensions de chaque relation pour tous les mondes possibles. Toutefois, cela est impossible dans la plupart des cas ou du moins très peu pratique. Néanmoins, dans certains cas, il est logique de spécifier partiellement une conceptualisation de manière extensible, au moyen d'exemples, en énumérant les extensions des relations conceptuelles en correspondance avec des états du monde sélectionnés et stéréotypés. En général, une manière plus efficace de spécifier une conceptualisation est de fixer un langage que nous voulons utiliser pour en parler, et de limiter les interprétations d'un tel langage de manière intentionnelle, au moyen d'axiomes appropriés. En bref, une ontologie n'est qu'un ensemble de tels axiomes, c'est-à-dire une théorie logique conçue pour saisir les modèles voulus correspondant à une certaine conceptualisation et pour exclure les modèles non voulus. Le résultat sera une spécification approximative d'une conceptualisation. En fait les modèles les mieux intentionnés seront capturés et les modèles non intentionnels seront exclus.

### **2.1.1.3 Importance du terme 'partagé'**

Comme indiqué dans la section 2.1.1, la prise en compte de l'aspect « partagé » dans les ontologies revient à Brost [42] qui exigea que la conceptualisation exprime un point de vue commun entre plusieurs parties, un consensus plutôt qu'un point de vue individuel. Cette conceptualisation ne peut pas nécessairement être entièrement partagée. En effet, pour une utilisation pratique et efficace des ontologies, il s'est très rapidement avéré que sans un engagement ontologique partagé, au moins aussi minimal de la part des ontologues, les

avantages d'avoir une ontologie seraient limités. En d'autres termes, l'ontologie peut s'avérer inutile si elle est utilisée d'une manière qui va à l'encontre de l'engagement ontologique partagé. En conclusion, toute ontologie sera toujours moins complète et formelle qu'il ne serait souhaitable en théorie. En l'occurrence, il est important que les ontologies qui sont destinées à soutenir l'interopérabilité soient bien construites avec des primitives bien choisies et axiomatisées pour être comprises.

### **2.1.2 Ontologie et terminologie, quelle différence ?**

Toutes les définitions qui vont suivre sont données dans le cadre d'un travail terminologique.

La norme ISO 1087 :2019<sup>12</sup> définit la terminologie comme étant un ensemble des désignations et des concepts appartenant à un domaine ou à un sujet. Une désignation, toujours selon l'ISO 1087 :2019, est une représentation d'un concept par un signe qui le dénote dans un domaine ou sujet. Elle peut être un terme, incluant les appellations, un nom propre ou un symbole.

L'objectif d'une terminologie est de faciliter l'échange des connaissances dans une langue et d'une langue à l'autre, par un texte en langue naturelle. Elle est caractérisée par différentes propriétés linguistiques ou grammaticales selon l'usage prévu de cette terminologie.

L'ontologie, qui est définie comme une « spécification d'une conceptualisation », est avant tout « une description (comme une spécification formelle d'un programme) des concepts et des relations qui peuvent exister » [40].

### **2.1.3 Mot, terme, concept, notion**

Le mot constitue indubitablement un maillon essentiel de compréhension lors de l'expression orale ou écrite. C'est un élément de la langue constitué d'un groupe de lettres ou de sons formant une unité sémantique. Cette unité est isolée par deux blancs typographiques, des signes de ponctuation, d'autres caractères hors de l'alphabet (comme les parenthèses et les

---

<sup>12</sup> <https://www.iso.org/obp/ui#iso:std:iso:1087:ed-2:v1:fr>, dernier accès le 17/03/2021.

guillemets) à l'écrit ou une pause à l'oral. D'après l'ISO 1087:2019<sup>13</sup>, un mot est une chaîne de caractères ou une chaîne de bits traitée comme une unité pour un objectif donné.

Un concept, selon ISO 1087:2019, est « une unité de connaissance créée par une combinaison unique de caractéristiques ». Il permet de regrouper sous une même appellation les objets qui partagent des caractéristiques ou propriétés communes. Ces caractéristiques peuvent être essentielles pour comprendre le concept, par exemple : pour définir le concept *fièvre*, la caractéristique, *température élevée* est essentielle. Elles peuvent aussi être non essentielles, c'est-à-dire pas indispensables pour comprendre le concept comme *le rythme cardiaque* n'est pas essentiel. Elles peuvent être distinctives, c'est-à-dire utilisées pour distinguer un concept d'autres concepts associés, par exemple : le milieu de contamination d'une pneumonie distingue une *pneumonie hospitalière* d'une *acquise en communauté*. En réalité, un concept peut être général ou individuel. Un concept partagé correspond à un nombre potentiellement limité d'objets qui grâce à leurs propriétés partagées forment un groupe, par exemple : *maladie, milieu hospitalier, médicament*. Un concept individuel, quant à lui, fait référence à un objet unique, par exemple : *l'hôpital Hopkins, l'Université du Québec en Outaouais, le prix Nobel de médecine 2019*.

ISO précise que le concept dépend de l'influence du contexte socioculturel et il n'est pas nécessairement lié à une langue particulière.

Le terme, toujours d'après ISO 1087:2019, est une désignation qui représente un concept général par des moyens linguistiques, par exemple : *coronavirus 2 du syndrome respiratoire aigu sévère, gel hydroalcoolique à 70% d'alcool, chloroquine*. Un terme peut être simple, c'est-à-dire composé d'un seul mot, ou peut être divisé en éléments distincts, par exemple *postnatal*. Aussi, un terme peut être complexe, c'est-à-dire constitué de plusieurs mots ou unités lexicales, par exemple *coronavirus 2 du syndrome respiratoire aigu sévère*, ou alors multiterme, c'est-à-dire un terme complexe constitué de plusieurs mots, par exemple : *tomate cerise, chauve-souris*.

Bien qu'en apparence similaire et souvent employée comme des synonymes, notion et concept présentent bel et bien de subtiles différences quant à leur signification. En effet,

---

<sup>13</sup> <https://www.iso.org/obp/ui/#iso:std:iso:1087:ed-2:v1:fr>, dernier accès le 17/03/2021.

notion provient du latin *notio*, c'est-à-dire action d'apprendre à connaître. Selon le dictionnaire Le Robert, on parle de notion lorsqu'on se rapporte à une connaissance élémentaire. Exemple : avoir des notions sur le web sémantique. Il la définit aussi comme étant une connaissance intuitive, assez imprécise, par exemple : perdre la notion du temps.

## 2.1.4 Composantes d'une ontologie

Nous retrouvons de façon générale les composantes suivantes d'une ontologie :

- 1) **Les classes** sont appelées également concepts ou termes. Elles représentent un objet, une notion ou une idée. Elles peuvent décrire entre autres une tâche, une fonction, une action, et un processus de raisonnement [48]. Selon ce dernier, elles peuvent être classifiées selon plusieurs dimensions : (i) niveau d'abstraction (concrète ou abstraite) ; (ii) atomicité (élémentaire ou composée) ; et (iii) niveau de réalité (réelle ou fictive).
- 2) **Les relations** traduisent les associations existantes entre les concepts d'une ontologie. Elles aident à la compréhension du domaine modélisé et à l'élimination de certaines ambiguïtés linguistiques pouvant exister entre les concepts, et qui peuvent induire en erreur.
- 3) **Les attributs**, qui sont appelés aussi propriétés, servent à décrire les classes et les relations.
- 4) **Les fonctions** qui présentent des cas particuliers de relations dans lesquelles le nième élément de la relation est unique pour les  $n-1$  éléments précédents [49].
- 5) **Les axiomes ou les règles** sont des assertions sous une forme logique permettant de définir la signification de certains concepts ; de définir des restrictions sur des propriétés ; de vérifier la cohérence logique d'une ontologie ; et d'inférer de nouvelles connaissances. Il est possible d'utiliser la logique du premier ordre pour les décrire. Les axiomes sont les principaux éléments de base pour fixer l'interprétation sémantique des concepts et des relations.
- 6) **Les instances**, qui sont aussi appelées individus, sont utilisées pour représenter des objets concrets des composants de l'ontologie et par conséquent, du domaine relatif au problème posé.

## 2.1.5 Ontologies et logiques de description

Les logiques de description sont des langages dédiés à la représentation des connaissances d'un domaine par des entités qui ont une description syntaxique à laquelle est associée une sémantique formelle. Les logiques de description furent fortement influencées par les travaux sur la logique des prédicats, les schémas (très connus sous le nom de frames) [50] et les réseaux sémantiques [51]. La syntaxe des logiques de description permet de représenter la connaissance de manière structurée en matière de concepts et de rôles, qui correspondent respectivement à des prédicats unitaires (individus) et à des prédicats binaires (relations entre les individus).

Les logiques de description ont été reconnues comme un formalisme pour la représentation des connaissances d'un domaine particulier [52]. De ce formalisme est né OWL (Ontology Web Language), un langage qui dérive de DAML+OIL<sup>14</sup>. OWL offre une sémantique riche et formelle. Il est conçu comme une extension de Resource Description Framework (RDF) et RDF Schema (RDFS).

Les logiques de description sont utilisées, entre autres, dans le cadre du web sémantique et de l'ingénierie des connaissances pour la représentation, entre autres, des ontologies et la recherche de l'information et/ou l'inférence basée sur la logique.

Une base de connaissances (par exemple les ontologies) présentée par une logique de description contient deux composantes : T-Box et A-Box. La T-Box (T pour terminological) introduit le vocabulaire d'un domaine d'application composé de concepts et de rôles (terminologie). A-Box (A pour assertional) contient des assertions sur les individus par rapport aux concepts et rôles qui ont été déjà définis dans la T-Box.

---

<sup>14</sup> <http://www.daml.org/2001/03/daml+oil-index.html>, dernier accès le 17/03/2021.

## 2.1.6 Raisonnement

Le raisonnement consiste à inférer des connaissances implicites à partir de celles explicites stockées dans une base de connaissances. Les inférences se font sur les T-Box ainsi que sur les A-Box.

Concernant les inférences sur la T-Box [52], les tâches de raisonnement se composent de : 1) la vérification de satisfiabilité d'un concept, c'est-à-dire s'assurer qu'il peut contenir au moins un individu ; 2) la vérification de relation de subsumption entre deux concepts, c'est-à-dire déterminer les relations des sur concepts et sous-concepts; 3) la vérification d'équivalence entre deux concepts ; et 4) la vérification de disjonction entre deux concepts.

Les tâches de raisonnement sur une A-Box [52] reposent sur : 1) la vérification d'instance, c'est-à-dire vérifier si un individu donné est une instance d'un certain concept ; 2) la vérification de consistance, c'est-à-dire une A-Box donnée est consistante par rapport à une certaine T-Box s'il existe une interprétation qui est un modèle des deux, c'est-à-dire de A-Box et de T-Box.

Par exemple, les axiomes suivants de la syntaxe DL décrivent certaines connaissances sur la pneumonie infectieuse :

$$\text{Infective Pneumonia} \sqsubseteq \text{infectious disease} \quad (1)$$

$$\text{Infectious disease} \equiv \text{disease causedBy pathogen} \quad (2)$$

On trouve que la pneumonie infectieuse est une maladie causée par un pathogène.

Il existe plusieurs programmes de raisonnement appelés communément raisonneurs qui sont conçus pour comprendre les ontologies et réaliser des inférences. Plus encore, au cours de ces dernières décennies, l'optimisation des programmes de raisonnement de l'ontologie a été poursuivie de manière compétitive, et un certain nombre de raisonneurs hautement optimisés ont été développés. Nous citons FaCT [53], Pellet [54], RacerPro [55], Hermit [56], Konclude [57] et OnTop [58]. Cela dit, l'ontologie permet de représenter formellement les connaissances, et le raisonneur est capable d'en tirer le maximum.

Le concours nommé « OWL Reasoner Evaluation », qui a eu lieu en 2015, a compté 14 raisonneurs conformes à OWL 2 en compétition sur diverses tâches de raisonnement. Konclude s'est avéré être le meilleur, car il est très optimisé et efficace et les développeurs le testent en permanence par rapport à un vaste ensemble d'ontologies disponibles. Pellet et RacerPro se sont mérité la deuxième et la troisième position [59] .

## **2.1.7 Quelques ressources terminologiques et ontologiques en médecine**

Les prochains paragraphes porteront sur des exemples de ressources terminologiques et ontologiques dans le domaine médical, dont CIM, SNOMED, MeSH et LOINC.

### **2.1.7.1 CIM**

L'appellation complète de la CIM est Classification Statistique Internationale des Maladies et des Problèmes de Santé Connexes ou en anglais International Statistical Classification of Diseases and Related Health Problems. Elle permet le codage des maladies, des traumatismes et de l'ensemble des motifs de recours aux services de santé. Elle est publiée par l'Organisation mondiale de la santé (OMS)<sup>15</sup> et est utilisée à travers le monde pour enregistrer les causes de morbidité et de mortalité recueillies dans différents pays ou régions, à des fins diverses. La version actuelle est la CIM-10 publiée en 1994. CIM-11 est une nouvelle version qui a été adoptée par l'OMS en 2019 et qui rentrera en vigueur le 1<sup>er</sup> janvier 2022.

Son histoire a commencé avec la classification des causes de décès de Jacques Bertillon (1893). Cette dernière a connu cinq révisions jusqu'en 1938. À sa création en 1945, l'OMS se vit confier l'évolution de la classification de Bertillon qui devint en 1948, avec la sixième révision, la classification statistique internationale des maladies, traumatismes et causes de

---

<sup>15</sup> <http://www.who.int/classifications/icd/en/>, dernier accès le 17/03/2021.

décès. Elle cessait de ne répertorier que les causes de décès pour s'intéresser de façon plus générale à la morbidité, alors que la CIM-10 permet le recueil de diagnostics à des fins de santé publique ou d'évaluation de l'activité hospitalière pour le codage médico-économique des dossiers patients à des fins statistiques et budgétaires. Toute entité hiérarchique ou rubrique dans CIM-10 possède un et un seul père (sauf les entités du niveau 1 au sommet de la pyramide qui n'ont pas de père). À tout moment, il est possible pour toute entité hiérarchique donnée de reconstituer la liste exhaustive de tous ses ancêtres. La CIM-10 est divisée en 21 chapitres couvrant l'éventail complet des états morbides classés par appareil fonctionnel. Les chapitres sont toujours au niveau le plus élevé de la hiérarchie, ils sont divisés en groupes, eux-mêmes divisés en sous-groupes.

Selon certains auteurs, la CIM-10 est mal adaptée à la médecine de soins primaires, car il est difficile de coder des plaintes ou des états de santé non pathologiques fréquemment rencontrés en médecine générale. En l'occurrence, il s'agit d'une classification centrée sur la maladie et non sur le patient [60]. La version CIM-10 est en révision pour une nouvelle version, la CIM-11. La CIM-10-CA est une version élargie de la CIM-10 élaborée par l'ICIS (Institut Canadien d'Information sur la Santé) pour la classification de la morbidité au Canada. Elle contient 23 chapitres et englobe des problèmes et des situations qui ne sont pas des maladies, mais qui sont des risques pour la santé tels que les facteurs professionnels et environnementaux, le mode de vie et les circonstances psychosociales [61].

### **2.1.7.2 SNOMED**

SNOMED est une terminologie clinique développée à l'origine par le Collège des Pathologistes Américains (CAP) en 1955. Elle est une des classifications médicales les plus complètes. Son objectif est de rendre les connaissances de soins de santé plus accessibles pour l'ensemble des spécialités médicales. Elle regroupe des concepts organisés en hiérarchies. Chacun est associé à un contenu sémantique unique qui est défini selon une logique formelle et qui lui attribue sa place dans la structure hiérarchique. SNOMED-CT est la terminologie multilingue en santé la plus compréhensible et précise dans le monde. Aussi, elle est le langage de santé commun dans cinquante pays et elle a été développée grâce à une forte collaboration pour assurer qu'elle réponde aux besoins ainsi qu'aux attentes des professionnels de la santé [62].

Un concept dans SNOMED peut être décrit de différentes façons et rien n'empêche de créer par combinaison des concepts inconsistants [63]. Ce modèle pose encore des problèmes, par exemple les termes des différents axes ne sont pas complètement indépendants entre eux, l'axe « maladie » fait souvent double emploi, et certains concepts peuvent apparaître dans plusieurs axes.

Pour remédier à ce défaut, la SNOMED a évolué en SNOMED-RT (RT pour Reference Terminology), puis en SNOMED-CT (CT pour Clinical Terms), fusion de la SNOMED-RT et d'une terminologie britannique Clinical Terms de la NHS (services de santé britanniques). Dans cette nouvelle configuration, elle suit un certain nombre de principes : structure hiérarchique de concepts, définitions de types ou de rôles pour des concepts, consistance, exploitation dans le cadre d'une logique de description qui en font une ontologie formelle [64].

SNOMED est constituée principalement de trois types de composantes [65] :

- 1) **Les concepts de SNOMED-CT** représentent un contenu médical s'étendant de |abcès| à |zygote|. Chaque concept est muni d'un identifiant numérique unique. Au sein de chaque hiérarchie, les concepts sont organisés du plus général au plus détaillé, constituant des arborescences. Cette structure permet d'enregistrer des données médicales cliniques, puis de les regrouper ultérieurement à un niveau plus générique ;
- 2) **Les descriptifs de SNOMED-CT** associent aux concepts des définitions en langage courant. Un concept peut être associé à différents descriptifs, chacun d'eux constituant un synonyme décrivant le même concept médical. Toute traduction d'un concept comporte une série de descriptifs qui associe les définitions dans une autre langue au même concept. Chaque descriptif est pourvu d'un identifiant numérique unique ;
- 3) **Les relations de SNOMED-CT** relient les concepts à d'autres apparentés d'une manière ou d'une autre. Ces relations précisent la définition formelle et d'autres particularités d'un concept. Un type de relation est la relation |est un(e)| qui relie un concept à d'autres, plus génériques. Ces relations |est un(e)| mises bout à bout déterminent la hiérarchie des concepts.

SNOMED-CT contient plus de 500 000 concepts couvrant 19 axes [66]. Chacun recense les termes d'un sous-domaine de la médecine et il est hiérarchisé en fonction de la spécialisation

des concepts qui sont reliés par des relations d'hyponymie et d'hyperonymie. Il est important de souligner que SNOMED existe sous différentes variantes linguistiques.

### **2.1.7.3 MeSH**

MeSH<sup>16</sup> (Medical Subject Headings) a été conçu à la NLM (National Library of Medicine) aux États-Unis. La première version, publiée en 1960, a été un support de l'Index Medicus, répertoire des principales publications scientifiques et est utilisé par les systèmes de recherche bibliographique Medlars et MEDLINE<sup>17</sup>. C'est un thésaurus biomédical de référence et un outil d'indexation, de catalogage et d'interrogation des bases de données. Les descripteurs MeSH sont disposés dans une structure à la fois alphabétique et hiérarchique. Depuis 2016, il possède 16 niveaux hiérarchiques avec des relations de spécialisation-généralisation et réparti en 16 arborescences thématiques auxquelles correspond un code spécifique : « A » pour « anatomie », « B » pour « organisme », « C » pour « maladie », on retrouve 27 883 descripteurs avec plus de 87 028 mots clés. En plus de cela, plus de 230 872 concepts supplémentaires se trouvent dans un fichier additionnel. Ces enregistrements touchent à des exemples spécifiques concernant des protocoles des médicaments, des maladies et des produits chimiques.

Chaque année, afin de l'enrichir et de tenir compte de l'évolution du domaine biomédical, de nouveaux descripteurs sont ajoutés et d'anciens sont modifiés, remplacés, supprimés ou déplacés. MeSH est traduit en français par l'Inserm<sup>18</sup>.

### **2.1.7.4 LOINC**

LOINC<sup>19</sup> (Logical Observation Identifiers Names and Codes) est une nomenclature qui permet au travers d'un identifiant universel de codifier précisément une analyse de laboratoire médicale afin de faciliter les échanges électroniques. La plupart des systèmes de laboratoire et cliniques envoient actuellement des données en utilisant le standard de messages HL7 version 2.

---

<sup>16</sup> <https://www.nlm.nih.gov/pubs/factsheets/mesh.html>, dernier accès le 17/03/2021.

<sup>17</sup> <https://pubmed.ncbi.nlm.nih.gov/>, dernier accès le 17/03/2021.

<sup>18</sup> <http://mesh.inserm.fr/FrenchMesh/>, dernier accès le 17/03/2021.

<sup>19</sup> <https://loinc.org/>, dernier accès le 17/03/2021.

Le but de LOINC est de faciliter l'échange électronique et la collecte des résultats cliniques tels que les tests de laboratoire et les observations cliniques. Il comprend la composante laboratoire qui couvre tout ce qui peut être testé, mesuré ou observé dans un spécimen. Il contient les catégories courantes de chimie, d'hématologie, de sérologie, de microbiologie et la toxicologie ainsi que la composante LOINC clinique qui contient un sous-domaine d'une ontologie de documents et une classification des attributs clés des documents cliniques.

Les codes LOINC distinguent une observation donnée (test demandé/rapporté, question d'enquête, document clinique) à travers les six dimensions suivantes : 1) composante qui est la substance ou l'entité étant mesurée ou observée ; 2) propriété qui est la caractéristique de la composante ; 3) temps qui indique l'intervalle de temps sur lequel une observation a été faite ; 4) système qui est le spécimen ou la chose sur lequel l'observation a été faite ; 5) échelle qui indique la manière avec laquelle la valeur de l'observation a été quantifiée ; et 6) méthode qui est une classification de haut niveau de la manière avec laquelle l'observation a été faite. Elle est uniquement nécessaire lorsque la technique affecte l'interprétation clinique des résultats.

## **2.1.8 Conclusion**

La représentation et la modélisation des connaissances avec les ontologies sont un processus qui tend à créer un modèle qui reflète notre perception d'un domaine. Elle peut être définie comme l'ensemble des concepts liés au domaine ou de construire une représentation structurée et hiérarchisée de ces concepts avec toutes les relations qui leur sont associées. Une telle représentation riche et détaillée permet aux machines (i) de raisonner avec un haut niveau d'abstraction ; (ii) d'offrir des interprétations liées au domaine d'application concerné ; et (iii) d'éviter d'avoir des ambiguïtés. Son application dans le domaine médical est de plus en plus courante.

Il existe des ontologies médicales comme SNOMED-CT et CIM-10 qui couvrent certains aspects du domaine de diagnostic de la pneumonie comme les types de pneumonie, mais leurs représentations ont des lacunes. Aussi, leur utilisation devient difficile, c'est-à-dire l'utilisateur sera obligé de naviguer dans plusieurs ontologies pour trouver ce qui l'intéresse et parfois, un concept peut avoir plusieurs représentations, ce qui mettra l'utilisateur dans l'embarras du choix. Encore, ces ontologies ne répondent pas aux requêtes de l'utilisateur comme les symptômes ou les signes cliniques de la pneumonie. Une seule ontologie a été construite pour le diagnostic de la pneumonie destinée à un SADC [67]. Néanmoins, elle

n'est pas disponible ni accessible pour une consultation et/ou une réutilisation, et les GBPC n'ont pas été pris en compte.

## **2.2 Construction des ontologies**

Il existe un ensemble de critères et de principes qui ont fait leurs preuves dans le processus de construction des ontologies. Gruber (1993) propose cinq critères qui sont les suivants : la clarté et l'objectivité, la complétude, la cohérence, l'extensibilité monotonique maximale et les engagements ontologiques minimaux [40]. Depuis les années 90, des méthodologies ont été proposées pour construire des ontologies qui satisfont les critères cités ci-dessus. Avant de les présenter dans la section 2.2.3, nous avons jugé utile d'aborder les questions de compétence qui sont inéluctables pour la construction d'une ontologie (section 2.2.1), ainsi que son cycle de vie (section 2.2.2).

### **2.2.1 Questions de compétence et leur rôle dans l'ingénierie des connaissances**

Les questions de compétence (QC) ont été introduites pour la première fois dans [68] pour l'ingénierie des ontologies. Ce sont des questions formulées en langage naturel auxquelles l'ontologie doit répondre. De ce fait, l'ontologie doit contenir un ensemble d'axiomes nécessaire et suffisant pour représenter et résoudre ces questions. Lorsqu'elles sont utilisées au début de la construction, elles permettent de bien définir pour quel objectif l'ontologie serait construite, et quelle direction de développement elle prendrait, et cela dès la première phase.

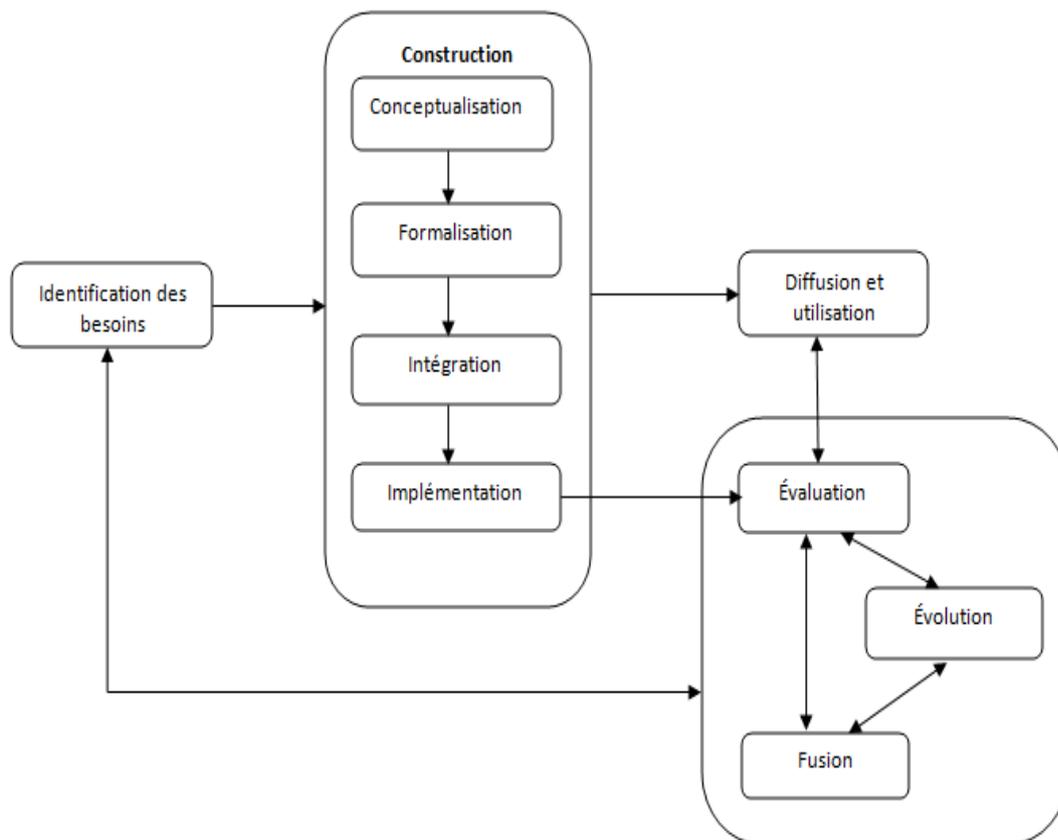
L'utilisation des QC n'est pas uniquement limitée à la première phase de construction, car elles peuvent aussi, entre autres être utilisées pour l'évaluation d'une ontologie. D'après Gomez-Perez [69], pour évaluer une ontologie, un cadre de référence est nécessaire et ce dernier peut être constitué de QC.

### **2.2.2 Cycle de vie d'une ontologie**

Les ontologies sont destinées à être utilisées comme des composants logiciels dans des systèmes répondant à des objectifs opérationnels. Leur développement doit s'appuyer sur les

mêmes principes que ceux appliqués en génie logiciel. En considérant une ontologie comme un objet technique évolutif, elle possède donc un cycle de vie qui doit être bien défini [70].

Le cycle de vie d'une ontologie rassemble plusieurs activités allant de l'identification des besoins pour la construction de l'ontologie jusqu'à sa maintenance [71]. La **Figure 1** présente les activités les plus importantes constituant le cycle de vie d'une ontologie que nous traiterons dans les prochains paragraphes, selon certains travaux [71-73].



**Figure 1** : Cycle de vie d'une ontologie [71-73].

L'identification des besoins est la phase la plus importante, celle qui permet d'identifier les besoins des utilisateurs et de cerner le domaine de connaissances. Lors de cette étape, des scénarios d'utilisation sont définis afin de justifier le développement de l'ontologie. La recherche des connaissances manquantes est dès lors effectuée afin de trouver un consensus nécessaire pour la phase de construction. Au terme de cette étape, un document de spécification de l'ontologie informel, semi-formel ou formel écrit en langage naturel est produit.

Lors de la phase de construction, l'ontologie est créée. Elle est subdivisée selon les quatre étapes suivantes :

- 1) **Étape 1** : La conceptualisation permet de structurer les connaissances du domaine dans un modèle conceptuel qui utilise le vocabulaire identifié dans l'étape de l'identification des besoins. Le but est de représenter le domaine de discours par des entités ontologiques qui sont les classes, les relations, les axiomes et les individus. Elle nécessite une bonne structuration des entités et un recours aux dictionnaires afin d'éviter toutes ambiguïtés liées aux termes utilisés et/ou aux domaines étudiés ;
- 2) **Étape 2** : La formalisation permet la représentation de la conceptualisation définie par un langage formel ;
- 3) **Étape 3** : L'intégration consiste à faire appel à des ontologies existantes et à les intégrer dans celle qui est créée dans le but de ne pas commencer à zéro, cela réduit l'effort et le coût de mise en œuvre ;
- 4) **Étape 4** : L'implémentation sert à créer l'ontologie physiquement. Le résultat doit être une ontologie codifiée qui doit refléter l'étape de l'identification des besoins. Selon certains auteurs, l'environnement de développement des ontologies doit avoir au minimum les éléments suivants [71] : un analyseur lexical et syntaxique pour garantir l'absence des erreurs (lexicales et syntaxiques) ; un traducteur pour assurer la portabilité de l'ontologie ; un éditeur pour ajouter, supprimer ou modifier des définitions; un navigateur pour inspecter la librairie des ontologies et leurs définitions ; un outil de recherche qui permet la recherche des définitions les plus appropriées ; un évaluateur pour détecter les incomplétudes, les inconsistances et les redondances ; un mainteneur automatique qui permet la suppression et la modification des définitions existantes.

La maintenance consiste à vérifier la qualité de l'ontologie construite, à l'adapter et à l'améliorer pour prendre en compte les nouveaux besoins. Cette phase englobe les quatre éléments suivants :

- 1) L'évaluation qui permet de juger et de vérifier la qualité de l'ontologie ainsi que sa capacité de répondre aux attentes des utilisateurs telles que définies lors de la première étape. Pour cela, elle utilise des tests et des requêtes, et se base essentiellement sur un ensemble de critères métriques tels que la clarté, l'accessibilité, la cohérence, la consistance de l'ontologie et la capacité d'inférence. À la fin de cette

étape, un document est livré par l'ontologue qui contient la description de la procédure suivie pour son évaluation, les techniques utilisées, les types d'erreurs trouvés et les sources des connaissances utilisées ;

- 2) L'évolution consiste à modifier la structure et/ou les instances de l'ontologie. Elle permet de la réviser et la faire évoluer pour répondre aux éventuels besoins des utilisateurs ;
- 3) La fusion consiste à créer une nouvelle ontologie en fusionnant deux ou plusieurs qui sont existantes. En effet, après la phase de construction et de vérification, l'ontologie peut faire partie du processus de construction d'autres ontologies que l'on nomme activité d'intégration ;
- 4) La diffusion et l'utilisation assurent le déploiement de l'ontologie et son utilisation par les utilisateurs finaux. Cette étape suit celle de la construction de l'ontologie.

### **2.2.3 Aperçu des méthodologies de construction manuelle**

Dans cette section, nous nous intéressons aux méthodologies de construction manuelle des ontologies qui ont réussi à s'imposer dans l'ingénierie des ontologies. Dans ce contexte, la méthodologie définit la stratégie en incluant un ensemble de méthodes, de principes et de règles à suivre pour la conception des ontologies. Elle indique ce qui doit être fait, à quel moment, et par qui, mais elle n'indique pas les aspects opérationnels concernant les technologies et les outils. La plupart des méthodologies proposées entament le processus de construction par l'identification des concepts, puis leurs organisations et leurs structurations des concepts et leurs relations. Néanmoins, les ontologies réalisées ayant le même objectif et construites par des ontologues variés sont différentes les unes des autres.

#### **2.2.3.1 Critères d'analyse**

Certains des critères d'analyse de chaque méthodologie sont repris des travaux de Fernandez Lopes [74] et ils s'inspirent de la norme IEEE 1074-1995 [75]. Cette dernière décrit le processus de développement de logiciels, les activités à mener et les techniques qui peuvent être utilisées pour développer des logiciels. La raison d'appliquer cette norme au développement des ontologies est qu'un logiciel est défini par le glossaire standard de l'IEEE sur la terminologie du génie logiciel [76] comme « computer programs, procedures, and possibly associated documentation and data pertaining to the operation of a computer

system » et les ontologies font partie des produits logiciels. Par conséquent, les ontologies devraient être construites selon les normes proposées pour les logiciels en général, avec certaines adaptations aux caractéristiques particulières des ontologies. Voici les cinq processus standards de l'IEEE qui doivent être appliqués dans une méthodologie de construction d'une ontologie : 1) modélisation du cycle de vie des logiciels ; 2) gestion de projet ; 3) processus orientés vers le développement de logiciels ; 4) développement ; et 5) post-développement.

Ces neuf critères d'analyse sont les suivants: 1) l'héritage de l'ingénierie de la connaissance ; 2) le détail de la méthodologie ; 3) les recommandations pour la formalisation des connaissances ; 4) la stratégie pour la construction d'ontologies ; 5) la stratégie pour l'identification des concepts ; 6) le cycle de vie recommandé ; 7) les différences entre la méthodologie et la norme IEEE 1074-1995 ; 8) les techniques recommandées ; et 9) les ontologies développées à l'aide de la méthodologie et les systèmes qui ont été mis en place en utilisant ces ontologies.

L'analyse des méthodologies, qui seront présentées ci-après, selon les critères 1, 3, 4, 6, et 9 n'est pas présentée dans cette thèse. Même si ces critères ne sont pas liés à notre problématique de recherche, ils ont été tout de même vérifiés pour chacune des méthodologies. Nous allons nous focaliser dans le cadre de cette thèse sur les critères 2, 5 et 8 auxquels nous rajoutons ces deux critères : 1) la nécessité d'implication des experts du domaine tout au long de la méthodologie et 2) la favorisation de la réutilisabilité et de l'interopérabilité.

### **2.2.3.2 Analyse des méthodologies**

#### **a) Méthodologie de Grüninger et Fox**

Il s'agit d'une méthodologie basée sur l'expérience de construction de l'ontologie dans le cadre du projet TOVE (Toronto Virtual Enterprise) dans le domaine des procédés commerciaux et la modélisation des activités de l'entreprise [77]. Selon les auteurs, le développement d'une ontologie doit être motivé par des problèmes qui se posent dans le domaine d'application. Cette méthodologie est inspirée des techniques de développement de systèmes à base de connaissances utilisant la logique de premier ordre. Le modèle n'est pas construit directement, c'est-à-dire qu'en premier lieu, une description informelle est établie

sur la base des spécifications de l'ontologie à construire, et ensuite, cette description est formalisée. Les étapes proposées sont les suivantes :

- **Spécification des scénarios** : Selon les auteurs, le développement d'une ontologie est motivé par des scénarios d'usage qui constituent des problèmes ou des exemples auxquels l'application qui va utiliser l'ontologie doit répondre. Ces scénarios peuvent donner une description informelle de la sémantique des objets et des relations à y inclure ;
- **Formulation informelle des questions de compétence** : Des questions auxquelles le système projeté est censé pouvoir répondre. L'ontologie doit être capable de représenter ces questions au moyen de sa terminologie et être capable d'y répondre en utilisant des axiomes et des définitions ;
- **Spécification de la terminologie dans la logique du premier ordre** : Les questions de compétences qui ont été préalablement identifiées vont déterminer une liste de termes qui va être utilisée pour spécifier une terminologie dans la logique du premier ordre ;
- **Formulation formelle des questions de compétence en utilisant la terminologie de l'ontologie** : Traduire les questions formulées précédemment dans un langage formel propre aux ontologies ;
- **Spécification des axiomes dans la logique du premier ordre** : Des axiomes exprimés dans la logique du premier ordre sont ajoutés à l'ontologie en suivant les questions de compétence.

Cette méthodologie a permis de développer des projets complexes dans le domaine de l'entreprise, mais son utilisation reste limitée vu le manque de précisions au niveau des étapes et techniques utilisées.

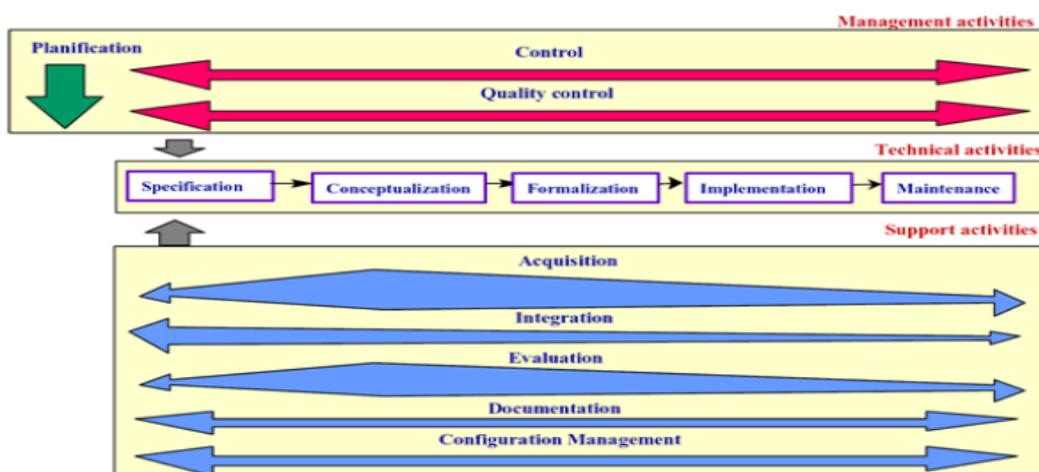
## **b) Méthodologie de Uschold et King**

Elle est basée sur l'expérience acquise lors du développement de l'ontologie intitulée « *Enterprise Ontology* ». Cette méthodologie fournit quatre principes pour développer des ontologies, dont (1) l'identification de l'objectif, c'est-à-dire définir clairement l'objectif, sa portée et les scénarios de son utilisation ; (2) la construction de l'ontologie qui est divisée en plusieurs étapes (l'identification des concepts et des relations clés du domaine concerné, la production de définitions précises non ambiguës de ces concepts et de ces relations,

l'identification des termes qui désignent ces concepts et ces relations, le codage de l'ontologie en la représentant explicitement dans un langage formel et l'intégration des ontologies existantes) ; (3) l'évaluation a pour objectif de produire des jugements techniques sur l'ontologie, son environnement logiciel et la documentation produite ; et (4) la documentation de l'ontologie, ayant pour but de faciliter sa réutilisation et son partage [78].

### c) METHONTOLOGY

Elle a été proposée en 1997 par l'équipe d'intelligence artificielle de l'Université Polytechnique de Madrid et elle permet de couvrir tout le cycle de vie d'une ontologie [71, 79]. Elle s'intéresse pratiquement à toutes les activités liées aux ontologies, c'est-à-dire aux activités de développement, de gestion de projet et de support. Elle adopte des techniques du génie logiciel dans son processus de développement (voir **Figure 2**). Elle est supportée par la plateforme ODE (Ontology Development Environment) qui a donné suite à WebODE. Nous retrouvons dans la littérature plusieurs ontologies qui ont été développées en utilisant cette méthodologie. Cette dernière est composée de sept étapes : (1) spécification qui comprend entre autres la définition du domaine et de la portée de l'ontologie ; (2) acquisition des connaissances à partir des ressources textuelles ; (3) conceptualisation qui consiste à structurer la connaissance du domaine dans un modèle conceptuel qui décrit le problème et sa solution en fonction du vocabulaire du domaine identifié dans l'étape de spécification de l'ontologie; (4) implémentation ; (5) intégration des ontologies existantes ; (6) évaluation ; et (7) documentation.



**Figure 2** : Processus de développement d'ontologie avec METHONTOLOGY [80].

#### **d) Ontology development 101**

Elle est proposée par l'Université Stanford [81]. Elle propose les sept recommandations suivantes : (1) déterminer le domaine et l'apport de l'ontologie en utilisant des questions de compétence (quel domaine va-t-elle couvrir ? Dans quel but serait-elle utilisée ? Qui sera son utilisateur ? Est-ce qu'elle sera utilisée et maintenue ?) ; (2) prendre en considération les ontologies existantes pour une éventuelle réutilisation ; (3) énumérer les termes importants du domaine ; (4) définir les classes et leur hiérarchie ; (5) définir les propriétés des classes et attributs ; (6) définir les facettes des attributs, et enfin (7) créer les instances.

#### **e) Méthodologie On-To-Knowledge**

Cette méthodologie a été développée par des chercheurs de l'université Karlsruhe, en 2001 [82]. Elle est constituée de six phases : (1) étude de faisabilité pour déterminer la faisabilité économique et technique du projet. Cette phase consiste aussi à choisir le domaine le plus prometteur et la meilleure solution pour les problèmes potentiels ; (2) lancement qui décrit le domaine de l'ontologie et de la portée, les applications qu'elle soutient, ses sources de connaissances (experts du domaine, organigrammes, plans d'entreprises, dictionnaires, etc.), et ses utilisateurs potentiels incluant les scénarios d'utilisation ; (3) affinement qui sert à rassembler tous les concepts pertinents, obtenir les connaissances des experts du domaine sur la base des concepts déjà obtenus auparavant et développer l'ontologie, et enfin traduire l'ontologie dans un langage de représentation formel ; (4) évaluation qui sert à vérifier si l'ontologie satisfait au document de spécification des exigences et si son utilisation auprès des utilisateurs finaux est satisfaisante ; et (5) maintenance dont l'objectif est d'entretenir l'ontologie et de la mettre à jour selon les nouveaux besoins. Cette méthodologie est principalement destinée aux entreprises.

#### **f) ARCHONTE**

Bachimont (2002) s'est basé sur la sémantique différentielle pour proposer la méthodologie ARCHONTE (ARCHitecture for ONTological Elaborating) [83]. Selon cette dernière, la construction d'une ontologie passe par trois étapes principales :

- 1) Choisir les termes pertinents du domaine et normaliser leurs sens, puis justifier la place de chaque concept dans la hiérarchie ontologique en précisant les relations de

similarités et de différences que chaque concept entretient avec ses concepts frères et son concept père ;

- 2) Formaliser les connaissances, ce qui implique par exemple d'ajouter des propriétés à des concepts, des axiomes et de contraindre les domaines d'une relation ;
- 3) Représenter l'ontologie dans un langage formel de représentation des connaissances.

### **g) OntoSpec**

Elle vise à apporter aux concepteurs une aide pour modéliser des connaissances ontologiques, en amont de leur représentation formelle [84]. Son objectif est d'aider l'ontologue à répondre à la question : qu'est-ce qu'une bonne définition de concept ou de relation ? La méthode repose sur un ensemble de principes de modélisation définis de façon rigoureuse. Son application conduit à élaborer une ontologie semi-informelle qui est indépendante de tout langage de représentation des connaissances. Cette méthode permet de définir les concepts et leurs relations comme des ensembles de propriétés structurés. Chaque propriété est ainsi présentée comme une proposition logique et des règles typographiques sont alors posées. Les définitions demeurent compréhensibles par chacun permettant à des experts des domaines concernés ou à de futurs utilisateurs de l'ontologie de coopérer avec le concepteur, en évaluant les choix de modélisation et la qualité des définitions produites. L'objectif d'OntoSpec est de prévoir à la fois une modélisation précise des connaissances de l'ontologie et une meilleure lisibilité.

Le **Tableau 1** présente l'analyse de chaque méthodologie selon les critères précédemment définis.

**Tableau 1** : Analyse des méthodologies de construction d'ontologies à partir de texte.

Méthodologie	Détail de la méthodologie	Stratégies pour l'identification des concepts	Techniques recommandées	Nécessité d'implication des experts du domaine	Favorisation de la réutilisabilité et de l'interopérabilité
Grüninger et Fox	Pas détaillée.	Une fois que les questions informelles de compétence ont été posées pour l'ontologie, les concepts sont spécifiés en utilisant la logique du premier ordre.	Vague, peu de recommandation. Par exemple, les techniques de formulation des questions de compétence ne sont pas mentionnées.	Les experts du domaine sont nécessaires.	La notion de réutilisation est complètement absente.
Ushold et King	Peu détaillée.	Les concepts clés sont établis en recherchant d'abord les plus importants (ceux qui sont liés à la portée de l'ontologie). Ensuite, les autres sont obtenus par généralisation et par spécialisation. On peut donc dire qu'une stratégie intermédiaire est utilisée pour identifier les concepts.	Vague. Par exemple, la méthodologie recommande que les concepts et relations clés dans le domaine étudié soient identifiés lors de l'acquisition. Cependant, aucun détail n'est donné sur la manière dont cela doit être fait, et seule une ligne directrice très vague, impliquant l'utilisation de techniques de remue-méninges, est donnée.	Les experts du domaine sont nécessaires.	Elle aborde la nécessité de l'intégration des ontologies existantes, mais sans fournir aucune indication sur la manière dont elle sera effectuée.
METHONTOLOGY	Peu détaillée.	Elle identifie d'abord les concepts les plus importants à partir du glossaire des termes, ensuite, elle en obtient d'autres par généralisation et par spécialisation.	Pas de recommandations. Par exemple, pour la construction de la taxonomie des concepts.	Les experts du domaine sont nécessaires.	La possibilité de réutilisation est brièvement évoquée dans l'étape de l'intégration sans fournir aucun détail sur la manière dont elle doit être faite.
Ontology development 101	Peu détaillée.	Aucune.	Peu de recommandations. Par exemple, le contenu du corpus de textes n'est pas indiqué et il n'y a aucune recommandation le concernant	Elle cite les experts du domaine quand il s'agit de l'utilisation de l'ontologie. Elle ne cite pas leur implication et suppose que l'ontologue a des connaissances du domaine à modéliser.	La méthodologie s'intéresse à la réutilisation des ontologies à l'étape (2), mais les auteurs considèrent qu'aucune ontologie n'existe et que la méthodologie est dédiée pour une construction ex nihilo c'est-à-dire, à partir de zéro.

Méthodologie	Détail de la méthodologie	Stratégies pour l'identification des concepts	Techniques recommandées	Nécessité d'implication des experts du domaine	Favorisation de la réutilisabilité et de l'interopérabilité
On-To-Knowledge	Pas détaillée.	Les concepts les plus importants sont identifiés, ensuite, la généralisation et la spécialisation sont utilisées pour identifier d'autres concepts.	Elle recommande la création du document de spécification des exigences de l'ontologie lors de l'étape du lancement. Ce document décrit ce que l'ontologie doit couvrir, le domaine concerné par la modélisation, etc. Ce document sera une référence à l'ontologue pour sa prise de décision concernant l'inclusion et l'exclusion	Ils sont indispensables pour la réalisation de la plupart des étapes de la méthodologie.	La réutilisation est citée brièvement.
ARCHONTE	Détaillée mais elle ne suit pas le cycle de vie d'une ontologie.	Elle utilise les principes différentiels (pour plus de détails, voir la section 4.3.1 du chapitre 4)	Peu de recommandations au niveau de la formalisation des connaissances. L'étape reste un peu vague.	Les experts du domaine sont nécessaires.	La réutilisation est citée brièvement.
OntoSpec	Détaillée.	Elle définit les concepts et leurs relations comme des ensembles de propriétés structurés. Pour chaque propriété, elle identifie son rôle par rapport à l'entité définie, en matière de conditions.	Distinguer les propriétés nécessaires des propriétés contingentes.	Les experts du domaine sont nécessaires.	La notion de réutilisation est complètement absente.

## 2.2.4 Aperçu des méthodologies de construction automatique des ontologies

La construction manuelle des ontologies est coûteuse en matière de temps et de ressources, ce qui a amené plusieurs chercheurs sur la manière d'alléger ce processus à travers l'automatisation de certaines étapes. C'est le cas des méthodes d'acquisition d'ontologie à partir de textes qui ont été largement utilisées ces quinze dernières années dans l'ingénierie ontologique [85-87]. Nous présentons dans ce qui suit quelques méthodologies de construction automatique et semi-automatique. Contrairement aux méthodologies de

construction manuelles, on ne va pas s'attarder sur leur analyse et on va se contenter de comprendre leur fonctionnement.

a) **TERMINAE** : Il s'agit d'une approche pour sélectionner les concepts, leurs propriétés, les relations et leur regroupement. Elle utilise des outils de traitement automatique des langues pour analyser les termes de textes et les relations lexicales. Les termes sont regroupés suivant leur contexte et facilitent la création de concepts et de relations sémantiques. Les concepts et relations sont ensuite formalisés dans un modèle [88].

b) **Text2Onto** : Il est un outil développé pour construire des ontologies à partir de textes de manière complètement automatique. Il est codé en java et est composé de modules qui extraient à partir des textes, des concepts, des relations entre eux (relation d'équivalence, hiérarchique, etc.) et des instances de concepts [89]. Chaque module peut combiner un ou plusieurs algorithmes : RTF (Relative Term Frequency), TF-IDF (Term Frequency – Inverse Document Frequency) [90], l'entropie, ou la méthode C-value/NC-value pour l'extraction des concepts [91], l'utilisation de WordNet<sup>20</sup> pour construire une hiérarchie et la définition de patrons lexico-syntaxiques pour l'extraction des relations de subsomption [92], utilisation de WordNet et définition des expressions régulières avec JAPE<sup>21</sup> pour l'extraction des relations partitives ; analyse syntaxique pour déterminer les relations sémantiques générales. Text2Onto utilise l'architecture GATE<sup>22</sup> pour prétraiter les textes. Les résultats sont dotés d'une mesure de confiance entre 0 et 1 obtenue à l'aide de différentes mesures combinables (TF-IDF, RTF, entropie). Cet outil possède une interface graphique utilisateur avec plusieurs volets et est indépendant du langage de formalisation utilisé. Il est intégré dans l'environnement d'ingénierie ontologique NeOn<sup>23</sup> [93]. Il est également disponible comme un plug-in de l'environnement de développement Eclipse<sup>24</sup> ce qui facilite son intégration dans d'autres éditeurs d'ontologies.

---

<sup>20</sup> <https://wordnet.princeton.edu/>, dernier accès le 17/03/2021.

<sup>21</sup> <https://gate.ac.uk/sale/tao/splitch8.html>, dernier accès le 17/03/2021.

<sup>22</sup> <https://gate.ac.uk/>, dernier accès le 17/03/2021.

<sup>23</sup> <http://neon-toolkit.org>, dernier accès le 17/03/2021.

<sup>24</sup> <http://www.eclipse.org/>, dernier accès le 17/03/2021.

**c) OntoLearn :** L'outil extrait des termes pertinents d'un domaine en combinant des outils linguistiques et statistiques dans le but de déterminer leur distribution dans le corpus. Il se sert de glossaires disponibles sur le Web. Pour découvrir les relations de subsomption entre concepts, il utilise des patrons lexico-syntaxiques décrits par des expressions régulières [94]. La structure interne des termes multimots est aussi utilisée pour extraire ce type de relations [85]. Il se sert également de la base de données lexicale WordNet à fin extraire des synonymes et d'autres types de relations. Comme on peut le constater, ces méthodologies de construction d'ontologies à partir de textes sont basées principalement sur des outils de traitement automatique des langues (TAL) et utilisent souvent des méthodes statistiques pour le filtrage des résultats. Une des limites principales qu'on peut noter est qu'elles peuvent fonctionner pour des tâches de construction où un corpus textuel suffisant est disponible pour faire fonctionner les outils de TAL.

**d) OntoGen :** Cet outil codé en langage .net implémente une approche semi-automatique pour la construction d'ontologies de thèmes à partir de collections de documents [95]. En effet, il construit l'ontologie en interagissant avec l'expert du domaine, il suggère à ce dernier des concepts sous la forme de classes de documents, il propose une dénotation et leur associe automatiquement des instances (les documents). Il permet de visualiser l'ontologie tout au long du processus de construction. OntoGen exploite des algorithmes de fouille de textes non supervisés dont k-means [96] et les techniques d'analyse sémantique latente LSI [97] pour suggérer des concepts, des relations entre ces concepts et des instances. L'utilisateur à travers l'interface peut interagir en ajoutant de nouveaux concepts ou de modifier ceux déjà existants.

**e) OntoLt :** C'est une approche qui fournit un plug-in pour l'outil de développement d'ontologie Protégé<sup>25</sup> [85]. Il extrait des concepts et des relations automatiquement à partir des collections de textes annotés en utilisant des règles linguistiques. Elle intègre l'analyse linguistique dans l'ingénierie ontologique et supporte l'acquisition semi-automatique et interactive d'ontologies à partir de textes, mais aussi, l'extension des ontologies existantes.

---

<sup>25</sup> <https://protege.stanford.edu/>, dernier accès le 17/03/2021.

## **2.2.5 Failles des méthodologies manuelles de construction des ontologies**

Les méthodologies de construction manuelle décrites dans la section 2.2.3 n'abordent pas la réutilisation des ontologies (ou brièvement), bien qu'elles la reconnaissent comme faisant partie du processus global de construction d'une ontologie. Ceci constitue en quelque sorte leur faiblesse. Nous avons aussi constaté que la plupart d'elles recommandent la participation des experts du domaine auprès des ontologues. Une chose qui n'est pas évidente notamment dans le domaine médical. D'ailleurs, ce sont des méthodologies conçues pour le domaine des entreprises sans prévoir d'autres domaines potentiels. Ces méthodologies présentent aussi un manque d'orientation de l'ontologue durant le processus de construction et en l'occurrence leur utilisation requiert une forte implication des experts du domaine. La seule méthodologie qui épargne les experts du domaine est ARCHONTE. Cette dernière contraint l'ontologue à l'engagement sémantique et d'explicitier clairement le sens de chaque concept selon le contexte en introduisant la « normalisation sémantique ».

Parfois, les méthodologies mentionnées dans la section 2.2.3 ne sont pas complètes dans le sens où elles ne reprennent pas le cycle de vie d'une ontologie. Ces méthodologies sont peu suivies dans la construction des ontologies médicales. Voici quelques publications concernant des ontologies construites sans suivre aucune de ces méthodologies : ontologie des signes vitaux [98], ontologie des symptômes [99], ontologie des maladies infectieuses [100], ontologie de l'épilepsie et de crise [101], et celle de diagnostic et de traitement [102]. Leur construction s'est limitée à suivre le cycle de vie d'une ontologie.

## **2.2.6 Conclusion**

La construction d'une ontologie médicale n'est pas une tâche aisée. Elle nécessite une double expertise, celle de l'ingénierie des connaissances et celle du domaine médical. Cette double expertise n'est pas facile à avoir. L'ontologue aura besoin d'une méthodologie qui lui permettra de construire une ontologie médicale en minimisant l'intervention des médecins. La seule méthodologie parmi celles qui sont citées dans cette section qui pourrait permettre cela est ARCHONTE avec les principes différentiels qu'elle préconise pour éclaircir les concepts. Cependant, ARCHONTE construit une ontologie à partir du texte uniquement et ne précise pas le type de sources textuelles qui serait plus efficaces et où le retrouver.

L'ontologue aura besoin d'autres indications et directives afin de bien mener cette méthodologie comme la manière dont la réutilisation des ontologies doit être effectuée.

## **2.3 Réutilisation des ontologies - entre théorie et application**

La réutilisation est une pratique reconnue et recommandée par la plupart des méthodologies de construction d'ontologies. En effet, elle réduit considérablement le temps et les efforts de développement puisque l'ontologue peut uniquement se concentrer sur la modélisation du nouveau contenu spécifique à son domaine d'ontologie plutôt que de créer et de gérer un contenu déjà créé et validé [103]. En outre, la réutilisation des ontologies augmente l'interopérabilité [104], réduit la redondance [14] et empêche la prolifération aléatoire des ontologies. La réutilisation peut être souple ou rude.

### **2.3.1 Définition de la réutilisation**

La réutilisation des ontologies a été définie pour la première fois de manière formelle par Katsumi et son collègue de l'université de Toronto en 2016 [105]. En effet, la réutilisation est souvent considérée comme un cas particulier de conception et de construction d'une ontologie. Elle renvoie à prendre une ou plusieurs ontologies existantes et de les manipuler d'une manière ou d'une autre afin de satisfaire aux exigences de conception. L'objectif de cette conception est de développer des axiomes qui capturent les modèles voulus. La définition fournie par ces chercheurs se base sur les modèles voulus.

Les modèles voulus illustrent les exigences qui résident dans l'esprit de l'ontologue. Certains de ces modèles peuvent être partiellement caractérisés par des axiomes ou des questions de compétence. D'autres ne peuvent être révélées que si l'ontologue reconnaît une erreur, c'est-à-dire un modèle que les axiomes permettent de reconnaître comme quelque chose d'indésirable ou un modèle qu'il s'attend à trouver vrai dans sa théorie, mais que les axiomes qu'il a définis excluent.

Pour qu'une ontologie  $T1$  soit réutilisable pour construire l'ontologie  $T2$ , certains des modèles de  $T1$  doivent s'aligner avec les modèles voulus de  $T2$ . L'ontologie  $T1$  peut-être :

- plus faibles que nécessaire, dans ce cas, ces modèles empêcheront certains aspects des exigences d'être satisfaits. Ils doivent être renforcés pour les réutiliser ;

- plus forte que nécessaire, dans ce cas, ces modèles dépassent ce qui a été spécifié dans les exigences. Ils doivent être affaiblis pour les réutiliser.

### 2.3.2 Défis de la réutilisation

En considérant la difficulté à choisir les ontologies à réutiliser, les principales limites et défis ont été identifiés lors du sommet de l'ontologie 2014 [106], qui sont les suivants :

- 1) ***Incompréhension et incohérence*** : Un manque de compréhension de l'ingénieur des connaissances d'une ontologie et une tentative de sa réutilisation conduiraient cette dernière à l'échec. Cela peut aussi parfois résulter d'un écart entre la signification que l'ontologue perçoit à partir du nom d'un concept et sa signification réelle résultant des axiomes de l'ontologie.
- 2) ***Trouver l'ontologie appropriée (pertinente)*** : Trouver la bonne ontologie, parmi toutes celles qui sont accessibles, qui répondent aux besoins de la conception présente un réel défi. En effet, dépendamment des différents domaines qu'elles couvrent, il est nécessaire de laisser tomber la méthode de recherche traditionnelle par mot clé et de disposer de métadonnées utiles pour identifier l'ontologie appropriée à réutiliser.
- 3) ***Cette ontologie ne convient pas*** : Dans certains cas, une ontologie peut être appropriée et peut répondre aux exigences, mais pas directement réutilisable. Cela est dû au manque de l'ontologie à une certaine partie, à une sémantique ou un concept et même parfois, elle peut ne pas être formalisée dans le langage souhaité. Pour mieux tirer profit des ontologies existantes, l'ontologue devrait être en mesure de déterminer quelles parties d'une ontologie ou quelles modifications peuvent être nécessaires pour répondre aux exigences.
- 4) ***Modularité*** : Commençons par définir la notion de la modularisation ontologique. Nous pouvons distinguer deux façons de la percevoir. Elle pourrait être vue comme un processus qui vise à la décomposition d'une ontologie en des modules de petite taille. Elle pourrait également être perçue comme une étape du processus de construction de l'ontologie qui s'accomplit à travers la conception d'un ensemble de modules indépendants les uns des autres. Modulariser revient donc à sélectionner les concepts et relations pertinents en fonction des besoins et spécifications de la conception de

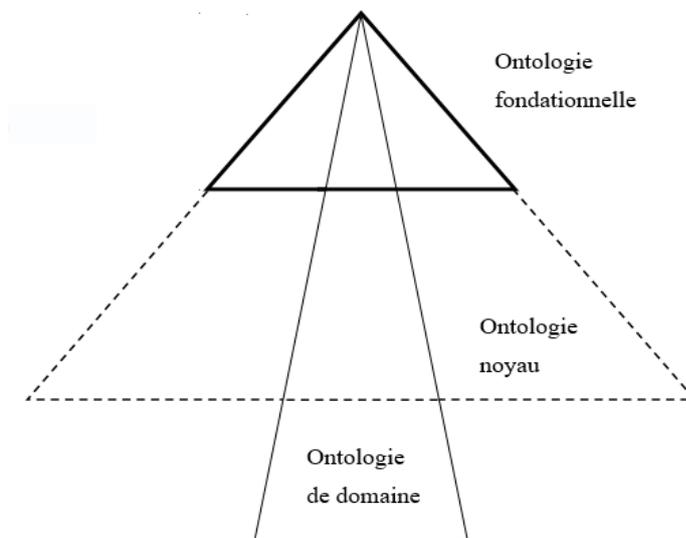
l'ontologie qu'on désire construire. La modularisation en l'occurrence pourrait être utilisée dans la réutilisation, mais elle reste problématique. L'assemblage, l'extension, la spécialisation, l'intégration, l'alignement et l'adaptation de petites ontologies modulaires doivent faire partie de la méthodologie de développement des ontologies.

- 5) **Intégration** : L'intégration de plusieurs ontologies est généralement utilisée dans la réutilisation. Parmi les moyens utilisés lors de l'intégration est l'alignement entre multiples ontologies qui peuvent potentiellement être réutilisées. En fait, comprendre comment les différentes ontologies d'un même domaine sont liées les unes aux autres est un élément essentiel pour déterminer si une ontologie peut ou non être intégrée aux autres.
- 6) **Problème de « faites-le vous-même »** : Parfois, la réutilisation peut s'avérer ne pas être une option faisable à cause du travail qu'il faut effectuer pour trouver et comprendre les ontologies existantes. Toutefois, en la comparant avec la méthode de construction à partir de zéro, elle représente un potentiel moyen pour réduire le temps de construction.
- 7) **Facteurs sociaux** : L'hypothèse que la langue dans laquelle l'ontologie a été développée serait aussi la même langue que l'ontologue voudrait utiliser n'est pas valable. Le problème se pose encore plus lorsque la sémantique prévue des concepts dans l'ontologie est principalement spécifiée dans la documentation au lieu d'être formellement saisie dans les axiomes de l'ontologie.

### 2.3.3 Les ontologies génériques ou de niveau supérieur

On distingue dans la littérature trois niveaux dans les ontologies, dépendamment du degré d'abstraction et de généralité des concepts qu'elles couvrent. Comme l'illustre la **Figure 3**, les ontologies de domaine, qui se trouvent au niveau le moins abstrait, contiennent des concepts spécifiques à un domaine particulier comme le diabète, la pneumonie et le diagnostic de l'hypertension. Au deuxième niveau, nous trouvons les ontologies noyaux, qui couvrent les concepts fondamentaux d'une discipline comme le diagnostic d'une maladie. Au niveau le plus haut et le plus abstrait se trouvent les ontologies génériques ou fondationnelles ou de niveau supérieur. Elles couvrent les concepts généraux qui pourront être utilisés dans plusieurs disciplines en leur fournissant une base ontologique commune. Généralement, une

ontologie noyau sert de lien entre une ontologie de domaine et une ontologie fondationnelle (générique).



**Figure 3 :** Les trois niveaux d'ontologies (schéma adapté de [107]).

### 2.3.4 Qu'est-ce qu'on réutilise ?

Selon la disponibilité, de la pertinence et de la qualité des ontologies médicales existantes, deux sortes de réutilisation peuvent être considérées :

- 1- Réutilisation rude (Hard reuse) :** Elle consiste à réutiliser une ontologie dans son intégralité en faisant son importation. Cette méthode est utilisée pour la réutilisation des ontologies génériques (voir section 2.3.5).
- 2- Réutilisation souple ou partielle (Soft reuse) :** Elle signifie qu'il faut uniquement réutiliser les parties qui répondent aux spécifications et besoins de la conception, et ignorer le reste. Elle réutilise les concepts en référençant leurs URI<sup>26</sup> (Uniform Resource Identifier).

Lors de la réutilisation rude, l'ontologie générique est importée dans sa totalité. Ensuite, l'ontologue détermine les classes et les relations à réutiliser concrètement. Cela revient à placer les nouvelles classes du domaine qu'il veut représenter à l'intérieur des classes qui leur

---

<sup>26</sup> URI est un identifiant qui désigne de manière unique une ressource sur le web.

correspondent le mieux. Le même processus s'applique aux relations. Habituellement, il y a des classes et relations de l'ontologie générique qui ne sont pas réutilisées. Cependant, on préfère les laisser pour d'éventuelles futures extensions.

Lors de la réutilisation partielle, l'ontologue recherche dans les ontologies existantes les concepts (classes et relations) qui correspondent au domaine de l'ontologie qu'il veut modéliser. De ce fait, il réutilise le concept en référençant son URI.

En plus des classes et des relations qui peuvent être réutilisées, il en va de même pour les propriétés ainsi que les annotations. Les annotations incluent les définitions, les synonymes, etc.

### **2.3.5 Pourquoi une ontologie de niveau supérieur ?**

Les ontologies de niveau supérieur ont été depuis plusieurs années l'objet de nombreuses discussions et études, dans la communauté de l'ingénierie des connaissances [108]. Au moins cinq fonctions leur sont octroyées : 1) faciliter le processus de conception des ontologies de domaine [109], directement ou indirectement par le biais d'une ontologie noyau ; 2) expliciter les choix de conceptualisation qui sont faits de manière implicite dans les ontologies de domaine (par exemple leurs engagements ontologiques) et offrir à ce titre un cadre de comparaison [110] ; 3) fédérer différentes ontologies de domaine en les intégrant à un système de représentation sémantique unique omnienglobant [111] ; 4) améliorer la qualité et l'utilisabilité, voire la réutilisabilité, la maintenance et l'évolutivité [112] ; et 5) améliorer l'interopérabilité sémantique [113, 114].

Quoique le plus large consensus au sujet de la fonction des ontologies de niveau supérieur concerne l'interopérabilité sémantique des ontologies de multiples domaines. L'interopérabilité selon le dictionnaire informatique standard d'IEEE est « l'aptitude de deux (ou plus) systèmes ou composants à échanger des informations et à utiliser ces dernières »<sup>27</sup> [115]. Selon les auteurs de [116], trois types d'interopérabilité sont distingués : l'interopérabilité syntaxique, technique et sémantique. L'interopérabilité syntaxique, qui est

---

<sup>27</sup><https://ieeexplore.ieee.org/browse/standards/dictionary?activeStatus=false&queryText=interoperability>, dernier accès le 17/03/2021.

généralement liée aux formats, est la capacité d'échanger les informations et d'y accéder sans tenir compte de leurs significations. L'interopérabilité technique, qui est fréquemment liée aux composants matériels/logiciels, aux systèmes et aux plateformes, est centrée sur la communication et le fonctionnement des protocoles entre machines. L'interopérabilité sémantique est la capacité de comprendre l'ontologie et de l'utiliser [117].

En effet, les ontologies de niveau supérieur assurent l'interopérabilité entre systèmes, car d'une part, en utilisant une description sémantique explicite des concepts dont ces entités sont l'instanciation linguistique, elles réduisent les risques de mésinterprétation des entités composant l'univers de discours considéré, et d'une autre part, en mettant en place une transparence conceptuelle qui assure la possibilité d'un partage de l'information [109].

### **2.3.6 Basic Formal Ontology (BFO)**

Basic Formal Ontology (BFO)<sup>28</sup> est une petite ontologie de niveau supérieur initiée par Barry Smith et Pierre Grenon dans le but de soutenir l'intégration des données obtenues de la recherche scientifique. Elle ne contient aucun des termes physiques, chimiques, biologiques ou d'autres termes issus des autres domaines. Elle est conçue pour aider à garantir que les ontologies de domaine construites sur sa base représentent les universels dans leurs domaines respectifs d'une manière cohérente et structurée. De ce fait, BFO aide les ontologues en fournissant une structure commune de haut niveau pour soutenir l'interopérabilité des multiples ontologies de domaine créées. Le projet BFO a été initié pour la première fois en 2002 sous l'égide du projet « Formes de vie » et parrainé par la Fondation Volkswagen. BFO se développe dans une orientation philosophique qui se chevauche avec celle des deux ontologies de niveau supérieur DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) et SUMO (Suggested Upper Merged Ontology). Elle est de ce fait utilisée par plus de 250 ontologies. En effet, de multiples ontologies travaillant ensemble dans le cadre de l'initiative OBO Foundry utilisent l'ontologie formelle de base (BFO) comme point de départ pour la catégorisation des entités et des relations dans leurs domaines de recherche respectifs.

Nous allons introduire les composantes de cette ontologie de niveau supérieur selon le livre écrit par son concepteur Barry Smith [118] et qui est aussi l'un des fondateurs d'OBO

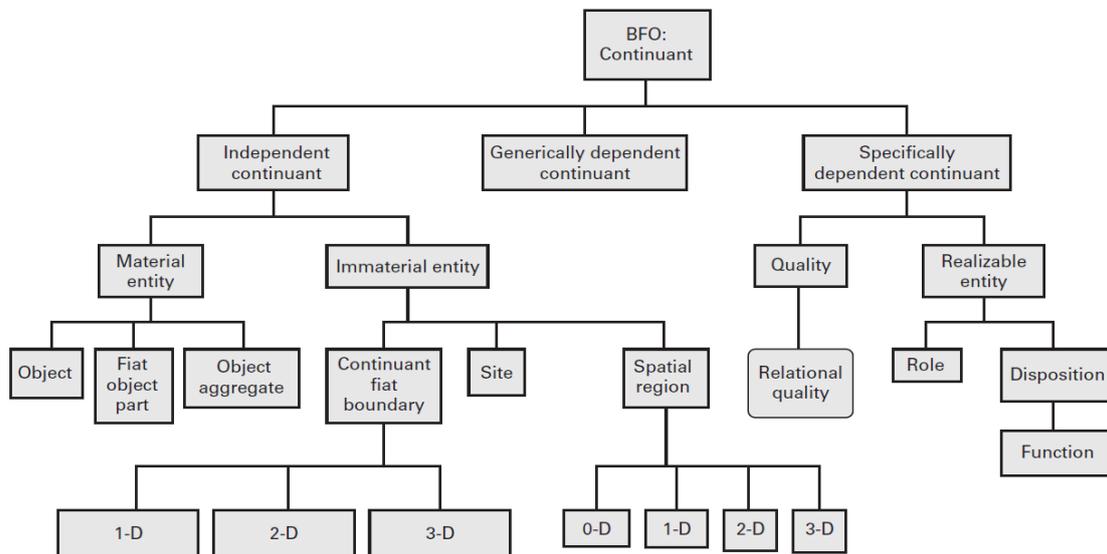
---

<sup>28</sup><http://ifomis.uni-saarland.de/bfo/>, dernier accès le 17/03/2021.

Foundry. Les entités de BFO sont divisées en deux catégories disjointes : « BFO : continuant » et « BFO : occurrent » (voir **Figure 4** et **Figure 5** respectivement).

**Continuant** : toute entité qui persiste ou continue à travers le temps, comme l'illustre la **Figure 4**, incluant (1) les objets indépendants (comme un être humain, un animal) ; (2) les continuités dépendantes, y compris les qualités (telles que la température), et les fonctions (telles que la fonction de cet interrupteur pour allumer cette lumière) ; ainsi que (3) les régions spatiales que ces entités occupent à un moment donné.

La catégorie « BFO : occurrent » modélise les événements auxquels les « BFO : continuant » participent (processus de mesure des biomarqueurs d'imagerie, application clinique des biomarqueurs d'images, etc.).

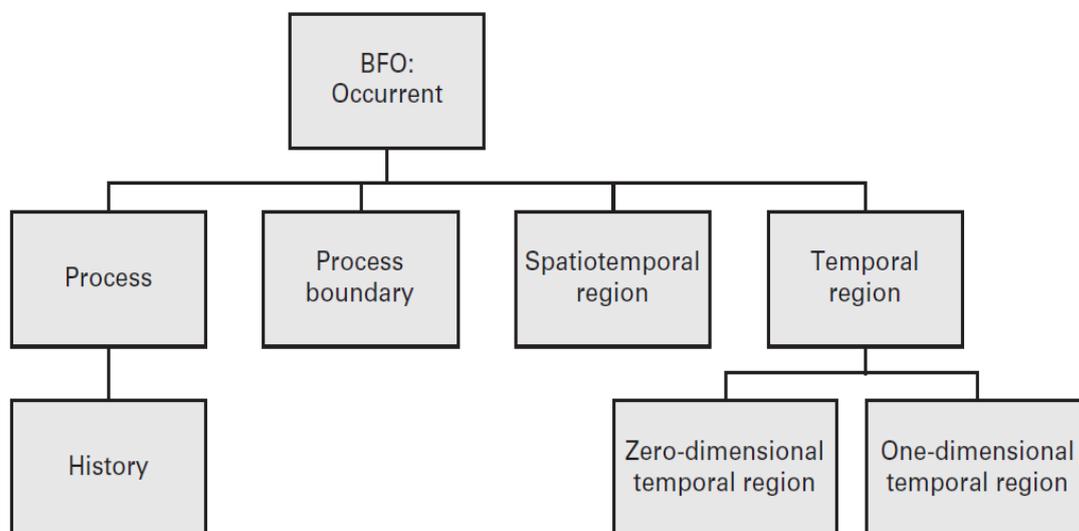


**Figure 4** : Hiérarchie de BFO continuant [118].

*Continuand* est subdivisée en trois sous-catégories qui sont : (1) *independent continuand* qui est toute entité de type *continuand* et qui est porteuse de qualité. Une entité *a* est *continuand* s'il n'y a aucune autre entité *b* telle que *a* dépend de *b*. Cela peut être *material entity* c'est-à-dire, *independent continuand* qui s'étend dans l'espace en trois dimensions, et qui continue d'exister pendant un certain temps, par exemple : un être humain, un animal, une plante, etc., ou *immaterial entity*, c'est-à-dire *independent continuand* qui ne contient aucune entité matérielle en tant que parties, par exemple : la taille d'un individu, la frontière du Pays de Galles (qui se déplace avec la rotation de la terre), etc. ; (2) *specifically independent continuand* qui est continuand dont l'existence dépend d'un ou de plusieurs *independent*

*continuant*, par exemple la couleur d'une tomate est *specifically independant continuant*, car l'existence de la couleur de cette tomate dépend de l'existence de cette tomate ; et (3) *generically independent continuant* qui est *continuant* qui dépend de l'un ou l'autre *independant continuant* qui peut lui servir de porteur, autrement dit, *a* dépend génériquement de *b* si et seulement si *a* existe et *b* existe et pour certains types *B*, *b* est un exemple de *B* et nécessairement si *a* existe, alors un certain *B* existe. De là, nous définissons *generically independent continuant* comme suit : *a* est *generically independent continuant* si et seulement s'il y a un certain *b* tel que *a* est génériquement dépendant de *b*. On peut citer ces deux exemples : motif de tourbillons rouges et blancs sur l'étiquette de cette bouteille de Coca-Cola, ce modèle de signature de cette personne, etc.

**Occurrent** : Cette catégorie modélise les événements et les processus auxquels les «BFO : continuant» participent (voir **Figure 5**). Il s'agit d'une entité qui se produit où se déroule. Elle comprend : (1) les processus qui se suivent de façon séquentielle dans leur exécution ; (2) les limites (début et fin) de chacun des processus ; et (3) les régions temporelles et spatiotemporelles dans lesquelles ces processus se produisent.



**Figure 5** : Hiérarchie de BFO occurrent [118].

*Occurrent* est subdivisé en quatre sous-catégories : (1) *process* est une entité qui dépend d'une ou plusieurs *material entity*, existe dans le temps en se produisant et ayant des parties temporelles, par exemple : le processus de la méiose, l'évolution de la maladie, ce vol de cet oiseau, etc.; (2) *process boundary* constitue la limite temporelle d'un processus, cette limite est le début ou la fin des processus qu'elle relie ; (3) *temporal region* qui est une entité

occurrente qui fait partie du temps, dans laquelle on trouve *zero-dimensional temporal region*, comme le moment de la mort de quelqu'un, ou *one-dimensional temporal region*, comme le dix-neuvième siècle ou la première heure de la journée ; et (4) *spatiotemporal region est occurrent entity* dans laquelle une *occurrent entity* peut être localisée, par exemple la région occupée par le développement d'une tumeur cancéreuse.

### **2.3.6.1 Quand et comment réutiliser BFO ?**

Il est toujours souhaitable qu'une ontologie médicale (ou non) soit réutilisable et interopérable avec d'autres ontologies. Certes, il existe des ontologies de bonne qualité reconnues au sein de la communauté des ontologies biomédicales, comme Gene Ontology (GO), qui n'utilisent pas BFO, mais il demeure fort utile de la réutiliser, pour assurer, entre autres, une meilleure interopérabilité avec d'autres ontologies.

La réutilisation cohérente de BFO implique beaucoup de réflexion et une très bonne compréhension de chaque concept du domaine de l'ontologie à construire. Aussi, la compréhension de chacun des concepts de BFO est indispensable. Ceci inclut aussi la prise en compte des axiomes qui concernent les concepts à réutiliser. Par exemple, un concept placé dans *continuant* ne peut pas avoir une sous-classe qui va avoir un héritage d'une classe issue de *occurrent*, car ce sont deux classes disjointes dans BFO.

### **2.3.7 Ontology for Biomedical Science (OGMS)**

OGMS est une ontologie qui couvre des entités impliquées dans une rencontre clinique [102]. Son objectif est de décrire comment les troubles physiques sont liés à des dispositions anormales du côté des patients, dispositions réalisées dans des processus pathologiques qui sont reconnus par le clinicien lors d'une rencontre clinique comme des signes ou des symptômes. Il s'agit d'une ontologie supérieure de domaine qui utilise l'ontologie générique de base (BFO). Elle contient des concepts très généraux qui sont utilisés dans toutes les disciplines médicales, y compris : maladie, trouble, évolution de la maladie, diagnostic, patient et fournisseur de soins. Son champ d'application est limité aux humains, mais de nombreux termes peuvent être appliqués à une variété d'organismes. OGMS a été construite pour remédier aux lacunes des ressources terminologiques comme SNOMED-CT et de NCIT quant aux définitions des termes qui ne capturent pas suffisamment les relations logiques

entre les termes définis, et cela peut être un inconvénient pour l'intégration des informations et le raisonnement.

OGMS est utilisée comme ontologie supérieure de domaine par plusieurs ontologies telles que Infectious Disease Ontology (IDO), Sleep Disease Ontology (DSO) et Vital Sign Ontology (VSO).

### **2.3.8 Conclusion**

Les avantages de la réutilisation ne sont plus à démontrer, mais le processus est loin d'être facile. Le choix de l'ontologie à réutiliser, le choix des concepts, le choix d'une représentation quand le concept en a plusieurs, les placer dans l'ontologie de niveau supérieur sont de sérieux défis. Réutiliser fait gagner du temps et de l'effort oui, mais, réutiliser efficacement produira une ontologie de meilleure qualité. Les méthodologies existantes énoncent des étapes de réutilisation, mais ne guident pas l'ontologue dans la réalisation de chaque étape. Il serait fort utile d'avoir une méthodologie qui anticipera les difficultés et les problèmes qui pourraient être rencontrés et comment les résoudre.

## **2.4 Répertoires d'ontologies : prolifération et chevauchement**

### **2.4.1 Des répertoires de données aux répertoires d'ontologies**

Un répertoire de données est une collection de données numériques accessible à une ou plusieurs entités (utilisateurs, systèmes) à des fins diverses (par exemple : processus administratifs, recherche) et qui présente les caractéristiques proposées par Heery et Anderson [119] :

- Le contenu est déposé dans un dépôt que ce soit par le créateur du contenu, le propriétaire ou un tiers ;
- L'architecture du répertoire gère le contenu ainsi que les métadonnées ;
- Le répertoire offre un ensemble minimum de services de base, par exemple, mettre, obtenir, rechercher, contrôler l'accès, etc. ;
- Le répertoire doit être durable et fiable, bien soutenu et bien géré.

Il n'est pas surprenant qu'il y a quelques années, la communauté du web ontologique et sémantique se soit intéressée à l'utilisation de dépôts pour conserver le contenu sémantique (par exemple, les ontologies). Au cours des dernières années, les ontologies ont connu un développement et une application énormes dans plusieurs domaines, en particulier dans le contexte du web sémantique. Les universités et l'industrie développent et utilisent les ontologies pour fournir de nouvelles technologies et soutenir les opérations quotidiennes. Par conséquent, il existe actuellement un grand nombre d'ontologies développées par de nombreuses parties différentes, ce qui rend nécessaires les moyens de les partager et de les réutiliser.

Les premiers efforts pour collecter la base des ontologies existantes ont proposé la création de systèmes de bibliothèques (c'est-à-dire, connus sous le nom de systèmes de bibliothèques d'ontologie) qui offraient diverses fonctions pour gérer, adapter et normaliser des groupes d'ontologies [120]. Ces systèmes ont défini un environnement important pour le regroupement et la réorganisation des ontologies en vue d'une réutilisation, d'une intégration, d'une maintenance, d'un mappage et d'un versionnage ultérieurs. Ils ont défini un modèle d'évaluation basé sur les fonctionnalités fournies par le système de bibliothèque. Voici quelques exemples de systèmes de bibliothèques : WebOnto, Ontolingua, DAML Ontology Library System, SchemaWeb, etc.

Actuellement, des efforts sont déployés pour la création des répertoires d'ontologies. Un répertoire d'ontologies est similaire à ce que Ding et al. (2001) ont défini comme un système de librairies d'ontologie [120].

Hartmann et ses collègues (2009) ont introduit le concept de répertoire d'ontologies, avec des fonctionnalités avancées telles que la recherche, la navigation, la gestion des métadonnées, la visualisation, la personnalisation, les mappages et la programmation d'une application pour interroger leur contenu/services. Un répertoire d'ontologies est une collection structurée d'ontologies, de modules et de métaconnaissances supplémentaires en utilisant un vocabulaire des métadonnées de l'ontologie. L'accès aux ressources est réalisé par le biais d'interfaces sémantiques applicables aux humains et aux machines. Un répertoire fournit donc un langage d'interrogation formel [121].

Un répertoire d'ontologies est géré par un système de gestion de répertoires d'ontologies qui est responsable de stocker, organiser, modifier et extraire de la connaissance à partir du répertoire.

## 2.4.2 Répertoires d'ontologies biomédicales

### 2.4.2.1 OBO Foundry

OBO Foundry<sup>29</sup> (Open Biomedical Ontologies) est un ensemble de développeurs d'ontologies qui s'engagent à collaborer et à respecter les principes partagés. Sa mission est de développer une famille d'ontologies interopérables qui sont logiquement bien construites et scientifiquement exactes. En 2001, OBO a été créé par Ashburner et Lewis en tant qu'organisme tuteur pour les développeurs des ontologies dans le domaine des sciences de la vie, appliquant les principes clés sous-jacents au succès de l'ontologie génétique [122].

En 2005, les développeurs d'un sous-ensemble d'ontologies OBO ont initié l'OBO Foundry, une expérience de collaboration fondée sur l'acceptation volontaire par ses participants d'un ensemble évolutif de principes enrichissant ceux de l'OBO original, et conçu pour maximiser le degré auquel les ontologies peuvent soutenir les besoins de scientifiques qui travaillent sur des ontologies concernant les sciences de la vie.

OBO Foundry représente une stratégie de réforme coordonnée des ontologies OBO existantes et de création de nouvelles ontologies, sur la base de principes partagés régissant leur développement [123]. Les principes de construction sont les suivants<sup>30</sup> :

- 1) **Utilisation ouverte (Open use)** : Chaque ontologie d'OBO Foundry doit être accessible à tout le monde sans contrainte avec deux conditions : (a) son origine doit être reconnue, et (b) elle ne doit pas être modifiée et redistribuée en utilisant le même nom d'origine ou/avec les mêmes identifiants ;
- 2) **Format commun (Common format)** : L'ontologie est disponible dans un langage formel commun avec une syntaxe concrète acceptée, par exemple OWL ;

---

<sup>29</sup> <http://obofoundry.org/>, dernier accès le 17/03/2021.

<sup>30</sup> <http://obofoundry.org/principles/fp-000-summary.html>, dernier accès le 17/03/2021.

- 3) **Espace identifiant (Identifier space)** : Chaque classe et relation ou propriété dans l'ontologie doit avoir un identifiant URI (Uniform Resource Identifier) unique. L'URI doit être construit à partir d'un URI de base, d'un préfixe unique dans OBO Foundry (par exemple GO, CHEBI, CL) et d'un identifiant local (par exemple 0000001). L'identifiant local ne doit pas être constitué d'étiquettes ou de mnémoniques ayant un sens pour l'être humain. L'ID-espace/préfixe doit être préalablement enregistré auprès de la bibliothèque OBO ;
- 4) **Version (Versioning)** : Les développeurs d'ontologies doivent documenter des procédures d'identification de versions successives distinctes, y compris la fourniture de métadonnées décrivant les changements. Ceux-ci peuvent survenir, par exemple, en réponse à des erreurs qui ont été identifiées dans les versions précédentes, en raison des progrès des connaissances scientifiques ou en raison de la nécessité d'intégrer le contenu d'autres ontologies dans la formulation des définitions ;
- 5) **Portée (scope)** : L'ontologie devrait fournir une portée bien claire et son contenu devrait lui correspondre. L'idée derrière ce principe est que chaque ontologie doit être orthogonale à (ne pas se chevaucher) d'autres ontologies, déjà hébergées dans OBO Foundry, faciliter la recherche de contenu spécifique et permettre une sélection rapide des ontologies d'intérêt, tout en permettant de créer de nouveaux termes en combinant des termes existants ;
- 6) **Définitions textuelles (Textual definitions)** : Les développeurs d'ontologies doivent fournir des définitions pour une fraction substantielle et représentative de termes dans l'ontologie, plus des définitions formelles équivalentes pour au moins un nombre substantiel de termes. Pour les termes qui manquent de définitions textuelles, il devrait y avoir des preuves de la mise en œuvre d'une stratégie pour leur fournir des définitions. Celles-ci doivent être uniques, c'est-à-dire que deux termes ne doivent pas partager une définition ;
- 7) **Relations (Relations)** : L'ontologie devrait utiliser des relations entre les concepts et les définir sans aucune ambiguïté tout en suivant le modèle de définitions proposé par OBO. Ce principe n'est pas encore détaillé sur le site de OBO Foundry <sup>31</sup> ;

---

<sup>31</sup> <http://obofoundry.org/principles/fp-007-relations.html>, dernier accès le 17/03/2021.

- 8) Documentation (Documentation) :** Les auteurs de l'ontologie doivent s'efforcer de fournir autant de documentation que possible. Celle-ci doit détailler les différents processus spécifiques au cycle de vie d'une ontologie et cibler différents publics (utilisateurs ou développeurs). Au cœur de la problématique de la documentation ontologique, on retrouve la transparence et la traçabilité du développement des artefacts. Pour chacune des étapes de développement, des procédures claires doivent être mises à leur disposition. La disponibilité de la documentation sera utilisée pour évaluer la qualité de la ressource ;
- 9) Pluralité documentée des utilisateurs (Documented Plurality of users) :** Les développeurs doivent documenter l'ontologie pour qu'elle soit utilisée par plusieurs individus ou organisations indépendants. À des fins de documentation, les développeurs doivent fournir des liens ou citations comme preuves d'utilisation en tant qu'intrants au processus de révision ;
- 10) Engagement envers la collaboration (Commitment to collaboration) :** Le développement des ontologies au sein d'OBO, en commun avec de nombreuses autres activités scientifiques axées sur les normes, devrait être réalisé de manière collaborative. Les avantages de la collaboration sont plusieurs : (1) éviter la duplication du travail ; (2) augmenter l'interopérabilité ; et (3) s'assurer que le contenu de l'ontologie est scientifiquement solide et répond aux besoins de la communauté scientifique ;
- 11) Locus d'autorité (Authority locus) :** Il devrait y avoir une seule personne responsable des communications entre la communauté et les développeurs pour communiquer avec OBO Foundry sur toutes les questions liées à la construction d'ontologies médicales, médiatiser les discussions concernant la maintenance pour des progrès scientifiques et s'assurer que tous les commentaires des utilisateurs sont parvenus ;
- 12) Conventions de nommage (Naming conventions) :** OBO Foundry propose un certain nombre de conventions de nommage pour éviter les problèmes liés à la lisibilité et à la navigation dans les hiérarchies des classes des ontologies et à leur alignement et intégration. Parmi ces conventions : 1) écrire du texte en anglais normalement, c'est à dire sans caractères spéciaux comme « \_ », utiliser les espaces pour séparer les mots ; 2) épeler les abréviations, elles peuvent être rajoutées comme des propriétés ; 3) faire en sorte que les noms primaires (labels primaires) soient le plus clairs que possible, car l'ontologie peut être utilisée dans un contexte différent de celui pour lequel elle a initialement été conçue ; 4) les labels doivent être uniques au sein de l'ontologie, etc. ;

**13) Maintenance (Maintenance) :** OBO est une communauté ouverte et en se joignant à l'initiative, les auteurs d'une ontologie s'engagent à la maintenir à la lumière des avancées scientifiques et à travailler avec les autres membres pour assurer l'amélioration de ces principes au fil du temps.

OBO Foundry héberge actuellement des ressources sous le format OBO et le format OWL.

Une des exigences clés pour OBO Foundry est de s'assurer que les ontologies réutilisent les définitions de termes que d'autres ont déjà créés plutôt que de créer leurs propres définitions. Donc, chaque terme doit être défini dans une seule ontologie. Les autres ontologies qui doivent utiliser le terme se réfèrent à sa définition dans l'ontologie source, rendant ainsi les ontologies orthogonales.

### **2.4.2.2 BioPortal**

Dans le domaine biomédical, le BioPortal<sup>32</sup>, développé par le National Center for Biomedical Ontologies (NBCO) à Stanford, est considéré aujourd'hui comme le répertoire ouvert de référence pour les ontologies biomédicales initialement réparties sur le web et dans différents formats.

La provenance des ontologies de BioPortal est diverse. Un certain nombre d'ontologies proviennent d'OBO Foundry, de l'UMLS<sup>33</sup> (Unified Medical Language System), un ensemble de terminologies qui sont manuellement intégrées et distribuées par la Bibliothèque nationale de médecine des États-Unis. BioPortal inclut également des ontologies qui sont développées dans une variété de formats, y compris le format de fichier OBO, le format RRF de l'UMLS et bien sûr les standards du web sémantique OWL, RDF(S), et plus récemment, SKOS (Simple Knowledge Organization System). OWL reste le format le plus utilisé. Ces ontologies comprennent le thésaurus NCI<sup>34</sup>, l'ontologie des maladies humaines, les Medical Subject Headings (MeSH), la nomenclature systématisée de la médecine - termes cliniques (SNOMED-CT), l'ontologie des gènes, le modèle fondamental d'anatomie, etc.

---

<sup>32</sup> <http://bioportal.bioontology.org>, dernier accès le 17/03/2021.

<sup>33</sup> <https://uts.nlm.nih.gov/home.html>, dernier accès le 17/03/2021.

<sup>34</sup> <https://ncithesaurus.nci.nih.gov/ncitbrowser/>, dernier accès le 17/03/2021.

Les informations fournies dans la page des métriques des ontologies du BioPortal contiennent des données sur 869 ontologies en juin 2020.

BioPortal offre un nombre de services, à savoir *Ontology Recommender* qui permet de prendre en entrée un ensemble de mots clés (ou un paragraphe) qu'il va utiliser pour recommander les ontologies qui couvrent ces mots clés. *Annotator* est un autre service qui génère des annotations pour le texte donné en entrée. Il fait ensuite correspondre les mots du texte aux termes des ontologies en effectuant une comparaison exacte des chaînes de caractères entre le texte et les noms, synonymes et identifiants des termes de l'ontologie. BioPortal offre aussi un service de mapping (correspondance) entre une ontologie que l'utilisateur aurait choisie avec d'autres ontologies existantes dans le répertoire. Il retournera comme résultats les ontologies avec lesquelles l'ontologie donnée en entrée a des correspondances ainsi que le nombre de correspondances trouvées.

### 2.4.2.3 UMLS

UMLS<sup>35</sup> (Unified Medical Language System) est un répertoire de vocabulaires du domaine biomédical développé par la Bibliothèque Nationale de Médecine des États-Unis.

La principale composante de l'UMLS est le métathésaurus, un répertoire de concepts biomédicaux interdépendants. Le métathésaurus constitue la base unifiée des concepts médicaux et il est considéré comme la plus grande base de données terminologiques (voir **Figure 6**). Il fournit un lien automatique entre les différentes terminologies biomédicales de plusieurs sources de données. L'un des objectifs de l'UMLS est de fournir une plateforme permettant de regrouper tous les thésaurus, nomenclatures et classifications existantes dans le domaine médical [124]. Il comprend des synonymes, des variations lexicales et des concepts associés. La première version du métathésaurus « Meta-1 » comprenait déjà 30 000 concepts avec plus de 60 000 termes et 100 000 relations. Actuellement, la version 2017AA du métathésaurus contient plus de 3.47 millions de concepts avec plus de 13.5 millions de termes appartenant à 201 terminologies biomédicales telles que MeSH, SNOMED-CT, CIM-9 et CIM-10. « En d'autres termes, le métathésaurus ne représente ni une ontologie biomédicale propre à la NLM, ni une seule vision du domaine biomédical. Il préserve toutes les visions

---

<sup>35</sup> <https://www.nlm.nih.gov/research/umls/>, dernier accès le 17/03/2021.

actuelles présentes dans toutes les terminologies »<sup>36</sup>. Les trois principaux composants du métathésaurus sont les suivants : concepts, termes et relations. Les synonymes des termes sont regroupés pour former un concept et les concepts sont liés à d'autres grâce à différents types de relations générant un graphe. Les relations entre eux peuvent être héritées de la structure des vocabulaires initiaux ou générées par les éditeurs du métathésaurus. Les relations symboliques peuvent être hiérarchiques par exemple, « est une sorte de » ou « est », « partie de » ou associative par exemple, « causé par ». Des relations statistiques entre les concepts sont dérivées de la cooccurrence des termes d'indexation MeSH dans les citations MEDLINE [124].

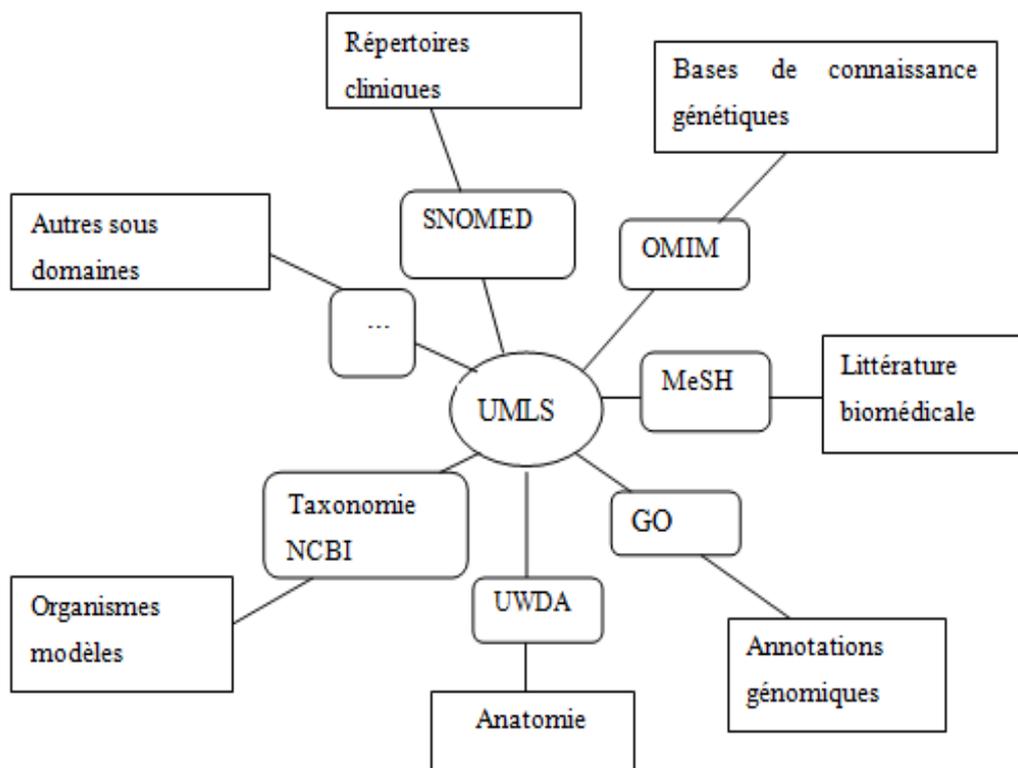
Les deux autres sources de connaissances de l'UMLS sont :

- 1) Le réseau sémantique, qui fournit des catégories de haut niveau, appelées types sémantiques, utilisées pour classer chaque concept du métathésaurus. Ce réseau élabore une partie de la sémantique des concepts du métathésaurus. Cela est fait en attribuant à chaque concept un ou plusieurs types sémantiques. Cette attribution des types sémantiques aux concepts joue également un rôle dans l'intégration de nouvelles terminologies dans l'UMLS [125] ;
- 2) Le lexique SPECIALIST et d'autres outils lexicaux<sup>37</sup> : Un grand lexique syntaxique de l'anglais biomédical et des outils pour normaliser les chaînes de caractères, générer des variantes lexicales et créer des index.

---

<sup>36</sup> <https://www.ncbi.nlm.nih.gov/books/NBK9676/>, dernier accès le 17/03/2021.

<sup>37</sup> <https://lexsrv3.nlm.nih.gov/Specialist/Home/index.html>, dernier accès le 17/03/2021.



**Figure 6 :** Les différents sous-domaines intégrés dans UMLS [124].

#### 2.4.2.4 AberOWL

AberOWL<sup>38</sup>, en plus d'être un répertoire d'ontologies, est aussi une infrastructure de raisonnement [126]. En effet, AberOWL peut non seulement permettre l'accès aux connaissances contenues dans les ontologies, mais aussi, être utilisée pour des requêtes sémantiques sur des données annotées avec des ontologies, y compris le grand volume de données qui sont de plus en plus disponibles par le biais des « endpoints » du langage SPARQL. Il contient environ 812 ontologies en juin 2020.

#### 2.4.3 Prolifération des ontologies, chevauchements et conflits

Comme mentionné ci-haut, chaque répertoire contient un certain nombre d'ontologies et ce nombre est en évolution continue, vu l'intérêt porté par la communauté de l'informatique médicale au développement d'ontologies. La qualité de ces ontologies reste encore un grand

<sup>38</sup> <http://aber-owl.net/#/>, dernier accès le 17/03/2021.

défi puisque leur développement exige une expertise dans l'ingénierie des connaissances et une autre dans le domaine de l'ontologie. Réunir une telle équipe n'est guère aisé. En l'occurrence, nous constatons que certaines des ontologies sont, du point de vue informatique, mal conçues et les concepts sont placés de manière non cohérente dans BFO et d'autres, manquant de cohérence sémantique médicale.

Les ontologies d'OBO Foundry se retrouvent dans BioPortal ainsi que dans AberOWL dans différentes versions. Relation Ontology (RO)<sup>39</sup> est un exemple d'ontologies dont les versions sont différentes dans les trois répertoires.

OBO Foundry, en plus des principes que le comité d'évaluation vérifie, se focalise aussi sur la nouveauté des concepts dans l'ontologie, c'est-à-dire si l'ontologie contient des concepts qui n'existent pas dans les ontologies existantes dans le répertoire, elle aura plus de chances d'être acceptée. Bien que la communauté de ce répertoire prône le principe du non-chevauchement des ontologies, c'est-à-dire le même concept ne doit pas avoir plusieurs identifiants et définitions, on trouve plusieurs concepts avec plusieurs représentations provoquant ainsi un conflit lors de la réutilisation.

## 2.4.4 Conclusion

Cette section sur les répertoires des ontologies médicales rejoint parfaitement la section sur la réutilisation des ontologies. En effet, plusieurs ontologies se retrouvent dans plusieurs répertoires et parfois dans différentes versions. OBO Foundry recommande un certain nombre de principes pour le développement d'une ontologie, mais il se base sur les ontologies qui contiennent de nouveaux concepts plus que les concepts réutilisés. Ceci peut être une limite puisque l'originalité d'une ontologie ne se traduit pas par la nouveauté de ses concepts, mais par la nouveauté du domaine qu'elle couvre et de la façon de la construire. Quant à BioPortal, il reprend toutes les ontologies d'OBO Foundry en plus d'autres ontologies. Pour chaque ontologie de ce répertoire candidat à la réutilisation, nous recommandons de vérifier son existence dans OBO pour l'utiliser puisqu'elle est plus à jour. UMLS est souvent utilisé pour l'enrichissement des concepts avec des annotations, et ce, dans le but de renforcer l'interopérabilité et la réutilisabilité.

---

<sup>39</sup> <http://www.obofoundry.org/ontology/ro.html>, dernier accès le 17/03/2021.

## 2.5 Guides de bonnes pratiques cliniques

La recherche de preuves scientifiques est une tâche qui nécessite un effort et un travail de revue et de synthèse importants. Ces dernières années, nous assistons de plus en plus à une production prolifique de publications scientifiques et cette augmentation n'est plus compatible avec l'activité quotidienne des médecins. En effet, une recherche d'un concept clinique sur PubMed<sup>40</sup> peut donner des milliers de publications et de références. Bien que des moyens sont mis en place pour mieux cibler les thématiques, le volume d'articles scientifiques reste non négligeable. Un travail sérieux de tri et de synthèse s'impose pour mieux servir les médecins et assurer leur formation continue. C'est dans cet objectif que se sont développés les guides de bonnes pratiques cliniques (GBPC). Ce chapitre est dédié exclusivement aux GBPC. Après les avoir définis (voir section 2.5.1), la section 2.5.2 présente les GBPC dans la pyramide des preuves, la section 2.5.3 est consacrée à l'élaboration des GBPC. Nous présentons dans la section 2.5.4, les différents obstacles liés à leurs utilisations pour enfin, finir avec les différents formalismes des GBPC (voir section 2.5.5).

### 2.5.1 Définition

Les GBPC appelés aussi recommandations pour la pratique clinique (RPC) sont des documents textuels médicaux édités dans le but de rassembler l'état de l'art de la prise en charge des pathologies et des cas cliniques, ce que nous appelons des recommandations élémentaires<sup>41</sup>. Celles-ci sont basées sur le concept de l'Evidence-Based Medicine (EBM) ou médecine factuelle (appelée aussi médecine fondée sur les preuves). Sackett (1996) définit l'Evidence-based Practice comme : « *the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patient* » [127] , p.71. Elles représentent un ensemble d'instructions spécifiques et méthodologiquement développées afin d'assister le praticien ainsi que le patient dans la décision d'un soin approprié, et ce, selon des circonstances cliniques spécifiques [128].

---

<sup>40</sup> <https://www.ncbi.nlm.nih.gov/pubmed>, dernier accès le 17/03/2021.

<sup>41</sup> La recommandation élémentaire désigne chacune des phases correspondant à une conduite à tenir pour une situation clinique donnée. La structure des recommandations de bonne pratique clinique correspond à un assemblage de recommandations élémentaires.

Les GBPC sont généralement construits par des sociétés savantes ou des agences de santé nationales comme NICE<sup>42</sup> (National Institute for Health and Clinical Excellence) pour l'Angleterre, NGC<sup>43</sup> (National Guideline Clearinghouse) pour les États-Unis, INESS (Institut National d'excellence en Santé et en Services Sociaux) et CMA<sup>44</sup> (Canadian Medical Association). Ces documents sont généralement mis à jour régulièrement [129] en fonction des avancées scientifiques, car ils ne sont pas rigides, mais ils s'adaptent aux progrès de la science [130].

Les agences de santé, sociétés savantes et entreprises spécialisées dans l'édition de GBPC y voient un moyen efficace pour l'amélioration générale des pratiques. L'objectif primordial est d'offrir au médecin l'ensemble des informations qui lui permettront de prendre des décisions de manière éclairée en lui évitant de réaliser lui-même une étude poussée de la littérature, incompatible avec sa pratique quotidienne de la médecine.

Vu leur importante intégralité, les GBPC peuvent être intégrés dans des SADC ou des systèmes d'alerte permettant ainsi de proposer des recommandations en s'adaptant au profil du patient au moment de la prise de décision [131].

## 2.5.2 GBPC dans la pyramide des preuves

La médecine factuelle distingue plusieurs niveaux de preuve qu'elle représente sous la forme d'une pyramide, appelée pyramide des preuves (voir **Figure 7**). Au niveau le plus bas de la pyramide se trouvent les études individuelles (single studies) qui représentent des recherches uniques menées pour répondre à des questions cliniques spécifiques. Les preuves de grande qualité issues de ces études sont synthétisées et se retrouvent au deuxième niveau (synopses of single studies). Le troisième niveau contient des synthèses (synthesis appelé aussi systematic reviews ou revues systématiques en français) qui sont des résumés complets de toutes les preuves entourant une question de recherche spécifique. Ces synthèses sont par la suite résumées pour obtenir des synopsis de synthèses qui se trouvent au quatrième niveau. Ces synopsis fournissent souvent des informations suffisantes pour soutenir l'action clinique

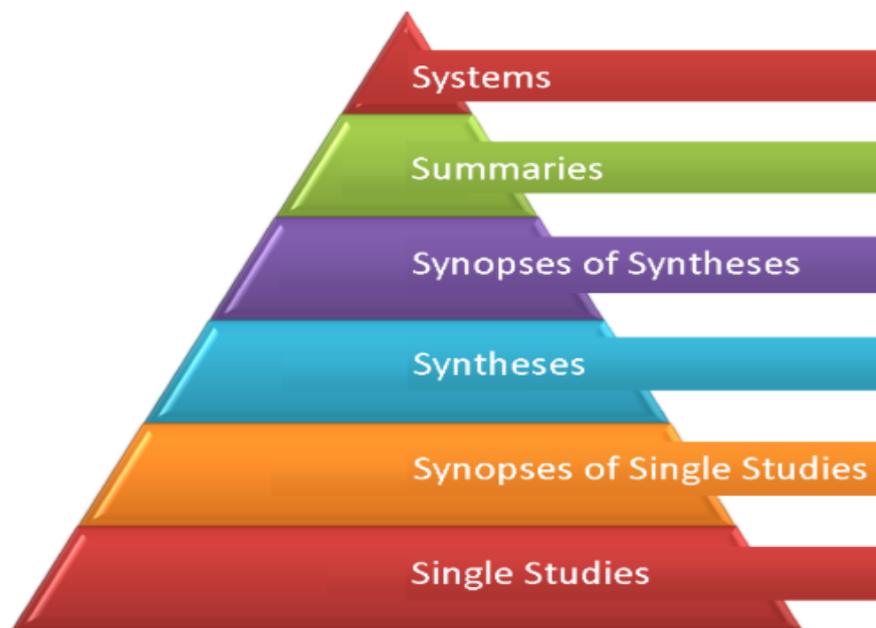
---

<sup>42</sup> <https://www.nice.org.uk/>, dernier accès le 17/03/2021.

<sup>43</sup> <https://www.guideline.gov/>, dernier accès le 17/03/2021.

<sup>44</sup> <https://www.cma.ca/>, dernier accès le 17/03/2021.

du fait qu'elles tirent des conclusions à partir de preuves aux niveaux inférieurs de la pyramide. Au deuxième niveau se trouvent des résumés (summaries) qui sont les GBPC. Ces derniers sont régulièrement mis à jour et intègrent des informations factuelles sur des problèmes cliniques spécifiques. Le premier niveau représente les systèmes (systems) qui intègrent les informations provenant des niveaux inférieurs de la hiérarchie aux dossiers individuels des patients.



**Figure 7 :** Pyramide des preuves<sup>45</sup>.

### 2.5.3 Élaboration des GBPC

L'élaboration des GBPC doit suivre une méthode précise pour qu'ils soient solides et exploitables. Ils sont définis en utilisant les données les plus récentes de la littérature médicale, étudiées et validées par un consensus d'experts spécialement réunis pour l'élaboration d'un GBPC pour une pathologie donnée.

Il existe principalement deux approches de rédaction des GBPC : 1) la méthode *top down* consiste à confier exclusivement la rédaction à des experts reconnus, au risque des critiques de la part des praticiens ; et 2) la méthode *bottom-up* qui consiste à inclure dans le groupe de travail des médecins afin de prendre en compte leurs opinions dans l'application quotidienne

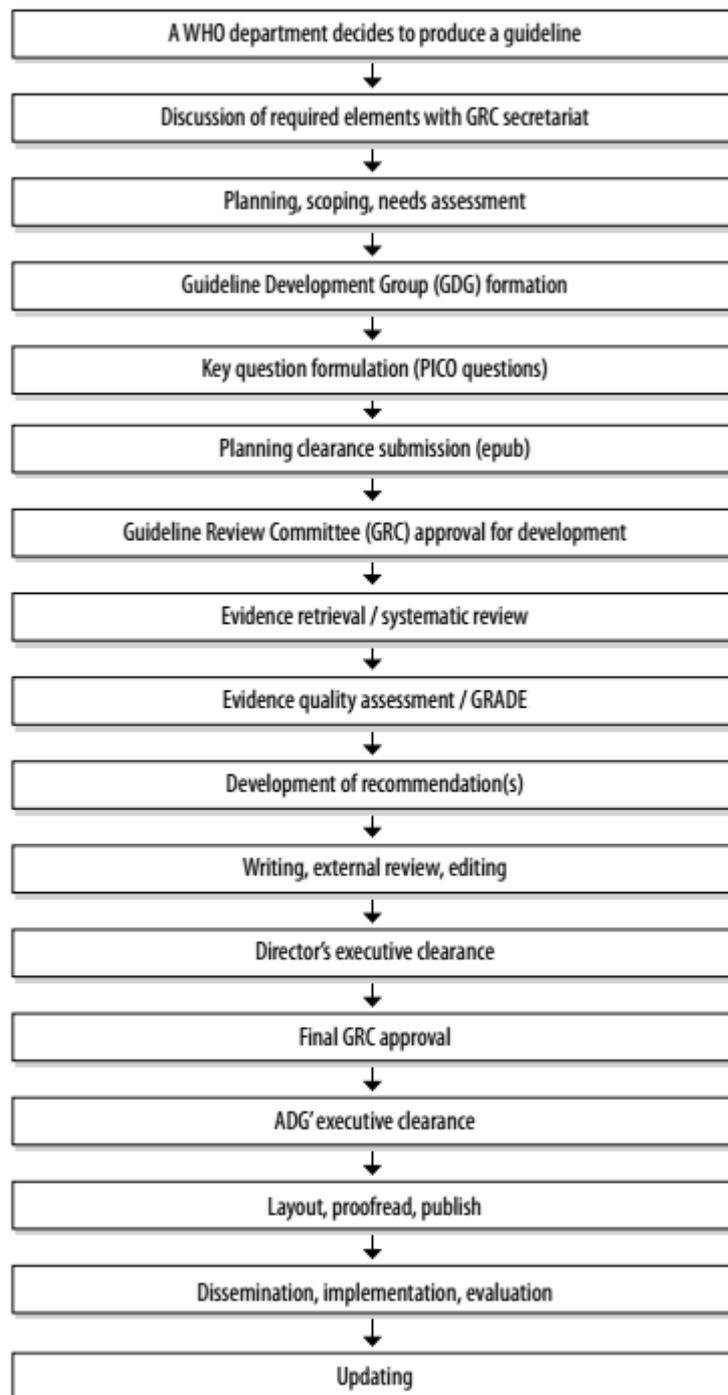
---

<sup>45</sup> <https://hslmcmaster.libguides.com/ebm>, dernier accès le 17/03/2021.

de ces recommandations [132]. L'OMS a publié un processus de développement d'un GBPC qui va de la réflexion sur son sujet à son éventuelle publication [133] (voir **Figure 8**).

Les GBPC sont généralement des documents pensés comme un outil éducatif pour les professionnels de la santé, et comme une cible pour la formation continue et permanente. Selon Patel et al. (2004), la majorité des médecins non spécialisés dans la dépression avaient de la difficulté à reconnaître les symptômes d'une maladie sans recourir à un GBPC [134].

Plusieurs enquêtes et interviews ont été menées, notamment auprès des cliniciens pour analyser leur utilisation ainsi que leur impact dans le milieu de la pratique clinique. Les résultats démontrent que 75% des praticiens considèrent les GBPC comme des sources qui fournissent de bons conseils, 71% les considèrent comme de bons outils éducationnels, 70% attestent qu'ils ont pour vocation l'amélioration de la qualité des soins, et 53% les décrivent comme un outil permettant de réduire les coûts [135].



**Figure 8 :** Processus de développement de GBPC selon l’OMS, extrait du Handbook for guideline development [133].

## 2.5.4 Obstacles à l'utilisation des GBPC

Les GBPC ont hérité de l'EBM ainsi que de leurs avantages et de leurs inconvénients. Une étude effectuée en 2011 présente les critiques qui ont été faites à l'égard d'EBM et des GBPC [136]. L'une d'entre elles est la généralisation abusive, ce qui est vrai à l'échelle d'une population pourrait ne pas l'être à l'échelle de l'individu. L'EBM tout comme le GBPC ne considèrent pas le patient. Les préférences du patient, qui est une des composantes clés de la prise de décision médicale et dont dépend l'observance thérapeutique, ne sont que rarement abordées. Pourtant, selon Woolf (1997), il faut parfois laisser au patient la liberté de décider [137]. Une autre critique est liée aux enjeux économiques pour les maladies cardiovasculaires qui ont une très forte prévalence et nécessitent bien souvent un traitement à long terme. Près de 80% des essais qui sont financés par l'industrie pharmaceutique [138], les médecins appréhendent une partialité lors de l'obtention des preuves [139]. En plus des conflits d'intérêts, Croft (2011) décrit un certain biais qui engendrerait une mise en avant des solutions thérapeutiques les plus simples à étudier. Enfin, une autre critique est notée, les GBPC sont généralement élaborés par les agences nationales de santé et sont pour certains, un procédé pour contrôler les coûts de cette dernière. Selon toujours Croft (2011), le manque de preuve pourrait servir à justifier le non-remboursement de certaines interventions. Or, l'absence de preuve d'efficacité est différente de celle de non-efficacité [136].

En 2013, Flottorp a effectué une revue de la littérature afin de proposer une liste de contrôle de 57 facteurs des GBPC qui influencent l'application d'une recommandation [140]. Ces éléments sont divisés en sept catégories de facteurs qui sont les suivantes :

- 1) **Recommandation (*Guideline factors*)** : Cette catégorie regroupe les éléments propres à la recommandation. Par exemple, de retrouver le niveau de la recommandation parmi d'autres, ou encore la cohérence avec d'autres recommandations ;
- 2) **Clinicien (*Individual health professional factors*)** : Elle regroupe les facteurs qui sont relatifs au praticien, à ses connaissances, à son accord avec la recommandation ou encore à son attitude envers les GBPC ;
- 3) **Patient (*Patient factors*)** : Les facteurs relatifs au patient sont des éléments clés de la prise de décision médicale. Ils comprennent ses besoins, ses préférences ou encore sa motivation ;

- 4) **Interactions interindividuelles** (*Professional interactions*) : Elle contient les facteurs qui portent sur la communication entre les praticiens, le partage d'avis sur les recommandations ou l'influence des leaders d'opinion ;
- 5) **Incitations et ressources** (*Incentives and resources*) : Les praticiens sont-ils encouragés à suivre les recommandations ? Des outils tels que des SADC sont-ils disponibles ?
- 6) **Évolution organisationnelle** (*Capacity for organisational change*) : Les gestionnaires ou les leaders connaissent-ils les recommandations ? Y a-t-il des opposants au changement ?
- 7) **Social, politique et légal** (*Social, political and legal factors*) : Cette catégorie concerne les facteurs désignant les obligations légales, les influences extérieures ou même la corruption.

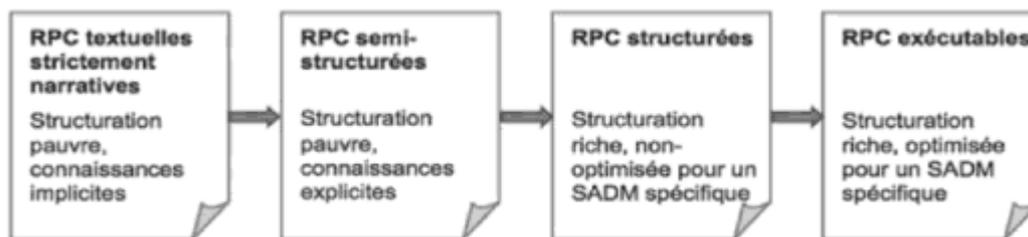
## 2.5.5 Formalisation des GBPC

Cette section couvre la formalisation des GBPC. La section 2.3.5.1 discute les différents niveaux de formalisation qui existent dans la littérature et qui vont du langage naturel aux codes, et la section 2.3.5.2, quant à elle, présente les différents langages utilisés pour leur représentation.

### 2.5.5.1 Différents niveaux de formalisation

La formalisation d'un GBPC dépend des lignes éditoriales et du public visé par les recommandations. Plus son niveau est élevé, plus le GBPC sera exploitable par un système informatique. Il est important de mentionner que le niveau de compréhension d'un GBPC par un humain dépend aussi de sa formalisation. Mais, dans ce cas, les extrêmes sont à éviter : un niveau trop élevé sera lourd et inassimilable tandis qu'un niveau trop faible ne sera pas assez clair.

La **Figure 9** présente les principaux niveaux de formalisation que peuvent adopter les GBPC et le passage du texte à un formalisme exécutable [141].

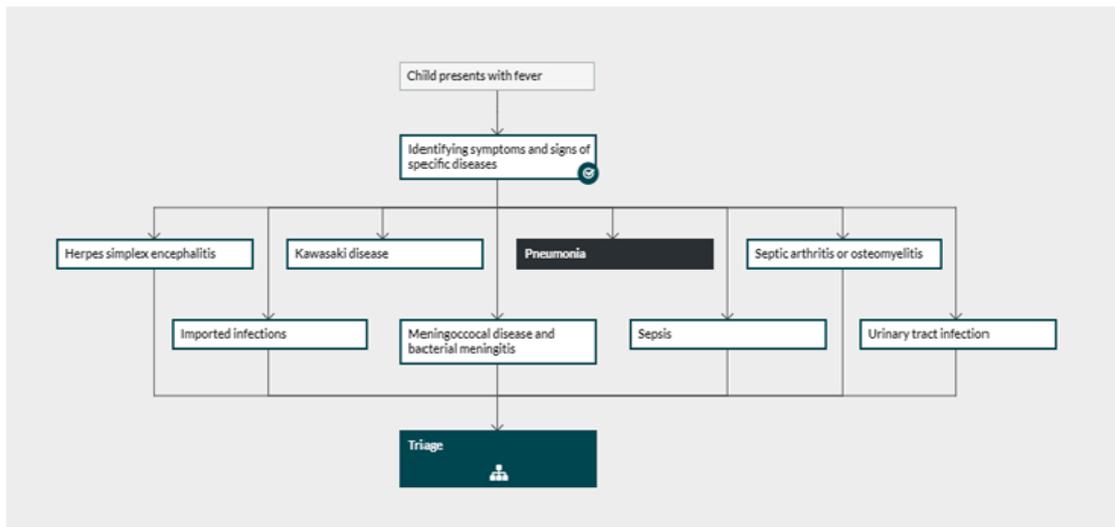


**Figure 9** : Niveaux de formalisation des GBPC.

- 1) **GBPC textuels narratifs** : Au plus bas niveau, ces GBPC sont écrits en langage naturel et contiennent peu ou pas d'éléments structurés. Ces guides contiennent plusieurs dizaines de pages. Voici un extrait du GBPC de IDSA (Infectious Diseases Society of America) concernant la gestion de la pneumonie acquise en communauté chez les enfants âgés plus de trois ans :

*« Les enfants suivant un traitement approprié doivent présenter des signes cliniques et biologiques d'amélioration dans les 48 à 72 heures. Pour les enfants dont l'état se détériore après l'admission et le début du traitement antimicrobien ou qui ne montre aucune amélioration dans les 48 à 72 heures, une recherche plus approfondie doit être effectuée »*  
[142], p.32.

- 2) **GBPC semi-structurés** : Une structuration trop basse des GBPC est nuisible à la fois à un traitement informatique, mais aussi, à un traitement par un humain. Les éditeurs de GBPC ont intégré des illustrations, des tableaux et des arbres décisionnels. Le texte y est écrit dans un style synthétique dans un but d'augmenter la clarté et de réduire les ambiguïtés. La **Figure 10** présente un exemple de GBPC semi-structuré.



**Figure 10 :** Extrait d'un NICE Pathway pour la description des symptômes et des signes cliniques pour les enfants<sup>46</sup>.

- 3) **GBPC structurés :** Il s'agit des recommandations qui sont un point de départ avancé vers une implémentation. En effet, ces GBPC sont particulièrement adaptés à une utilisation par un système informatique. En plus des nombreux formalismes génériques existants pour représenter la connaissance tels que les règles décisionnelles de production, les arbres de décision ou les graphes conceptuels, des formalismes spécifiques aux GBPC ont été développés. Ces derniers enrichissent les modèles génériques.
- 4) **GBPC exécutables :** À ce niveau, les GBPC sont capables d'être exécutés par un ordinateur.

### 2.5.5.2 Langages de représentation des GBPC

L'informatisation des GBPC représente une innovation importante par rapport à leur diffusion sur support papier, car elle peut fortement faciliter certains types de traitements susceptibles d'améliorer leur qualité. Par exemple, la détection des erreurs contenues dans les GBPC et l'amélioration de la clarté du document deviennent plus faciles. Une revue de littérature sur leurs formalismes a été effectuée et elle décrit en détail leurs approches [143]. En voici quelques-uns :

<sup>46</sup> <https://pathways.nice.org.uk/pathways/pneumonia>, dernier accès le 17/03/2021.

- **Arden Syntax** : C'est l'un des langages les plus connus pour représenter et standardiser la connaissance médicale. Il permet de définir des séquences de règles conditionnelles qui, si elles sont vérifiées, aboutiront à une action. Il a été développé en 1989 par l'équipe du professeur Jenders et a été adopté comme standard par le HL7 (Health Level 7) en 1999 [144]. La version initiale a été à la base de nombreux schémas d'encodages pour des systèmes experts tels que celui utilisé dans le système HELP (Health Evaluation through Logical Processing) à l'hôpital LDH dans l'Utah [145] qui permettait de signaler aux praticiens des résultats de laboratoire anormaux ou des interactions médicamenteuses potentiellement dangereuses [146]. Le niveau élevé de formalisation couplé à une certaine simplicité syntaxique permet à des médecins ayant ou pas de compétences informatiques de créer leurs propres programmes exécutables. Ce langage est particulièrement adapté à la génération d'alertes et de rappels (par exemple, « si PAS>180 mmHg alors alerter le médecin »).

- **GLIF** : Le GuideLine Interchange Format est un modèle pour représenter les GBPC [147]. C'est un projet qui a été initié par Intermed (Harvard, Columbia et Stanford) en 1998. L'idée était de créer un modèle de GBPC s'appuyant sur le modèle orienté-objet qui soit adapté à la prise en charge de pathologies chroniques et partageables entre institutions médicales. Il consiste en un ensemble de classes, d'attributs, de types de données et des valeurs. Il propose une modélisation par la planification des soins et permet de modéliser les GBPC sous la forme de diagrammes algorithmiques selon une syntaxe spécifique en décrivant des étapes ordonnées dans le temps. Celles-ci sont décrites sous la forme de conditions, d'actions, de décisions, de modalités d'exécution. Ces étapes sont décrites par : (i) des conditions à évaluer (*conditional step*), équivalentes à la partie (Si) d'une règle ; (ii) des actions (*action Step*) équivalentes à la partie (Alors) d'une règle ; de la simultanéité (*branch step*) qui permet de spécifier l'exécution simultanée des différentes étapes ; et (iii) de la synchronisation (*synchronisation step*) qui permet d'exécuter les différentes branches.

## 2.5.6 Conclusion

Ce chapitre a présenté les GBPC et leur utilité. Les GBPC sont des énoncés systématiques qui intègrent des informations factuelles sur des problèmes cliniques spécifiques. Ils sont construits à partir d'une synthèse des meilleures et plus récentes preuves scientifiques. Leur objectif est d'améliorer la qualité des soins en aidant les médecins et les patients à prendre des décisions appropriées dans des circonstances cliniques précises et cela en leur offrant de

l'information nécessaire. Encore plus, ils peuvent aider un ontologue à comprendre le domaine de son intérêt sans nécessairement faire appel à un médecin. Nous avons aussi constaté que la littérature parle de la difficulté des médecins à suivre ces GBPC et en l'occurrence, construire une ontologie à partir de ces documents et ensuite extraire des règles décisionnelles pourraient constituer un SADC qui aidera le médecin à mieux les intégrer dans sa pratique clinique.

## **2.6 Conclusion de la revue de littérature**

Nous avons présenté dans ce chapitre une revue de littérature couvrant les concepts les plus importants qui touchent à notre problématique de recherche et à la gestion de la connaissance médicale. En effet, le langage médical contient un vocabulaire extrêmement riche et difficile à manipuler d'autant plus qu'il n'existe aucun consensus établi sur la définition des termes utilisés. Les textes médicaux sont parfois imprécis, ambigus et ils contiennent souvent des abréviations et des acronymes. Les synonymes sont nombreux et le même terme peut avoir plusieurs significations selon l'auteur ou le contexte. Pour une description et une communication efficace et dépourvue d'ambiguïté, l'utilisation des ontologies est nécessaire. De ce fait, nous avons commencé notre revue de littérature par introduire l'ontologie puisqu'elle est le cœur de notre travail. Nous avons confirmé l'absence d'une ontologie couvrant le domaine de diagnostic de la pneumonie bien que certaines ontologies en couvrent certains aspects. Nous nous sommes ensuite intéressés à sa construction dans la section 2.2, qui peut être manuelle ou automatique. Cette section nous a permis de comprendre que ces méthodologies présentent, entre autres, le manque d'orientation de l'ontologue durant le processus de construction d'une ontologie médicale et nous avons jugé qu'il serait utile de lui fournir les indications nécessaires. Cette section nous a encore menés à nous étaler sur la réutilisation des ontologies et à ses défis dans la section 2.3. La section 2.4 a touché à l'état des répertoires d'ontologies médicales et la réalité des ontologies médicales. La dernière section de ce chapitre fut dédiée aux GBPC, à leur utilité dans la pratique clinique, mais aussi leur exploitation dans la construction d'une ontologie médicale.

## CHAPITRE 3 OBJECTIFS DE LA RECHERCHE

Après avoir situé notre problématique générale de recherche au premier chapitre et présenté une revue de littérature exhaustive au second chapitre, nous décrivons dans ce présent chapitre les objectifs de notre recherche.

Deux objectifs principaux sont distingués :

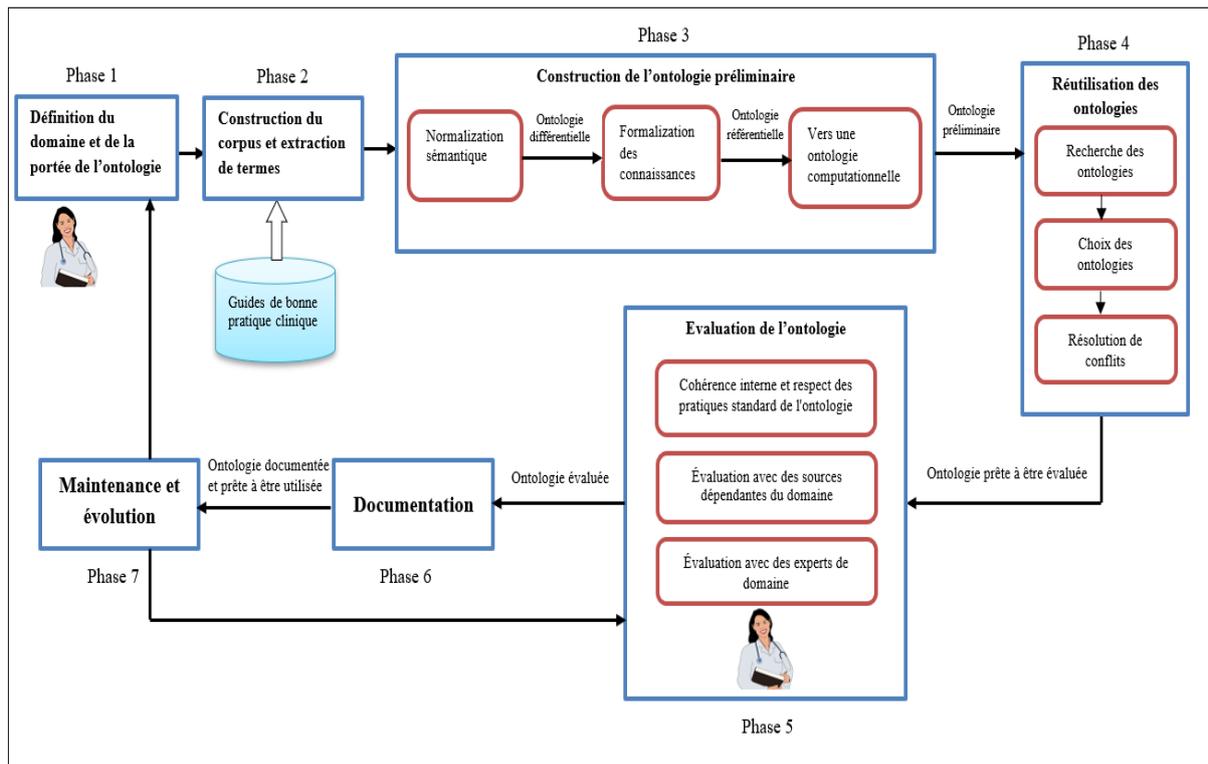
1. Proposition d'une nouvelle méthodologie de construction d'ontologies médicales en :
  - a. utilisant un corpus de textes pour construire une ontologie préliminaire (sections 4.1, 4.2 et 4.3) et en réutilisant les ontologies médicales existantes (section 4.4) ;
  - b. guidant l'ontologue dans toutes les étapes de construction. La méthodologie permettra de fournir des guides précis à chaque étape de construction ;
  - c. réduisant l'implication des experts du domaine, c'est-à-dire limitant leur intervention à la définition du domaine et de la portée de l'ontologie (section 4.1) et de son évaluation (section 4.5) ;
  - d. réutilisant efficacement les ontologies médicales existantes (section 4.4.2). Ceci implique le choix des ontologies à réutiliser, la réutilisation complète ou partielle, et la résolution de conflits quand un concept a plusieurs représentations (section 4.4.3) ;
  - e. garantissant l'interopérabilité avec les ontologies médicales existantes ainsi que sa réutilisabilité.
2. Développement d'une ontologie couvrant le domaine de la pneumonie et des domaines apparentés, et ce, au niveau du processus de diagnostic. La construction suivra notre méthodologie fournissant ainsi une preuve de concept. L'ontologie sera soumise à une évaluation rigoureuse.

Nous allons aussi considérer les deux points suivants :

- 1) Évaluation des ontologies médicales existantes réutilisées dans l'ontologie de diagnostic de pneumonie ;
- 2) Proposition de certaines améliorations au niveau de ces ontologies.

# CHAPITRE 4 MÉTHODOLOGIE DE LA RECHERCHE

Dans ce chapitre, nous présenterons d'une part, la méthodologie que nous proposons pour le développement d'une ontologie médicale et d'une autre part, nous fournissons une preuve de concept. En fait, pour valider la méthodologie, nous l'appliquons au développement d'une ontologie couvrant le diagnostic de la pneumonie. La méthodologie en question se compose de sept principales phases comme le montre la **Figure 11**. Chaque phase comporte une ou plusieurs étapes, et pour chacune nous fournissons une preuve de concept. C'est dans la première phase que l'ontologue définit le domaine et la portée de l'ontologie, en ayant recours à l'expertise des médecins (section 4.1). Dans la deuxième phase, l'ontologue construira son corpus de textes qui sera composé des guides de bonnes pratiques cliniques (GBPC) et extraira les termes (section 4.2). L'ontologie préliminaire sera le résultat de la troisième phase dans laquelle l'ontologue applique la méthodologie ARCHONTE qui est fondée, entre autres, sur les principes de la sémantique différentielle (section 4.3). Ensuite, l'ontologie préliminaire sera enrichie à la quatrième phase grâce à la réutilisation des ontologies médicales existantes. Cette phase explique clairement les étapes à suivre pour trouver les ontologies, les choisir et les réutiliser efficacement pour avoir une ontologie de qualité (section 4.4). La cinquième phase de la méthodologie est dédiée à l'évaluation (section 4.5). La sixième et la septième phase sont réservées respectivement à la documentation de l'ontologie (section 4.6) et à sa maintenance (section 4.7). La méthodologie suit aussi les principes rigides d'OBO (Open Biomedical Ontologies) Foundry pour la construction d'ontologies biomédicales collaboratives.



**Figure 11:** Les phases de construction d'une ontologie médicale.

Le **Tableau 2** résume la nature et les réalisateurs des phases dans la méthodologie proposée.

**Tableau 2 :** Résumé de la nature et réalisateurs des phases de la méthodologie.

Phases	Nature de la phase	Réalisateur de la phase
<b>Phase 1 :</b> définition du domaine et de la portée de l'ontologie	Manuelle	Ontologue et expert de domaine
<b>Phase 2 :</b> Constitution d'un corpus de textes et extraction des termes	Manuelle et automatique	Ontologue
<b>Phase 3 :</b> Construction de l'ontologie préliminaire	Manuelle	Ontologue
<b>Phase 4 :</b> Réutilisation des ontologies	Manuelle	Ontologue
<b>Phase 5 :</b> Évaluation de l'ontologie	Manuelle et automatique	Ontologue et expert de domaine
<b>Phase 6 :</b> Documentation	Manuelle et automatique	Ontologue
<b>Phase 7 :</b> Maintenance et évolution	Manuelle	Ontologue

## 4.1 Définition du domaine et de la portée de l'ontologie

Cette phase consiste à établir le domaine d'intérêt clinique de l'ontologie à modéliser ainsi que sa portée. Cela peut se faire soit en consultant des experts du domaine, soit en identifiant une lacune dans la littérature scientifique. Tout d'abord, il faut répondre à deux questions fondamentales :

1. Quel est le domaine que l'ontologie va couvrir ?
2. Quel est l'objectif principal du développement de l'ontologie ?

Deuxièmement, les exigences fonctionnelles de l'ontologie ou les questions auxquelles une ontologie devrait pouvoir répondre sont définies. Pour ce faire, des questions de compétences (QC) établies par les médecins et/ou dérivées des GBPC peuvent être utilisées. En effet, ces QC sont présentées sous la forme de questions informelles auxquelles une communauté particulière d'utilisateurs pense que l'ontologie devrait répondre. L'ensemble des QC doit être le plus exhaustif possible.

À la fin de cette phase, l'ontologue a une vision claire de la portée prévue de l'ontologie, de son utilisation et des concepts qu'elle devrait couvrir.

### 4.1.1 Cas de PNADO

Pour définir le domaine et la portée de l'ontologie à développer, nous avons eu des discussions avec quelques médecins de l'hôpital de Gatineau, Québec, Canada. Ces derniers étaient préoccupés par un certain nombre de diagnostics erronés de pneumonie. Ces erreurs sont récurrentes et causent souvent des décès. Ces médecins ont attribué cette situation à une dispersion des connaissances sur le diagnostic de la pneumonie et ont souhaité « résumer » ces connaissances pour une utilisation future. Cela nous a incités à travailler à l'élaboration de l'ontologie PNADO (rappel : PNA est l'abréviation médicale de pneumonia, DO pour Diagnostic Ontology). Les exigences fonctionnelles de PNADO ont été définies à l'aide d'un ensemble de QC établi par les médecins et provenant des GBPC couvrant le diagnostic de la pneumonie. Les questions suivantes, entre autres, doivent être répondues par les médecins pour diagnostiquer la pneumonie :

1. Quels sont les symptômes et les signes cliniques de la pneumonie ?

2. Quels sont les types de pneumonie ?
3. Quels sont les agents pathogènes de la pneumonie ?
4. Quels sont les antécédents cliniques du patient ?
5. Quels sont les tests de laboratoire permettant de diagnostiquer une pneumonie ?
6. Quels sont les résultats de l'examen physique du patient ?
7. Quel est le résultat de l'imagerie pulmonaire du patient ?
8. Quels sont les résultats des examens de laboratoire du patient ?
9. Quelles sont les maladies qui peuvent être confondues avec la pneumonie ?
10. Est-ce que le patient revient d'une croisière ?
11. Est-ce que le patient revient d'un voyage ?
12. Est-ce que le patient a été en contact avec une personne malade de la pneumonie ?
13. Est-ce que le patient prend des médicaments au moment de l'apparition des symptômes ?
14. Est-ce que le patient est fumeur ?
15. Quelles sont les complications de la pneumonie ?
16. Quelle est la saison où le patient manifeste ses symptômes ?

## **4.2 Constitution d'un corpus de textes et extraction de termes**

### **4.2.1 Élaboration du corpus des connaissances**

Dans cette phase, l'ontologue extrait et identifie les termes pertinents pour le domaine de l'ontologie défini dans la phase 1, à partir d'un corpus de textes appelé aussi corpus des connaissances.

Un corpus des connaissances est un ensemble de textes collectés pour être la base d'une démarche d'acquisition des connaissances. Celui-ci est utilisé par des outils de traitement automatique de la langue (TAL) pour extraire une majorité des termes et des relations du domaine à représenter. Les avantages d'une analyse de corpus sont nombreux :

- Les textes sont des sources de savoirs privilégiées, car ils contiennent des connaissances explicites mises à disposition sur un support physique, par opposition

aux connaissances des experts souvent tacites, difficilement exprimables et accessibles ;

- L'analyse automatisée de corpus est un gain de temps par rapport aux entretiens avec des experts du domaine ;
- Les textes sont des ressources fiables et stables puisqu'ils fixent par écrit, à un instant  $t$ , un certain nombre de connaissances sur le domaine ;
- Les modèles sont *a priori* plus facilement compréhensibles et donc maintenables, car le retour aux textes est possible [148].

Nous recommandons dans la méthodologie l'utilisation de la connaissance codifiée de haute qualité (GBPC, revues systématiques, etc.).

Nous avons déjà décrit l'importance majeure des GBPC dans la pratique médicale dans le chapitre 2, section 2.5. En outre, l'utilisation des GBPC permet de compenser l'implication limitée du médecin. Nous nous basons sur ces GBPC pour extraire les concepts qui concernent le domaine de l'ontologie précédemment défini. Selon le domaine concerné par l'ontologie, des répertoires comme NICE et Cochrane peuvent être utilisés. Il existe aussi des répertoires spécialisés comme pour les maladies respiratoires<sup>47</sup>, les maladies infectieuses<sup>48</sup>, les maladies cardiovasculaires<sup>49</sup>, le cancer<sup>50</sup>, etc.

#### 4.2.1.1 Cas de PNADO

Nous avons effectué une recherche approfondie dans des répertoires nationaux et internationaux des GBPC qui couvrent la maladie de la pneumonie. Le **Tableau 3** présente quelques répertoires auxquels nous avons eu accès ainsi que leur pertinence pour la construction de notre ontologie envisagée. Cette pertinence est mesurée en fonction du degré de couverture de ces GBPC à la pneumonie ainsi qu'à son diagnostic. À part les GBPC de NICE qui sont semi-structurés, les autres sont textuels narratifs.

---

<sup>47</sup> <https://www.aarc.org/resources/clinicalresources/clinical-practice-guidelines/>, dernier accès le 10/08/2020.

<sup>48</sup> [http://www.idsociety.org/jdsa\\_practice\\_guidelines/](http://www.idsociety.org/jdsa_practice_guidelines/), dernier accès le 17/03/2021.

<sup>49</sup> <https://www.ccs.ca/en/guidelines>, dernier accès le 17/03/2021.

<sup>50</sup> <https://www.partnershipagainstcancer.ca/tools/cancer-guidelines-database/>, dernier accès le 17/03/2021.

**Tableau 3 :** Tableau illustrant des répertoires nationaux et internationaux contenant des GBPC sur la pneumonie.

<b>Guide de bonnes pratiques cliniques</b>	<b>Pays</b>	<b>Pertinence</b>
Cochrane: <a href="https://www.cochranelibrary.com/">https://www.cochranelibrary.com/</a> , dernier accès le 10/08/2020.	Royaume-Uni	Pertinent
National Institute for Health and Care Excellence (NICE): <a href="https://www.nice.org.uk/guidance/">https://www.nice.org.uk/guidance/</a> , dernier accès le 10/08/2020.	Royaume-Uni	Pertinent
Infectious Diseases Society of America (IDSA): <a href="https://www.idsociety.org/">https://www.idsociety.org/</a> , dernier accès le 10/08/2020.	États-Unis	Pertinent
Australasian Society for Infectious Diseases, national (ASID): <a href="https://www.asid.net.au/resources/clinical-guidelines">https://www.asid.net.au/resources/clinical-guidelines</a> , dernier accès le 10/08/2020.	Australie	Pas pertinent
European Society of Clinical Microbiology and Infectious Diseases (ESCMID): <a href="https://www.escmid.org/escmid_publications/medical_guidelines/escmid_guidelines/">https://www.escmid.org/escmid_publications/medical_guidelines/escmid_guidelines/</a> , dernier accès le 10/08/2020.	Europe	Pertinent
American Association for Respiratory Care (AARC): <a href="http://www.aarc.org/resources/clinicalresources/clinical-practice-guidelines/">http://www.aarc.org/resources/clinicalresources/clinical-practice-guidelines/</a> , dernier accès le 10/08/2020.	États-Unis	Pas pertinent
Pulmonary & Critical Care Medicine (pulmccm): <a href="http://pulmccm.org/main/pulmonary-and-critical-care-guidelines-clinical-practice-recommendations-ats-accp-bts-ers/">http://pulmccm.org/main/pulmonary-and-critical-care-guidelines-clinical-practice-recommendations-ats-accp-bts-ers/</a> , dernier accès le 10/08/2020.	États-Unis	Pertinent
Canadian Thoracic Society (CTS): <a href="https://cts-sct.ca/">https://cts-sct.ca/</a> , dernier accès le 10/08/2020.	Canada	Pertinent
Association of Medical Microbiology and Infectious Disease (AMMI): <a href="https://www.ammi.ca/">https://www.ammi.ca/</a> , dernier accès le 10/08/2020.	Canada	Pas pertinent
American Thoracic Society (ATS): <a href="https://www.thoracic.org/statements/tuberculosis-pneumonia.php">https://www.thoracic.org/statements/tuberculosis-pneumonia.php</a> , dernier accès le 10/08/2020.	États-Unis	Pertinent
British Thoracic Society (BTS): <a href="https://www.brit-thoracic.org.uk/standards-of-care/guidelines/">https://www.brit-thoracic.org.uk/standards-of-care/guidelines/</a> , dernier accès le 10/08/2020.	Angleterre	Pertinent
Institut National d'excellence en Santé et en Services Sociaux (INESSS): <a href="http://www.inesss.qc.ca/en/publications/inessss-guides.html">http://www.inesss.qc.ca/en/publications/inessss-guides.html</a> , dernier accès le 10/08/2020.	Canada	Pertinent mais en français
Canadian Society of Respiratory Therapist (CSRT): <a href="http://www.csrt.com/catalogue-of-guidelines/">http://www.csrt.com/catalogue-of-guidelines/</a> , dernier accès le 10/08/2020.	Canada	Pas pertinent

## 4.2.2 Extraction des termes

Dans cette étape, un outil peut être utilisé pour une extraction automatique des termes. Parmi ces outils Text2Onto [89] qui propose différents algorithmes pour extraire à partir des textes : des « concepts » (ou plutôt des candidats-termes), des relations entre ces concepts (relations d'équivalence, hiérarchiques, etc) et des instances de concepts. Dans notre travail, nous nous intéressons uniquement à l'extraction des concepts et des instances avec Text2Onto.

Il utilise trois mesures de pertinence des concepts extraits :

- **RTFConceptExtraction** : calcule la fréquence relative du terme qui est obtenue en divisant le nombre d'occurrences du terme  $i$  dans le document  $d$  par la fréquence du terme qui apparaît le plus dans le document  $d$ .

$$TF(\text{terme}, d) = \frac{\text{fréquence du terme}}{\text{fréquence du terme le plus utilisé}}$$

- **TFDIFConceptExtraction** : le TF-IDF (Term Frequency-Inverse Document Frequency) est une mesure statistique qui évalue l'importance d'un terme dans un document  $d$  par rapport aux autres documents de la même collection  $D$ .

$$TF\text{-IDF} = (\text{terme}, d, D) = TF(\text{terme}, d) \times IDF(\text{terme}, D)$$

où :

$$IDF(\text{terme}, D) = \log\left(\frac{|D|}{DF(\text{terme})}\right)$$

$|D|$  est le nombre de tous les documents.

$DF(\text{terme})$  est le nombre de documents contenant le terme.

Cette mesure est utilisée pour pondérer les termes candidats : plus la valeur TF-IDF d'un terme candidat est élevée, plus celui-ci est important dans le document analysé. S'il y a un

besoin de prendre en compte plusieurs documents dans le corpus, cette méthode présente de meilleurs résultats. C'est le cas de notre travail.

- **EntropyConceptExtraction** : calcule l'entropie en

$$\text{Entropie du terme} = P(\text{terme}) \times \log(P(\text{terme}))$$

où :

$$P(\text{terme}) = \frac{\text{fréquence du terme}}{\text{Somme des fréquences de tous les termes}}$$

Les algorithmes d'extraction sont combinables, les résultats produits sont entre 0 et 1. Pour l'extraction des instances, Text2Onto utilise l'algorithme TFIDFInstanceExtraction.

Après l'extraction de l'ensemble de termes à partir du corpus, suivra l'étape de la sélection des termes candidats du domaine, c'est-à-dire, ceux qui véhiculent les connaissances nécessaires du domaine concerné, ainsi que leur analyse. Ces termes peuvent être des verbes, des noms, des adverbes, des syntagmes et des adjectifs.

#### 4.2.2.1 Cas de PNADO

Nous avons appliqué Text2Onto sur le corpus de textes précédemment construit (section 4.2.1.1) pour l'extraction des termes à partir du corpus. Ensuite, nous sélectionnons les termes du domaine qui sont liés au diagnostic de la pneumonie :

- Symptômes et signes cliniques de la pneumonie ;
- Imagerie, tests de laboratoires et leurs résultats ;
- Pathogènes (agents et autres) causeurs de la pneumonie ;
- Type de la pneumonie ;
- Quelques termes du domaine qui peuvent être polysémiques ayant un sens médical et un autre sens : taux, opération, culture, etc. ;
- Antécédents de la pneumonie ;
- Maladies qui peuvent être confondues avec la pneumonie (diagnostic différentiel).
- Complications de la pneumonie.

Nous avons choisi d'examiner les adjectifs, les adverbes, les noms et les noms propres, les verbes, et les syntagmes, car nous pensons qu'ils véhiculent les connaissances nécessaires à l'identification des objets du domaine de la pneumonie (voir **Tableau 4**)

**Tableau 4** : Exemples de syntagmes extraits par Text2Onto.

Catégorie	Exemple
Adjectif	Pleuritic
Adverbe	Timely
Nom propre	Klebsiella
Nom	Bacteria
Verbe	To wheeze
Syntagme adjectival	Microbiological tests
Syntagme nominal	Suspicion of an anatomic anomaly chest mass
Syntagme participial	Buffered charcoal yeast extract agar
Syntagme verbal	Inability to take oral medication

### 4.3 Construction de l'ontologie préliminaire

Cette phase est dédiée à la construction de l'ontologie préliminaire. Nous avons vu à la section 2.2 du chapitre 2 que peu de méthodologies proposent de guider l'ontologue dans le processus d'organisation des connaissances d'un domaine et par la suite, des concepts entre eux. Le point central de leurs démarches repose sur l'intuition de l'ontologue quant à la manière de modéliser le domaine ou sur l'avis d'un expert. Aucune d'entre elles, sauf le système ARCHONTE, ne définit les directives précises pour expliciter véritablement les concepts à l'aide du langage.

Bachimont (2000) propose de contraindre l'ontologue à un engagement sémantique, c'est-à-dire à expliciter clairement le sens de chacun des concepts de l'ontologie, en introduisant une normalisation sémantique :

*« Les primitives nécessaires à la représentation des connaissances doivent être modélisées à partir des données empiriques dont on dispose, à savoir l'expression linguistique des connaissances. Le travail de modélisation doit s'effectuer à partir de documents attestés dans la pratique d'un domaine et*

rassemblés en un corpus. Le corpus est constitué de documents produits dans le contexte où le problème à résoudre se pose » [149], p.4.

Comme l'illustre la, ARCHONTE comporte initialement trois étapes : la normalisation, la formalisation et l'opérationnalisation [83]. Les prochaines sous-sections porteront sur chaque étape.

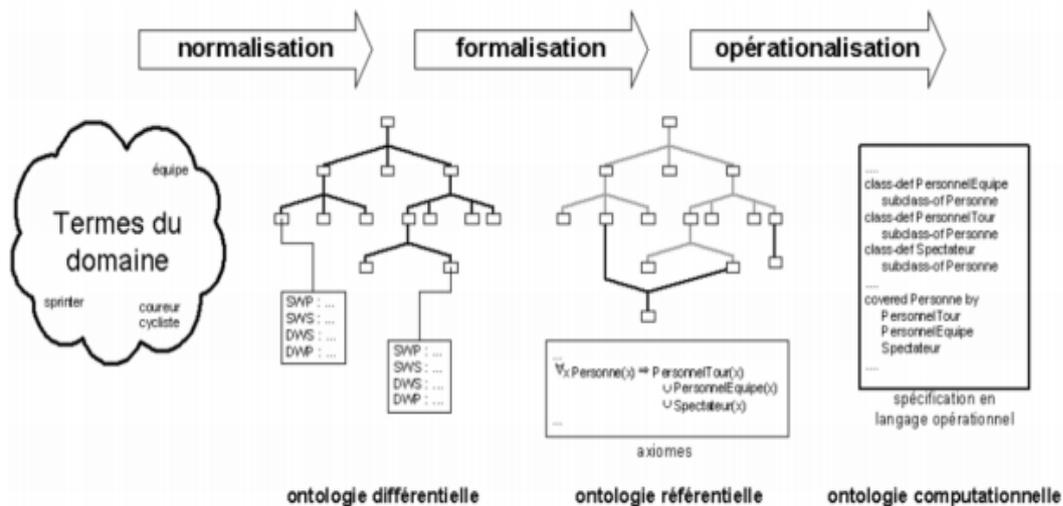


Figure 12 : Les trois étapes d'ARCHONTE [150].

### 4.3.1 Normalisation sémantique

L'objectif de cette étape est de parvenir à un accord sémantique sur la signification des termes et de construire une ontologie différentielle. Cette dernière est une ontologie où chaque concept et relation doit être défini en fonction de ses similitudes et différences avec son parent et ses frères. Bachimont (2000) propose pour cela d'utiliser la sémantique différentielle présentée dans les travaux d'Attal (1995) [151]. Cette théorie attribue un sens aux termes grâce à la définition de traits sémantiques génériques et spécifiques. Ceux-ci permettent de fixer le cadre interprétatif en fonction de l'objectif que s'est donné l'ontologue.

Sur le terrain, l'ontologue exprime en langue naturelle les identités (traits sémantiques génériques) et les différences (traits sémantiques spécifiques en opposition les uns avec les autres) que chaque concept entretient avec ceux qui lui sont proches. La structure d'une ontologie s'apparente à celle d'un arbre, ce qui facilite la détermination de la signification que doit avoir un concept selon sa position. Selon le paradigme différentiel de Bachimont (2000), la signification d'un concept (nœud dans l'arbre) se détermine en fonction de ses plus proches

voisins. Dans un arbre, les voisins les plus proches d'un concept sont le concept parent et les concepts frères. Quatre principes différentiels sont définis pour imposer d'explicitier en fonction des voisins les identités et les différences qui définissent chaque nœud dans l'arbre (ontologique). Ces principes doivent être respectés par tous les concepts (caractérisés par un terme/libellé) :

- 1) Le principe de communauté avec le père : il faut expliciter en quoi le fils est identique au père qui le subsume ;
- 2) Le principe de différence avec le père : puisque le fils existe, donc il est différent du père. Il faut expliciter en quoi le fils est différent du père ;
- 3) Le principe de communauté avec les concepts frères : tous les concepts fils d'une unité parente possèdent par définition un même trait générique, celui qu'elles partagent avec l'unité parente. Les frères sont situés au même niveau. Il faut expliciter la communauté existante entre les concepts considérés et chacun de ses frères ;
- 4) Le principe de différence avec les frères : il faut préciser la différence entre chaque concept considéré avec chacun des concepts frères.

À la fin de cette étape, l'ingénieur des connaissances obtient une ontologie différentielle. La signification de chacun des concepts la composant s'obtient de manière compositionnelle en parcourant les identités et les différences qui définissent l'ensemble des notions de l'arbre, et ce, allant de la plus générique à la notion cible considérée. En l'occurrence, la position d'un nœud dans l'arbre ontologique conditionne sa signification.

**Rappel :** Dans ce qui suit, les concepts ontologiques (qui proviennent des ontologies) sont écrits en italique pour les distinguer du reste des concepts.

#### **4.3.1.1 Cas de PNADO**

Pour construire la version différentielle de PNADO, nous appliquons les principes différentiels sur les termes précédemment choisis. Par exemple : *bacterial pneumonia*, *fungus pneumonia*, *infective pneumonia acquired prenatally*, *pneumonia due to parasitic infestation*, et *viral pneumonia* sont des concepts frères, car ils ont la similarité d'être des maladies infectieuses. Cette étape est la plus coûteuse en termes de temps.

## 4.3.2 Formalisation des connaissances

Cette étape consiste à introduire des axiomes logiques qui définissent le comportement des individus qui constituent les extensions des concepts formels. Une ontologie référentielle est ainsi construite. Désormais, l'ontologue peut mettre en œuvre des opérations ensemblistes telles que la réunion et l'intersection qui vont lui permettre de composer de nouveaux sens et concepts formels. C'est ici qu'intervient l'idée d'engagement ontologique :

*« Respecter le sens d'un concept, c'est s'engager à ce que lui corresponde une extension d'objets existants dans l'univers d'interprétation. Il s'agit donc bien d'un « engagement ontologique », puisque c'est l'existence d'objets qui est prescrite par le sens du concept » [149], p.10.*

### 4.3.2.1 Cas de PNADO

L'exemple suivant illustre l'ajout d'un axiome à PNADO. *Sign* (également appelé *clinical sign*) est un concept dans PNADO. Il est défini comme "quality of a patient, a material entity that is part of a patient, or a processual entity that a patient participates in, any one of which is observed in a physical examination and is deemed by the clinician to be of clinical significance". *Symptom* est un autre concept de PNADO ; il est défini comme "quality of a patient that is observed by the patient or a processual entity experienced by the patient, either of which is hypothesized by the patient to be a realization of a disease". Nous ajoutons un axiome pour indiquer que les deux concepts sont disjoints.

## 4.3.3 Opérationnalisation

Cette dernière étape consiste à implémenter l'ontologie référentielle dans un langage opérationnel de représentation des connaissances tel que l'OWL (Ontology Web Language). Ceci est réalisé en adoptant un formalisme de représentation (comme la logique de description). Cette étape marque le passage à l'ontologie computationnelle.

Dans cette étape, nous recommandons de suivre les principes d'OBO Foundry. Pour des fins d'interopérabilité et de réutilisabilité, l'utilisation d'une ontologie générique est fortement préconisée. BFO est largement utilisée dans OBO Foundry [118, 152] et même dans

Bioportal. Cela fait de BFO le meilleur choix pour la construction de toute ontologie médicale afin d'améliorer l'interopérabilité voire même la réutilisabilité.

Une ontologie générique de domaine qui couvre les concepts généraux spécifiques à ce domaine, comme OGMS, pourrait être utilisée. OGMS, comme mentionné dans la section 2.3.7, utilise BFO.

Au niveau du développement, l'éditeur d'ontologie Protégé<sup>51</sup>, qui offre la possibilité de visualiser les différents aspects de l'ontologie, peut être utilisé.

Pour que notre ontologie soit facilement réutilisable et interopérable avec d'autres ontologies existantes, nous lui intégrons l'identifiant UMLS de chaque concept, quand c'est possible. Pour cela, nous exploitons l'outil MetaMap<sup>52</sup> pour faire la correspondance entre l'ontologie et UMLS.

MetaMap fait partie des outils lexicaux fournis par la NLM [153]. Il permet de faire la correspondance entre du texte et les termes UMLS. Le texte à traiter passe par une série de modules. Il subit une analyse syntaxique, le décomposant en plusieurs éléments (phrases, expressions et mots clés). MetaMap propose des termes UMLS aux éléments identifiés dans le texte.

#### **4.3.3.1 Cas de PNADO**

Au cours de ce processus, en utilisant Protégé, nous opérationnalisons l'ontologie avec le langage OWL, car il répond parfaitement à nos besoins en termes d'expressivité et de simplicité. Nous utilisons l'ontologie médicale supérieure OGMS. Le choix de l'OGMS sera expliqué dans la prochaine phase (section 4.4). Pour enrichir davantage PNADO, nous extrayons les définitions, synonymes, acronymes et autres annotations de l'UMLS en utilisant MetaMap. À la fin de cette étape, nous obtenons le PNADO préliminaire.

---

<sup>51</sup> <https://protege.stanford.edu/>, dernier accès le 17/03/2021.

<sup>52</sup> <https://metamap.nlm.nih.gov/>, dernier accès le 17/03/2021.

## 4.4 Réutilisation des ontologies

Dans cette phase, nous finalisons l'ontologie préliminaire en ajoutant et en détaillant ses concepts. À cette fin, nous réutilisons certaines ressources terminologiques que nous considérons pertinentes pour le domaine de l'ontologie. La réutilisation peut être rude, c'est-à-dire que l'on importe l'ontologie complète pour la réutiliser, ou souple (partielle), c'est-à-dire que l'on importe uniquement quelques concepts. Les ontologies à réutiliser doivent être évaluées pour leur couverture du domaine et la profondeur des connaissances, ainsi que le potentiel de soutenir une inférence. La phase se compose de trois étapes : 1) trouver les ontologies ; 2) choisir les ontologies pertinentes, et 3) résoudre les conflits.

### 4.4.1 Recherche des ontologies

Dans cette étape, nous sélectionnons une ontologie pour la réutiliser lorsque nous supposons que ses axiomes pourraient être utilisés pour répondre à certaines des exigences définies dans la phase 1. Nous utilisons quatre critères principaux de recherche afin d'identifier les ontologies candidates à la réutilisation :

- i) le langage de l'ontologie candidate est le même que celui de l'ontologie que nous construisons ;
- ii) il existe une correspondance entre un sous-ensemble non vide des nouvelles exigences de l'ontologie (définies dans la phase 1) et un ensemble robuste de concepts de l'ontologie candidate ;
- iii) l'ontologie candidate est acceptée au sein de sa communauté d'utilisateurs. Cette acceptation est évaluée en utilisant certaines mesures telles que le nombre de membres de la communauté qui approuvent l'ontologie, le nombre de fois que l'ontologie a été réutilisée, etc. ;
- iv) l'ontologie candidate a été publiée dans une publication évaluée par des pairs.

Les répertoires d'ontologies, tels que NCBO BioPortal, OBO Foundry, AberOWL, Protégé library, Swoogle (Ontology Lookup Service de l'Institut européen de bioinformatique),

UMLS qui intègre de nombreuses terminologies et normes de codage, sont utilisés pour trouver des ontologies pertinentes.

#### 4.4.1.2 Cas de PNADO

Deux répertoires d'ontologies biomédicales ont été choisis : Open Biomedical Ontology (OBO) Foundry et BioPortal. La langue anglaise et le format OWL sont requis. En utilisant les critères de recherche, nous avons trouvé 26 ontologies candidates qui pourraient être réutilisées.

#### 4.4.2 Choix des ontologies

Cette étape consiste à sélectionner les ontologies les plus pertinentes à réutiliser. Nous recherchons les ontologies qui traitent chaque thème significatif trouvé dans les exigences définies dans la phase 1. Un thème peut correspondre par exemple aux symptômes de la pneumonie ou aux tests de laboratoire. Le processus de sélection s'inspire du travail de Guarino et al. et utilise des critères d'exactitude et de précision [154]. Avant de définir ces critères, nous devons introduire la notion de modèle voulu tel que défini par Guarino et al. [154]. Habituellement, une ontologie peut avoir de nombreuses interprétations. Afin de limiter le nombre de modèles possibles, une ontologie utilise des axiomes.

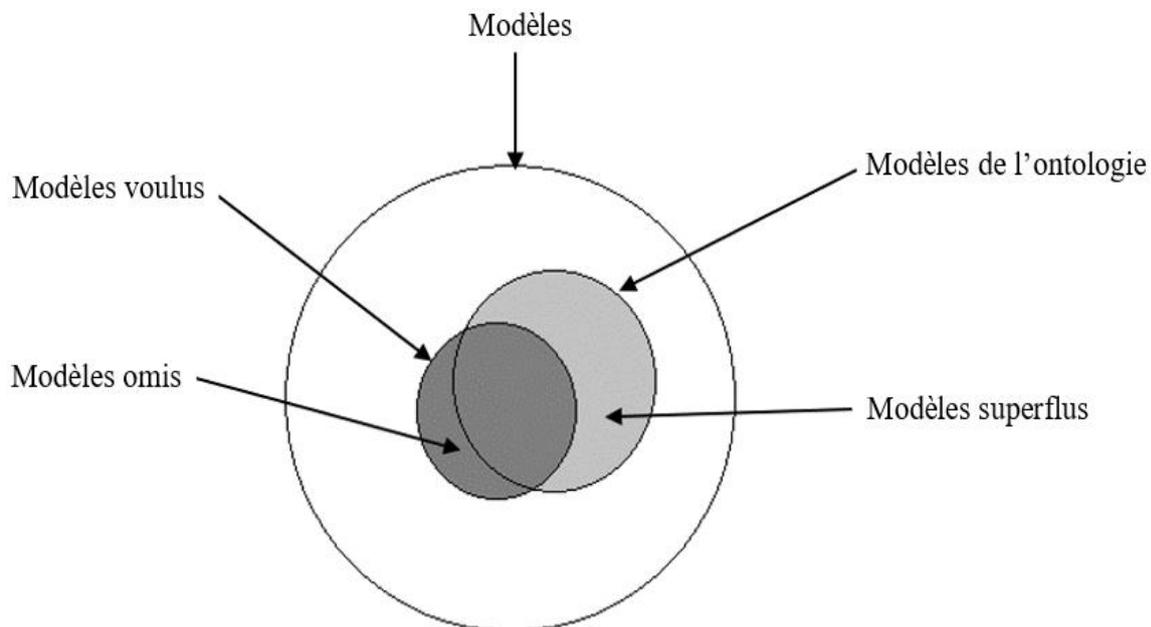
Néanmoins, il est généralement impossible de créer une ontologie idéale, en limitant le nombre de modèles possibles à un minimum, appelés modèles intentionnels. Une ontologie idéale (définie par les exigences) est une ontologie abstraite dont les modèles coïncident avec les modèles voulus. Ils excluent certaines interprétations inexactes sans vraiment décrire la signification du vocabulaire.

Les critères de préférence pour le choix des ontologies pertinentes sont l'exactitude et la précision [155]. Pour comparer la précision des ontologies candidates, deux types de modèles sont considérés :

- i) Les modèles superflus (SUP) qui correspondent à une situation où un aspect du candidat est plus faible que requis, c'est-à-dire qu'ils empêchent certains aspects des exigences d'être satisfaits (voir **Figure 13**) [155] ; par exemple, le modèle du

Coronavirus Infectious Disease Ontology (CIDO<sup>53</sup>) est plus faible que le modèle voulu par PNADO, parce qu'il ne satisfait pas à certaines exigences de l'ontologie définies dans la phase 1 (c'est-à-dire que certains concepts définis dans PNADO ne sont pas couverts par CIDO).

- ii) Les modèles omis (MO) qui correspondent à une situation où un aspect du candidat est plus fort que ce qui est requis, c'est-à-dire que cette ontologie candidate dépasse ce qui a été spécifié dans les exigences (voir **Figure 13**) ; par exemple, le lexique radiologique (RadLex<sup>54</sup>) couvre plusieurs concepts inutiles pour le diagnostic de la pneumonie.



**Figure 13** : La relation entre les modèles voulus d'une ontologie et les modèles des axiomes de l'ontologie (basée sur [155]).

#### 4.4.2.1 Cas de PNADO

En utilisant les critères de précision et d'exactitude, nous identifions les ontologies à réutiliser. Nous avons choisi la BFO qui est plus précise que l'ontologie formelle générale (GFO), car elle comporte moins de modèles omis et superflus. L'OGMS est plus précise que

---

<sup>53</sup> <http://www.obofoundry.org/ontology/cido.html>, dernier accès le 17/03/2021.

<sup>54</sup> <https://bioportal.bioontology.org/ontologies/RADLEX>, dernier accès le 17/03/2021.

Diagnostic Ontology (DIAGONT<sup>55</sup>) qui ne contient que les éléments nécessaires pour effectuer des méta-analyses pour les études diagnostiques. Par conséquent, le PNADO suit la représentation de l'OGMS.

L'ontologie SYMP<sup>56</sup> qui couvre principalement les symptômes et les signes cliniques, est plus précise que l'ontologie des signes et symptômes cliniques (CSSO<sup>57</sup>). Nous avons également choisi Relation Ontology (RO<sup>58</sup>) contenant les relations entre les entités, NCBITaxon<sup>59</sup> qui est une classification et une nomenclature de tous les organismes dans les bases de données publiques de séquences, Computer Based Record Ontology (CPRO<sup>60</sup>) pour la description du profil du patient, Radiological Lexicon (RadLex<sup>61</sup>) qui est une terminologie complète pour la radiologie, Human Phenotype Ontology (HPO<sup>62</sup>) fournit un vocabulaire structuré et contrôlé pour les caractéristiques phénotypiques des maladies humaines héréditaires et autres, et les noms et codes de Logical Observation Identifiers Names and Codes (LOINC<sup>63</sup>) pour l'identification des observations de laboratoire médical. Human Disease Ontology (DOID<sup>64</sup>) est une représentation des maladies humaines organisée par étiologie. Infectious Disease Ontology (IDO<sup>65</sup>) couvre le domaine des maladies infectieuses. Nous réutilisons SNOMED-CT<sup>66</sup> qui est la terminologie la plus utilisée dans le domaine des soins de santé et qui couvre le plus grand nombre de concepts pour représenter les connaissances cliniques [23, 24]. SNOMED-CT ne suit aucune ontologie de niveau supérieur et elle est polyhiérarchique dans la mesure où certains concepts ont plusieurs héritages. Cela entraîne souvent une ambiguïté de la part de l'utilisateur dans la sélection des concepts et une situation désordonnée dans leur classification [25]. SNOMED-CT est utilisée comme une ontologie plutôt qu'une

---

<sup>55</sup> <https://bioportal.bioontology.org/ontologies/DIAGONT>, dernier accès le 17/03/2021.

<sup>56</sup> <http://www.obofoundry.org/ontology/symp.html>, dernier accès le 17/03/2021.

<sup>57</sup> <https://bioportal.bioontology.org/ontologies/CSSO>, dernier accès le 17/03/2021.

<sup>58</sup> <http://www.obofoundry.org/ontology/ro.html>, dernier accès le 17/03/2021.

<sup>59</sup> <http://www.obofoundry.org/ontology/ncbitaxon.html>, dernier accès le 17/03/2021.

<sup>60</sup> <https://bioportal.bioontology.org/ontologies/CPRO>, dernier accès le 17/03/2021.

<sup>61</sup> <https://bioportal.bioontology.org/ontologies/RADLEX>, dernier accès le 17/03/2021.

<sup>62</sup> <http://www.obofoundry.org/ontology/hp.html>, dernier accès le 17/03/2021.

<sup>63</sup> <https://bioportal.bioontology.org/ontologies/LOINC>, dernier accès le 17/03/2021.

<sup>64</sup> <http://www.obofoundry.org/ontology/doid.html>, dernier accès le 17/03/2021.

<sup>65</sup> <http://www.obofoundry.org/ontology/ido.html>, dernier accès le 17/03/2021.

<sup>66</sup> <https://browser.ihtsdo.org/>, dernier accès le 17/03/2021.

terminologie, elle est basée sur la logique de description, et elle peut soutenir le raisonnement [26].

### 4.4.3 Résolution de conflits

Une fois que les ontologies de réutilisation ont été sélectionnées, nous procédons à l'identification des concepts qui enrichiront notre ontologie préliminaire (voir **Figure 14**). Nous souhaitons ajouter, compléter et détailler des concepts déjà représentés dans notre ontologie.

Si le concept pertinent se trouve dans plusieurs ontologies, nous choisissons l'ontologie qui fournit la meilleure représentation du domaine, c'est-à-dire celle qui fournit des réponses aux questions de résolution de conflits (QRC) liées au concept et ces réponses satisfont les exigences de l'ontologie, c'est-à-dire les modèles voulus. En effet, ces QRC sont présentées sous la forme de requêtes informelles et elles sont définies par l'ontologue. Elles sont liées à la signification et à la structure de l'arbre conceptuel de l'ontologie, c'est-à-dire à la hiérarchie des classes, et à d'autres concepts de la logique de description (DL) comme l'intersection et l'union de classes, les classes équivalentes, les classes universelles, la quantification universelle et existentielle, la restriction has-value et la restriction de cardinalité. Voici quelques exemples de QRC :

- ✓ Le concept a-t-il une définition (sous forme d'annotation) ? Si oui, sa définition correspond-elle à un modèle prévu de l'ontologie ?
- ✓ Le concept comporte-t-il une superclasse ? Si oui, est-elle pertinente pour l'ontologie ?
- ✓ Le concept a-t-il des enfants ? Si oui, chacun d'eux est-il pertinent pour l'ontologie ?
- ✓ Le concept a-t-il des synonymes ?
- ✓ Le concept a-t-il des noms alternatifs ?

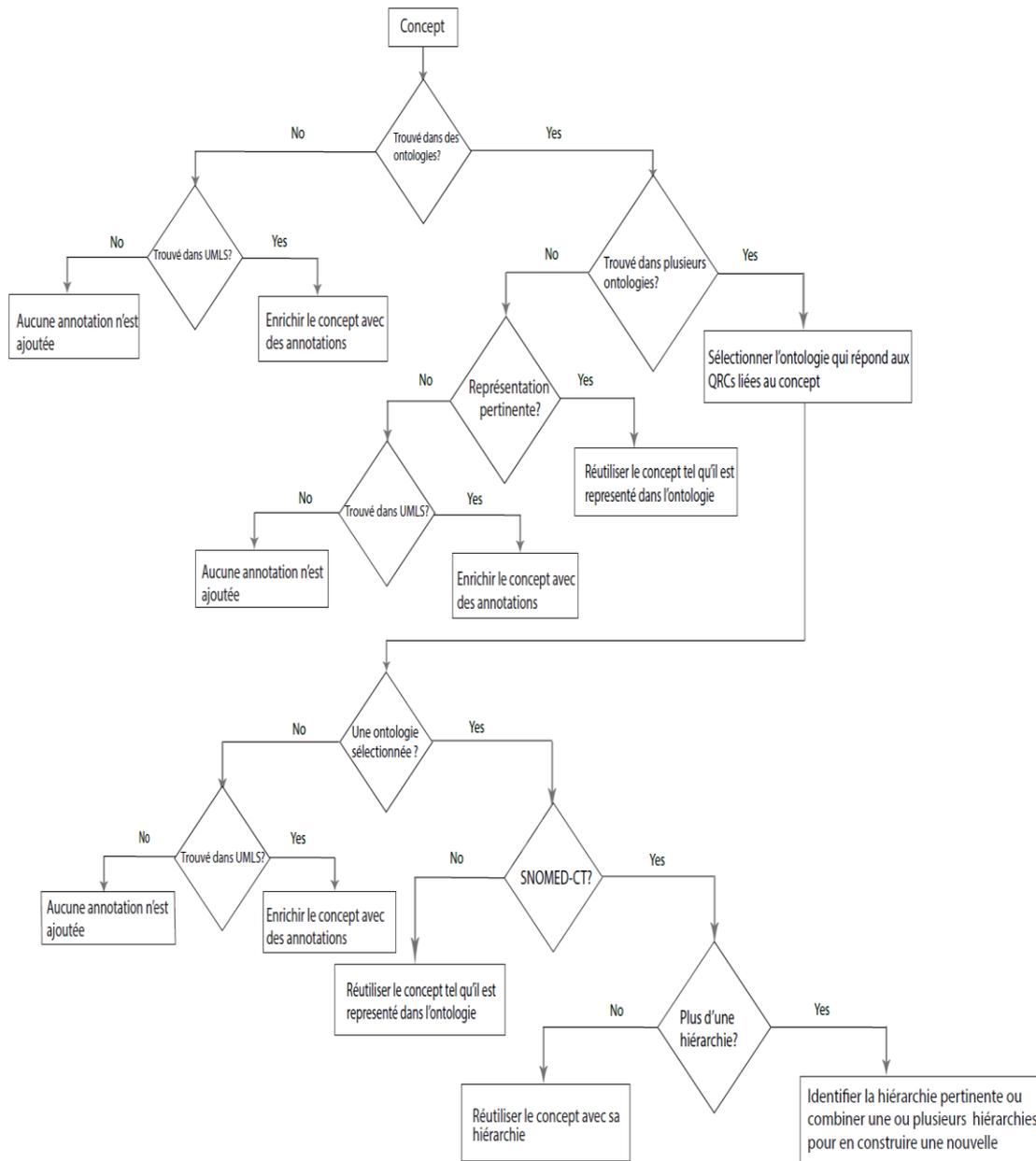


Figure 14 : Approche pour la sélection des concepts.

#### 4.4.3.1 Cas de PNADO

Nous donnons ici quelques exemples d'identification des concepts dans d'autres ontologies. Le concept hypothermia se trouve à la fois dans SYMP et SNOMED-CT. Pour choisir la meilleure représentation de ce concept, l'une des QRC utilisées est : l'hypothermie a-t-elle des concepts enfants ? La réponse est oui pour SNOMED-CT et non pour SYMP (voir **Tableau 5**). Le concept *patient* est représenté différemment dans CPRO et SNOMED-CT. L'une des QRC utilisées est : quel est le concept parent du patient ? Dans CPRO, c'est *organism* qui est

fourni par OGMS. Dans SNOMED-CT, le parent est *un social context* qui ne correspond pas aux exigences d'OGMS. Une autre QRC utilisée, est-ce que toute personne est un patient ? La réponse du CPRO est l'axiome : *human/person and ('plays role' some patient role) and ('participates\_in' some clinical act)*. Aucune réponse n'est trouvée dans SNOMED-CT.

Si le concept à réutiliser est bien représenté dans le SNOMED-CT, c'est-à-dire si sa représentation satisfait les exigences de PNADO, et s'il n'a pas plusieurs classes parentales, nous le réutilisons tel quel ; sinon, nous identifions la hiérarchie la plus pertinente ou nous en combinons deux ou plus pour créer une nouvelle hiérarchie, en tenant toujours compte des exigences de PNADO. Si le concept ne correspond à aucun des concepts des ontologies sélectionnées, nous lui proposons des annotations, telles qu'une définition, des synonymes, des commentaires, etc.

**Tableau 5 : Exemples de conflits résolus avec les QRC.**

Concept	Ontologies	QRC	Réponses aux QRC	
Hypothermia	SNOMED-CT SYMP	- Are there any <i>hypothermia's</i> children?	- SNOMED-CT: <i>hypothermia</i> has six children. - SYMP: <i>hypothermia</i> has no children.	
		- Is each child relevant for pneumonia diagnosis?	- Yes, <i>hypothermia's</i> children provided by SNOMED-CT are relevant for pneumonia diagnosis.	
		- What is the parent of <i>hypothermia</i> ?	- SNOMED-CT: <i>body temperature finding</i> is the parent of <i>hypothermia</i> . - SYMP: <i>neurological and physiological symptom</i> is the parent of <i>hypothermia</i> .	
		<b>Analyse:</b> <i>hypothermia's</i> children in SNOMED-CT correspond to PNADO scope and respond to its intended model. The <i>hypothermia's</i> parent which is <i>neurological and physiological symptom</i> responds to PNADO intended model.		
Hypotension	HPO SNOMED-CT SYMP	- Is <i>hypotension</i> a <i>symptom</i> or a <i>vital sign</i> ?	- HPO: <i>hypotension</i> is a phenotypic abnormality. - SNOMED-CT: <i>hypotension (low blood pressure)</i> is considered as a <i>disorder of the cardiovascular system</i> . - SYMP: <i>hypotension</i> is considered as a <i>hemic system symptom</i> .	
		- What is the synonym of <i>hypotension</i> ?	- <i>Hypotension</i> has two synonyms in HPO, three synonyms in SNOMED-CT, and no synonym in SYMP.	
		- Are there any <i>hypotension's</i> children?	- <i>Hypotension</i> has three children in HPO, twelve children in SNOMED-CT, and one child in SYMP.	
		- Is any child relevant for diagnosis of pneumonia?	- Only <i>hypotension's</i> children provided by HPO and SYMP are relevant for diagnosis of pneumonia.	
		<b>Analyse:</b> <i>blood pressure</i> is a measurement of the pressure in arteries. We consider it in PNADO as a <i>vital sign</i> . <i>Hypotension's</i> children in HPO combined with SYMP's child correspond to PNADO intended model. We reuse <i>hypotension's</i> synonyms of SNOMED-CT.		
Respiratory rate	SNOMED-CT LOINC	- Is <i>respiratory rate</i> a <i>symptom</i> or a <i>vital sign</i> ?	- In both SNOMED-CT and LOINC, it is mentioned that the scale type of <i>respiratory rate</i> is quantitative.	
		- Are there any <i>respiratory rate's</i> children?	- SNOMED-CT: <i>respiratory rate</i> has two children. - LOINC: <i>respiratory rate</i> has no children.	
		- Is each child relevant for diagnosis of pneumonia?	- Yes, <i>respiratory rate's</i> children provided by SNOMED-CT are relevant for diagnosis of pneumonia.	
		- Is there any measurement unit?	- No measurement for <i>respiratory rate</i> is given in SNOMED-CT. - LOINC provides "{breaths}/min" as an example of <i>respiratory rate</i> measurement unit.	
		- Are there any synonyms? - Are there any related names?	- SNOMED-CT provides two synonyms and no related names. - LOINC doesn't provide any synonym for <i>respiratory rate</i> but provides 13 related names.	
		<b>Analyse:</b> <i>respiratory rate</i> is represented as a <i>vital sign</i> in PNADO. We reuse SNOMED-CT's children because they respond to PNADO scope and its intended model. We reuse the unit measurement of LOINC and some of the related names. We also reuse the synonyms of SNOMED-CT.		

## **4.5 Évaluation de l'ontologie**

Nous présenterons dans les prochaines sous-sections l'évaluation de l'ontologie en adaptant certains critères et suivant certaines approches dans le but d'avoir une ontologie de qualité.

### **4.2.1 Approches et critères d'évaluation**

L'évaluation de l'ontologie est une phase intrinsèque du processus de développement d'une ontologie (ainsi que des bases de connaissance en général) et de son cycle de vie (voir section 2.2.2). Elle consiste à aider les ontologues à déterminer si l'ontologie est de bonne qualité. Plusieurs méthodes ont émergé dans le cadre de l'évaluation des ontologies et qui sont reconnues dans plusieurs travaux [158-160]. Quatre principales approches ont été distinguées : 1) les approches qui utilisent l'ontologie dans une application et évaluent les résultats obtenus, appelées basées sur l'application ou la tâche ; 2) les approches qui comparent l'ontologie à un gold standard ; 3) celles qui sont basées sur le corpus ou sur les données d'un domaine ; et 4) celles qui recommandent l'évaluation par les utilisateurs finaux de l'ontologie qui sont des experts du domaine.

Selon Cimino (1998), une ontologie médicale (ainsi que tout autre vocabulaire contrôlé) de bonne qualité devrait avoir les 12 caractéristiques (qu'il réclame comme desiderata) suivantes : le contenu (complétude), l'orientation des concepts, la permanence des concepts, l'identifiant non sémantique du concept, la polyhiérarchie, la définition formelle des concepts, le rejet de « non classé ailleurs », la granularité multiple, les vues multiples consistantes, la représentation du contexte, la redondance reconnue, la maintenance non déstabilisante [161]. Nous allons évaluer notre ontologie selon ces 12 critères. La cohérence interne doit être vérifiée en utilisant l'un des raisonneurs offerts par Protégé.

### **4.5.2 Cohérence interne et respect des pratiques standard de l'ontologie**

Dans cette première étape, nous vérifions la cohérence interne et la conformité totale de l'ontologie aux directives de OBO Foundry et aux desiderata de Cimino [161]. La cohérence interne concerne l'évaluation au niveau syntaxique et peut être obtenue en utilisant les raisonneurs fournis par Protégé.

### **4.5.3 Évaluation avec des sources dépendantes du domaine**

Cette méthode est également appelée méthode basée sur le corpus. Le corpus est constitué de documents qui couvrent de manière significative un domaine donné tel que les GBPC, les revues systématiques et les dossiers médicaux des patients. Les revues systématiques contiennent plus de détails concernant une question clinique donnée par rapport aux GBPC, d'où l'intérêt de leur utilisation. Nous introduisons une méthode d'évaluation de l'ontologie en utilisant l'annotation textuelle. L'annotation textuelle est le processus qui consiste à relier les concepts dans les documents à leurs représentations correspondantes dans l'ontologie. L'objectif de cette méthode est de permettre à l'ontologue de vérifier chaque concept d'un document qui peut se référer au domaine, qu'il soit abordé dans l'ontologie ou non. Les concepts concernés par le processus d'annotation comprennent les adjectifs, les adverbes, les noms propres, les noms, les verbes, les phrases verbales, les phrases nominales, les phrases adjectivales et les phrases participatives. Nous recommandons également de relever le défi d'identifier les relations dans les GBPC et de trouver leurs équivalences dans l'ontologie.

### **4.5.4 Évaluation avec des experts du domaine**

L'évaluation est effectuée par des experts du domaine qui tentent d'évaluer dans quelle mesure l'ontologie répond à un ensemble d'exigences prédéfinies. La technique de l'échelle [162] et des questions de compétence peuvent être utilisées. WebProtégé<sup>67</sup> peut être utilisé pour partager l'ontologie avec les experts du domaine. Nous recommandons d'écrire un document pour expliquer l'ontologie, ses objectifs et ce que l'on attend d'eux. Ce document devrait inclure des questions utilisant la technique de l'échelle.

### **4.5.5 Cas de PNADO**

Malheureusement, l'approche d'évaluation basée sur l'application n'est pas réalisable pour le moment, mais elle sera intégrée dans les travaux futurs. Nous avons l'ambition de l'intégrer au sein de SEKMED<sup>68</sup> (Software for the Evolution of Knowledge in MEDicine) qui est un système d'information médical canadien dont le développement a été entamé au Québec en

---

<sup>67</sup> <https://webprotege.stanford.edu/>, dernier accès le 14/08/2020.

<sup>68</sup> <https://www.sekmed.site/>, dernier accès le 14/08/2020.

2012, quoique le concept de la plateforme ait été présenté en 2007 à un congrès sur l'information de la pratique par son pionnier. Ce système est considéré comme un outil d'appui à la démarche clinique et une base de connaissances médicale évolutive. Il permet à des communautés de pratique de créer et de partager des ressources médicales et de transférer des connaissances. Ces ressources sont des notes, prises d'histoire, examens physiques, établissement d'un diagnostic différentiel, établissement des ordonnances, prescriptions, algorithmes décisionnels, vidéos, images et procédures. Elles sont créées et révisées par les communautés médicales. À l'instar d'un wiki, la fraîcheur des données et des connaissances et le suivi des standards les plus récents sont parmi les points forts de SEKMED. Nous avons un accord de principe avec son fondateur, le Dr Sylvain Croteau qui a participé dans la définition du domaine et la portée de l'ontologie ainsi que dans son évaluation.

Il n'est pas possible non plus d'utiliser un gold standard pour évaluer PNADO, car il n'existe pas d'ontologie de référence qui nous permettrait de faire la comparaison. Nous utilisons un ensemble de données cliniques qui contiennent de multiples cas de pneumonie et un corpus de connaissances cliniques composé de GBPC et de revues systématiques, selon la troisième approche. Les revues systématiques sont considérées comme le type de revue scientifique le plus important, car elles sont fondamentales pour une pratique fondée sur des données probantes [163]. Des experts du domaine sont impliqués dans ce processus d'évaluation. Nous nous considérons chanceux d'avoir des médecins impliqués dans cette phase dans cette période du COVID19.

Nous utilisons le raisonneur Pellet fourni par Protégé pour vérifier la cohérence interne de PNADO, et aucune erreur n'a été détectée. Pour le moment, le raisonneur Konclude n'est pas encore très recommandé, car, entre autres, son utilisation sans faille avec Protégé n'est pas garantie. Nous vérifions la conformité totale de PNADO aux principes d'OBO Foundry et aux desiderata de Cimino.

Outre les GBPC choisis à la phase 4.2, nous avons rassemblé 43 revues systématiques provenant des répertoires de la médecine factuelle nationaux et internationaux. Voici un exemple (voir **Tableau 6**) d'annotation du GBPC "Management of Community-Acquired Pneumonia in Adults" [164] extrait du répertoire américain Infectious Diseases Society of America/American Thoracic Society Consensus Guidelines.

“A *chest radiograph* is required for the routine evaluation of *patients* who are *likely to have pneumonia*, to establish the *diagnosis* and to aid in *differentiating CAP* from other *common causes* of *cough* and *fever*, such as *acute bronchitis*.”

**Tableau 6** : Exemple d'annotation d'un paragraphe issu d'un GBPC avec PNADO.

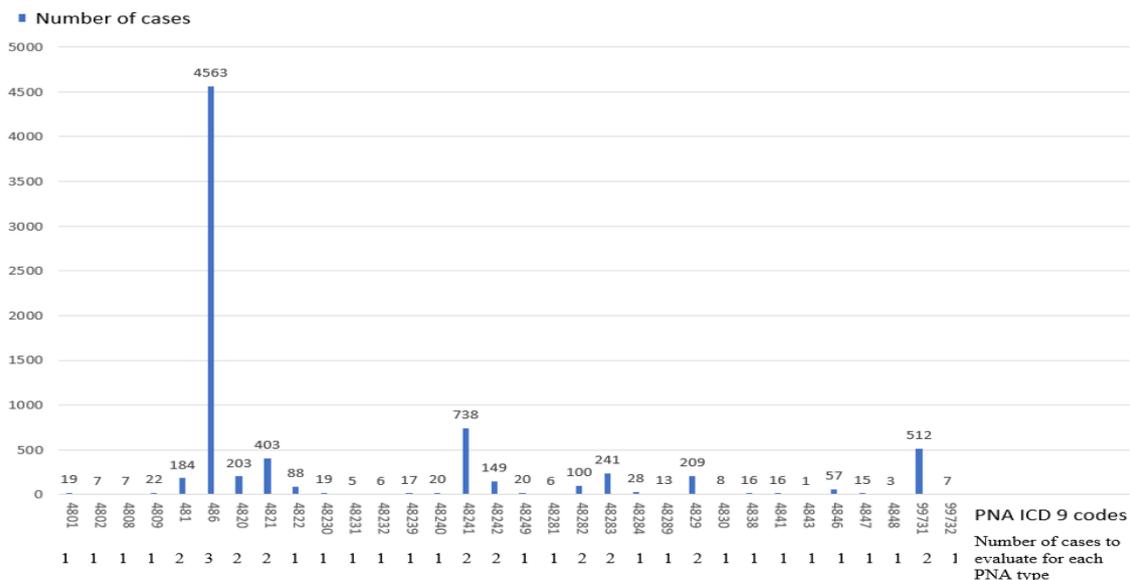
Concepts du texte	Équivalences dans PNADO			
	Classe	Relation	ID	Entité parente dans PNADO
chest radiograph	<i>chest radiography</i>		PNADO:0000783	<i>imaging of lung</i>
patients likely to have pneumonia		<i>has a diagnosis (patients, pneumonia)</i>	PNADO:0001261	
Diagnosis	<i>diagnosis</i>		OGMS:0000073	<i>data item</i>
differentiating CAP from...acute bronchitis		<i>has a differential diagnosis (CAP, acute bronchitis)</i>	PNADO:0001157	
CAP	<i>community-acquired pneumonia</i>		PNADO:0001171	<i>infective pneumonia</i>
Cough	<i>cough</i>		HP:0012735	<i>respiratory system and chest symptom</i>
Fever	<i>fever</i>		PNADO:0000538	<i>neurological and physiological symptom</i>

Nous utilisons les dossiers médicaux des patients qui contiennent les connaissances nécessaires à la pratique médicale dans la vie hospitalière réelle, pour vérifier si PNADO contient la connaissance nécessaire pour le diagnostic de la pneumonie. Nous utilisons la base de données MIMIC III<sup>69</sup> (Multi-parameter Intelligent Monitoring in Intensive Care III) [165]. L'accès à cette base de données a nécessité une formation à distance auprès de l'Institut de Technologie du Massachusetts (MIT). Une formation qui consistait en 8 cours qui portaient sur les données ou spécimens uniquement pour la recherche. Chaque cours a été suivi par un examen. Le score minimum demandé pour avoir accès à cette base de données était de 90%. Suite à un score de 97%, nous avons réussi à avoir accès à cette base de données.

Cette base de données clinique comprend 38 597 patients adultes distincts et désidentifiés et 49 785 admissions à l'hôpital. Elle comprend, entre autres, des données de laboratoire, des rapports de radiologie, des signes vitaux, des profils d'intervention thérapeutique, des réglages de ventilateurs, des notes infirmières, des codes de diagnostic, des résumés de sortie

<sup>69</sup> <http://physionet.org/>, dernier accès le 17/03/2021.

et des données d'entrée d'ordonnances de prestataires. Elle est constituée de 26 tables relationnelles reliées par des identifiants tels que HADM\_ID se rapportant à une admission unique à l'hôpital et SUBJECT\_ID se rapportant à un patient unique. Parmi ces tables, nous considérons la table nommée "Diagnoses\_icd" qui contient les diagnostics des patients issus de la CIM-9 et la table "note events" qui contient toutes les notes concernant les patients depuis leur admission jusqu'à leur sortie, y compris les notes des infirmières et des médecins, les rapports d'ECG (électrocardiogramme), les rapports de radiologie et les résumés de sortie. En interrogeant la table 'diagnoses\_icd' avec une requête SQL, 7702 cas de diagnostic de la pneumonie ont été trouvés. Comme le montre la **Figure 15**, 32 types de pneumonie sont distingués. Nous voulons nous assurer que PNADO couvre tous les cas qui se présentent dans MIMIC-III. Le nombre de cas de pneumonie avec le code 860 représente plus que la moitié de tous les cas. Par conséquent, nous utilisons la fonction logarithmique pour obtenir le nombre de cas à inclure dans l'évaluation pour chaque type de pneumonie, soit 43 cas au total.



**Figure 15** : Cas de pneumonie dans MIMIC-III.

L'étape suivante consiste à extraire les enregistrements de la table 'noteevents' en utilisant HADM\_ID déjà obtenu à l'étape précédente. Nous sélectionnons les cas au hasard en fonction du nombre obtenu pour chaque type et nous complétons l'annotation.

Concernant les médecins impliqués dans l'évaluation de PNADO, une pneumologue de l'hôpital Charles-Le Moyne<sup>70</sup> de Montréal, un urgentologue de l'hôpital de Gatineau<sup>71</sup> et un médecin de famille de la clinique médicale de l'UQO (Université du Québec en Outaouais)<sup>72</sup> ont accepté d'évaluer PNADO. Nous leur avons remis un document (voir Annexe A) pour leur expliquer l'ontologie, son objectif et ce que l'on attendait d'eux. Nous avons également préparé des questions en utilisant la technique de l'échelle [162]. Pour partager PNADO avec eux, des comptes dans WebProtégé ont été créés. Une fois qu'ils se sont familiarisés avec l'environnement, ils ont utilisé des questions de compétences pour l'évaluation et ont commenté directement PNADO sur WebProtégé.

## 4.6 Documentation

Une ontologie doit être documentée pour une meilleure compréhension et utilisation des futurs utilisateurs, et pour qu'elle soit facilement maintenue et mise à jour par n'importe quel ontologue. Dans notre méthodologie, la documentation est en réalité entamée lors de la première phase, lors de la définition du domaine et de la portée de l'ontologie. À la fin de phase, un document qui détaille son déroulement ainsi que ses livrables est établi. Dans cette phase, il faudrait rajouter la documentation de l'ontologie, c'est-à-dire ses classes, relations, propriétés, instances, axiomes et annotations. Pour avoir une documentation lisible, il existe des outils qui peuvent la générer à partir d'une ontologie. Parmi ces outils, OWLDoc<sup>73</sup> qui exporte l'ontologie vers HTML et offre une vue dynamique qui permet de naviguer à l'intérieur de l'ontologie. Il organise la documentation de chaque entité ontologique dans différentes sections, montrant sa taxonomie, son usage dans l'ontologie et ses axiomes. OWLDoc est intégré dans l'éditeur protégé et son utilisation est relativement simple. LODÉ<sup>74</sup> (Live OWL Documentation Environment) [166] est un autre outil qui génère automatiquement une description du fichier OWL en prenant en compte les axiomes et les annotations. Cette documentation est présentée à l'utilisateur sous la forme d'une page HTML avec des liens intégrés pour faciliter la navigation.

---

<sup>70</sup> <https://santemonteregie.qc.ca/installations/hopital-charles-le-moyne>, dernier accès le 17/03/2021.

<sup>71</sup> <https://ciss-ouaouais.gouv.qc.ca/>, dernier accès le 17/03/2021.

<sup>72</sup> <https://ssuqo.ca/>, dernier accès le 17/03/2021.

<sup>73</sup> <https://protege.wiki.stanford.edu/wiki/OWLDoc>, dernier accès le 17/03/2021.

<sup>74</sup> <https://essepuntato.it/lode/>, dernier accès le 17/03/2021.

## 4.6.1 Cas de PNADO

Nous avons utilisé OWLDoc pour documenter PNADO. Nous avons obtenu une documentation dont nous montrons quelques illustrations dans l'Annexe B.

## 4.7. Maintenance et évolution

Les connaissances médicales (ainsi que d'autres domaines) évoluent et il faut en prendre compte. En l'occurrence, une ontologie médicale est appelée à être évoluée et à être mise à jour pour faire face aux incomplétudes, tout en préservant la consistance et la cohérence. D'abord, les changements à effectuer doivent être spécifiés, c'est-à-dire : 1) le type des entités concernées (classe, relation, propriété, instance) ; 2) les informations nécessaires pour la gestion du changement ; 3) les conditions à satisfaire pour que le changement puisse être appliqué ; et 4) les inconsistances potentielles qui peuvent être causées par ce changement. Ces changements doivent être en concordance avec le domaine et la portée de l'ontologie. Une fois que les changements sont appliqués à l'ontologie, celle-ci doit être évaluée pour vérifier sa consistance.

En plus des nouvelles connaissances qui peuvent nécessiter leur intégration dans l'ontologie (à partir de nouveaux GBPC, revues systématiques, etc.), il ne faut pas omettre les concepts qui proviennent des ontologies réutilisées. Ces derniers peuvent être supprimés ou modifiés sans avoir une incidence sur les ontologies qui les réutilisent. Malheureusement, il n'y a aucun moyen automatisé de le faire à part celui de vérifier fréquemment ces ontologies.

### 4.7.1 Cas de PNADO

Dans le contexte de la COVID19, nous avons trouvé deux nouveaux GBPC, dans NICE et CEBM<sup>75</sup> (the centre for Evidence-Based Medicine), traitant le diagnostic de pneumonie. Nous avons extrait 16 nouveaux concepts pertinents pour PNADO.

OGMS a fait l'objet de plusieurs mises à jour qui ont eu un impact sur PNADO. Par exemple, *symptom* a été déplacée dans la classe *process*, *vital sign* a été déplacée dans la classe

---

<sup>75</sup> <https://www.cebm.net/>, last access 17/03/2021.

*material entity, sign* a été supprimé. Nous avons reflété tous ces changements au niveau de PNADO.

## **4.8 Conclusion**

Nous avons présenté dans ce chapitre notre méthodologie de construction d'une ontologie médicale. Elle est constituée de sept phases dont certaines sont composées des étapes détaillées. L'objectif de cette méthodologie est d'offrir à l'ontologue les directives nécessaires qui l'aideront durant tout le processus, en réduisant l'intervention des médecins. Cette dernière est intrinsèque pour la définition du domaine et la portée de l'ontologie ainsi que son évaluation. Nous avons appliqué la méthodologie à la construction de PNADO pour le diagnostic de la pneumonie. Nous présentons et discutons les résultats dans le chapitre qui suit.

## CHAPITRE 5 RÉSULTATS ET DISCUSSION

Alors que les connaissances médicales continuent de s'accroître et que leur gestion devient de plus en plus difficile pour les cliniciens, des solutions telles que la représentation appropriée des connaissances à l'aide d'ontologies deviennent de plus en plus importantes. La construction d'une ontologie a toujours été un défi, que ce soit à partir d'un texte ou en réutilisant des ontologies existantes. En effet, bien que plusieurs méthodes ont été proposées, elles ne fournissent pas assez d'indications sur la manière dont les différentes étapes doivent être réalisées. De plus, pour construire une ontologie médicale, il est nécessaire d'avoir des compétences en médecine et en informatique, ce qui représente un sérieux défi. C'est la raison pour laquelle la plupart des ontologies médicales existantes présentent des lacunes. Nous proposons dans ce travail une méthodologie pour construire une ontologie médicale de qualité. La méthodologie proposée a été appliquée avec succès à la construction d'une ontologie couvrant le diagnostic de la pneumonie PNADO. Les experts du domaine n'ont été sollicités que deux fois : lors de la première phase pour déterminer le domaine et la portée de l'ontologie et lors de la phase de l'évaluation de l'ontologie.

À la fin de la première phase, l'ontologue a une vision claire de ce que l'ontologie couvrira. Dans notre cas d'étude, le domaine est le diagnostic de la pneumonie, et l'ontologie devrait couvrir les domaines suivants : symptôme, signe, processus de diagnostic, diagnostic, patient (antécédents, etc.), résultat clinique et agent pathogène. PNADO sert d'ontologie d'application qui saisit la terminologie contrôlée pour le diagnostic de la pneumonie ainsi que les connaissances du domaine communément utilisées dans le milieu hospitalier dans le but d'une réutilisation et d'une interopérabilité avec d'autres ontologies médicales.

La deuxième phase, qui consiste à construire le corpus de connaissances et à en extraire les termes, est relativement simple, car il suffit de cibler les référentiels, les GBPC et les termes pertinents. PNADO est créé à partir d'un corpus de GBPC afin de garantir des preuves de haute qualité en suivant les recommandations actuelles, sans aucune implication des experts du domaine.

Dans la troisième phase, la méthodologie ARCHONTE a été appliquée pour construire la version PNADO préliminaire à partir du texte. Les engagements ontologiques et sémantiques recommandés (phase 3) ont été très utiles à chaque étape du développement. À la fin de la

phase 3, qui a été coûteuse en termes de temps, PNADO préliminaire contenait les concepts les plus importants pour diagnostiquer la pneumonie. Comme ces concepts avaient besoin d'être enrichis par d'autres ressources, nous avons choisi de réutiliser les ontologies existantes.

Au cours de la quatrième phase, les ontologies médicales existantes ont été identifiées (section 4.4.1), les plus pertinentes ont été choisies pour être réutilisées (section 4.4.2) et les conflits entre les différentes représentations des concepts ont été résolus (section 4.4.3). Durant cette phase 4, il est démontré que l'étape de la recherche des ontologies (section 4.4.1) est décisive pour le développement ultérieur de la méthodologie. L'omission d'obtenir des ontologies appropriées aura un impact sur le choix des ontologies appropriées à réutiliser (section 4.4.2). L'étape du choix des ontologies ne doit pas être prise à la légère. C'est un véritable défi, car toutes les ontologies existantes et publiées ne sont pas nécessairement fonctionnelles, au moins pour l'ontologie que nous voulons construire (comme démontré dans la section 4.4). Certaines ontologies sont exclues à première vue, mais d'autres nécessitent une analyse approfondie. Après cette étape, un autre défi est relevé. La représentation de chaque concept à réutiliser doit être examinée en raison des différentes représentations possibles dans les différentes ontologies. Le **Tableau 7** résume les améliorations apportées à METHODOLOGY au niveau de chaque phase.

**Tableau 7 : Résumé des améliorations apportées à METHONTOLOGY.**

<b>Phases de METHONTOLOGY</b>	<b>Améliorations</b>
Phase 1	-
Phase 2	Nous recommandons la construction du corpus de connaissances en utilisant la source de connaissances la plus fiable qui consiste en une connaissance codifiée (GBPC, revues systématiques, etc.)
Phase 3	Nous recommandons l'utilisation de la méthodologie ARCHONTE et de l'application des principes différentiels.
Phase 4	L'implémentation ou l'opérationnalisation est incluse dans la précédente phase (d'ARCHONTE).
Phase 5	Nous proposons des outils pour choisir les ontologies à réutiliser. Nous recommandons que l'ingénieur des connaissances utilise les QRC pour résoudre les conflits entre les représentations des concepts.
Phase 6	Nous détaillons la documentation.
Phase 7	Nous détaillons l'évaluation.
-	Nous ajoutons la phase de maintenance et d'évolution.

Le **Tableau 8** présente les statistiques sur les concepts réutilisés. PNADO est construit selon les principes d'OBO Foundry, en utilisant un OGMS qui suit le paradigme BFO. OGMS fournit un ensemble de classes de référence générales liées aux maladies, aux patients et aux diagnostics. PNADO est uniquement axé sur le diagnostic. Par conséquent, seuls les concepts OGMS liés au diagnostic sont utilisés. D'autres termes OGMS de haut niveau pourraient être utilisés dans des extensions ultérieures. Nous avons remarqué que SNOMED-CT était fréquemment impliquée dans la résolution de conflits, car il chevauche avec la plupart des ontologies médicales existantes.

**Tableau 8 : Réutilisation d'autres ontologies dans PNADO.**

	Ontologie	Utilisation dans PNADO	Classes	Relations	Total	Conflits
<b>Réutilisation rude</b>						
1	BFO	Ontologie de niveau supérieur	35	5	40	0
2	OGMS	Ontologie noyau	149	0	149	0
<b>Réutilisation partielle</b>						
1	SYMP	Symptômes	57	0	57	22
2	NCBITAXON	Virus et bactéries	184	0	184	0
3	RO	Relations	0	22	22	0
4	CPRO	Rôles dans le domaine des soins de santé	4	0	4	2
5	LOINC	Laboratoire	26	0	26	1
6	SNOMED_CT	Maladies et signes cliniques	796	0	796	115
7	RADLEX	Imagerie	55	0	55	5
8	DOID	Maladies	35	0	35	13
9	HPO	Symptômes et signes cliniques	96	0	96	73
10	IDO	Pathogènes	11	0	11	0

La phase d'évaluation utilise différentes approches et implique des experts du domaine. L'évaluation par ces derniers est essentielle, car elle permet de valider la méthodologie proposée. PNADO est cohérente et adhère aux desiderata de Cimino suivants : identifiant de concept non sémantique, orientation du concept, polyhiérarchie, définitions formelles, reconnaissance de la redondance, représentation du contexte, rejet des "non classés ailleurs" et granularités multiples.

L'évaluation à l'aide d'un corpus constitué de GBPC et de revues systématiques a nécessité la vérification de la représentation de chaque concept du corpus, lié au diagnostic de la

pneumonie. La fin de cette étape a révélé que tous les concepts annotés se trouvaient dans PNADO.

## 5.1 Caractéristiques de PNADO

À la fin de l'évaluation de PNADO avec la base de données clinique MIMIC III, 36 concepts sur 988 n'étaient pas couverts par PNADO. L'analyse de chaque concept a révélé que 16 concepts sont des annotations, et 20 d'entre eux sont des extensions des classes existantes. Voici quelques annotations qui comprennent des étiquettes (labels) et des synonymes alternatifs (alternative labels) : pulmonary embolus, rhonchi sound, rhonchus sound, pneumococci, pneumonic infiltrates, MSSA PNA, MSSA pneumonia, MRSA PNA, MRSA pneumonia, et trouble breathing. Les nouvelles classes incluent les types de pneumonie à savoir lobar pneumonia, multilobar pneumonia, polymicrobial pneumonia, multifocal pneumonia, bilateral pneumonia ; les résultats de l'examen physique comme respiratory failure ; les symptômes comme recurrent cough et occasional cough ; et l'historique clinique comme tracheostomy.

Les médecins impliqués dans l'évaluation ont suggéré une dizaine de modifications : *radiological signs of consolidation*; déplacer *fever* à *general symptom*, *purulent tracheobronchial secretions* à *respiratory system and chest symptom*, *headache* à *neurological and physical symptom*; supprimer *transitory tachypnea of the newborn*, *sputum eosinophilia* et *lithophytic*.

Les résultats de l'évaluation sont satisfaisants et prouvent que notre méthodologie est efficace. PNADO dans sa dernière version contient 1598 classes (1448 sont réutilisées et 150 nouvelles), 42 relations (27 sont réutilisées et 15 sont nouvelles), 1951 axiomes, et 83 annotations.

La **Figure 16** montre un exemple de PNADO montrant la hiérarchie de la classe *infective pneumonia* et, en particulier, la classe *viral pneumonia*, ses annotations et ses axiomes. La **Figure 17** illustre la hiérarchie des symptômes avec comme exemple la classe *respiratory system and chest symptom*.

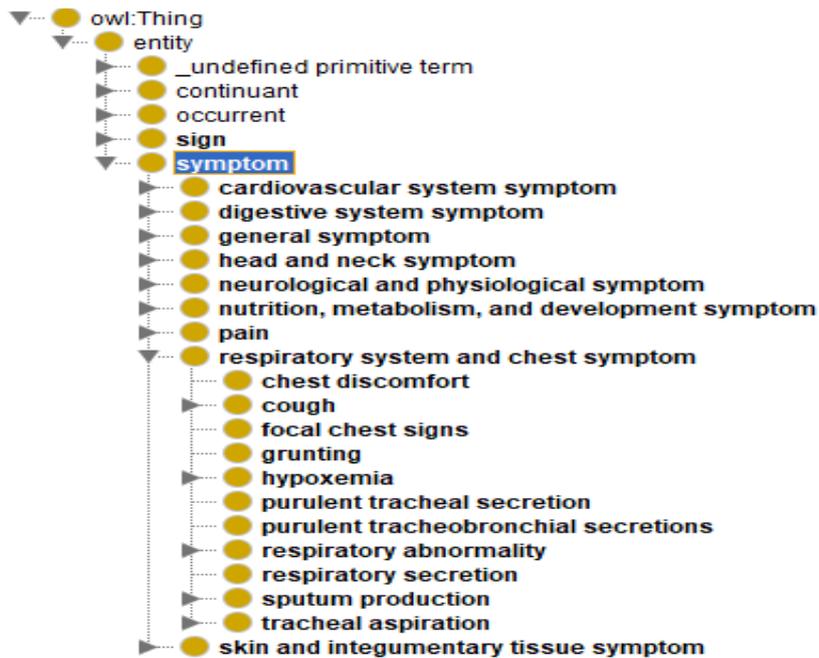
Les relations sont présentées dans le **Tableau 9**. Par exemple, la relation *causes condition* relie la classe *pathogen* avec la classe *infectious disease* et elle est définie par l'existence d'un lien de causalité entre la classe *pathogen* et la classe *infectious disease*.

**Tableau 9** : Quelques relations de PNADO.

Object property	Domain class	Range class
<i>Has symptom</i>	<i>Patient</i>	<i>Symptom</i>
<i>Has diagnosis</i>	<i>Patient</i>	<i>Diagnosis</i>
<i>Has clinical finding</i>	<i>Patient</i>	<i>Clinical findings</i>
<i>Has sign</i>	<i>Patient</i>	<i>Physical sign</i>
<i>Detect</i>	<i>Antigens tests</i>	<i>Pathogens</i>
<i>g depends on</i>	<i>Generically dependent continuant</i>	<i>Independent continuant</i>
<i>s depends on</i>	<i>Specifically dependent continuant</i>	<i>Independent continuant</i>
<i>Precedes</i>	<i>Process</i>	<i>Process</i>
<i>Has participant</i>	<i>Occurrent</i>	<i>Continuant</i>
<i>Participates in</i>	<i>Continuant</i>	<i>Occurrent</i>
<i>Occurs in</i>	<i>Independent continuant</i>	<i>Occurrent</i>
<i>Complication of</i>	<i>Disease</i>	
<i>Differential diagnosis of</i>	<i>Diagnosis</i>	<i>Diagnosis</i>
<i>Causes condition</i>	<i>Pathogen</i>	<i>Infectious disease</i>
<i>Is diagnosis of</i>	<i>Diagnosis</i>	<i>Disease</i>
<i>Finding of</i>	<i>Physical sign</i>	<i>Physical examination</i>

The screenshot displays an ontology editor interface. On the left, a tree view shows the hierarchy of classes under 'Infective pneumonia', with 'Viral pneumonia' highlighted. The right pane shows the details for the selected class 'Viral pneumonia'. It includes annotations such as 'rdfs:label [language: en] Viral pneumonia' and 'id [type: xsd:string]'. The 'Description' field contains 'Viral pneumonia'. Under 'Equivalent To', there is a plus sign. Under 'SubClass Of', 'Infective pneumonia' is listed. Under 'General class axioms', there is a plus sign. Under 'SubClass Of (Anonymous Ancestor)', a complex axiom is shown: 'Complication of disease' some ('Systemic inflammatory response syndrome' or Shock or osteomyelitis or 'Septic arthritis' or Arrhythmia or Endocarditis or Sepsis). The bottom of the right pane shows 'Instances' and 'Target for Key' sections, both with plus signs.

**Figure 16** : La classe *infective pneumonia* avec *viral pneumonia* comme exemple de sous-classe.



**Figure 17** : La classe *symptom* avec *respiratory system and chest symptom* comme exemple de sous-classe.

La **Figure 18** présente la classe *diagnostic process* ainsi que les relations avec d'autres classes dans PNADO.

La **Figure 19** illustre la structure supérieure de PNADO. Les principales classes de PNADO sont représentées sous les classes de niveau supérieur les plus appropriées de BFO et OGMS. Par exemple, la classe *symptom* (9) est une classe OGMS qui est alimentée par de nouvelles classes PNADO et des classes réutilisées de SYMP, HP et SNOMED-CT.

PNADO est publiée dans le répertoire de BioPortal. Elle est accessible via ce lien : <https://bioportal.bioontology.org/ontologies/PNADO>.

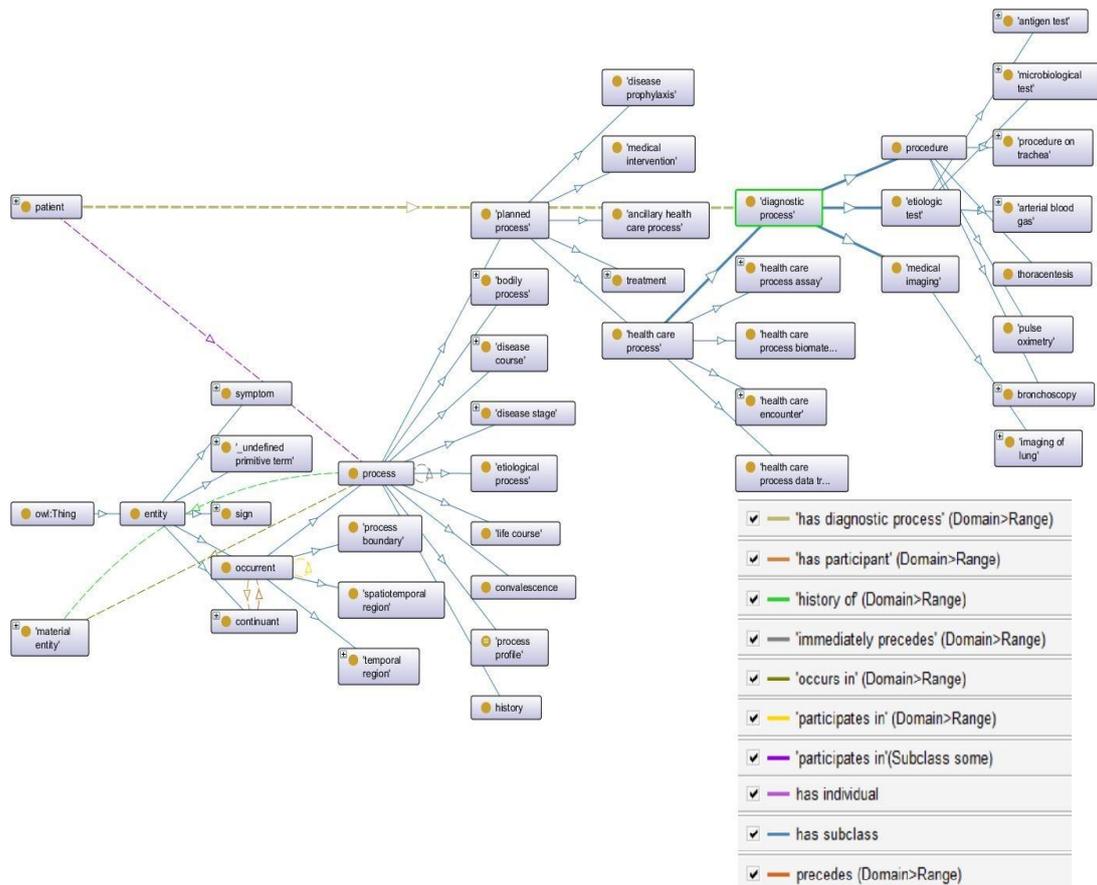
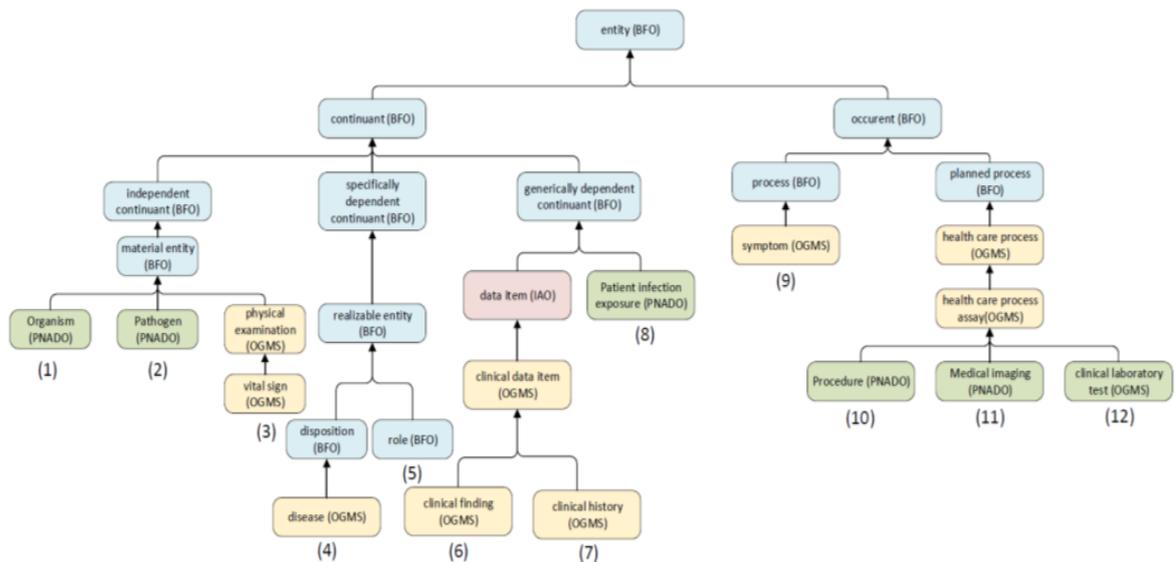


Figure 18 : Représentation de la classe *diagnostic process* dans Protégé.



- |                                       |                                |
|---------------------------------------|--------------------------------|
| (1) NCBITAXON, SNOMED-CT              | (7) PNADO                      |
| (2) IDO                               | (8) PNADO                      |
| (3) LNC, SNOMED-CT                    | (9) SYMP, HP, SNOMED-CT, PNADO |
| (4) DOID, SNOMED-CT, HP, PNADO        | (10) SNOMED-CT, PNADO          |
| (5) CPRO                              | (11) Radlex, PNADO             |
| (6) LNC, HP, SNOMED-CT, Radlex, PNADO | (12) SNOMED-CT, PNADO          |

Figure 19 : les entités fondationnelles de PNADO.

## 5.2 Exemples de problèmes liés à l'intégration avec d'autres ontologies médicales

Dans ce qui suit, nous passerons à travers quelques nouveaux concepts (*symptom* et *pathogen*) et relations (*complication of*, *has symptom*, *is diagnosis of*, *differential diagnosis of* et *finding of*) ajoutés à PNADO.

### **symptom**

*symptom* est défini dans OGMS comme "a quality of a patient that is observed by the patient or a processual entity experienced by the patient, either of which is hypothesized by the patient to be a realization of a disease". Il est aussi défini dans SYMP comme "a perceived change in function, sensation, loss, disturbance or appearance reported by a patient indicative of a disease". Néanmoins, lors de la réutilisation du SYMP, nous avons remarqué qu'il contient également des maladies et des signes cliniques alors qu'il était destiné à ne contenir que des symptômes. Par exemple, *hypotension*, *arythmia*, *bronchitis*, *endocarditis* et le *shock* sont définis dans le SYMP comme des symptômes, alors qu'il s'agit de maladies. En effet, il existe souvent une confusion entre les signes cliniques, les maladies et les symptômes dans SYMP et d'autres ontologies. Dans PNDAO, la classe *symptom* inclut uniquement ce qui est rapporté par le patient.

### **pathogen**

Avant d'aller plus loin dans l'analyse du concept *pathogen*, nous allons analyser le concept *organism* car les deux concepts sont liés. Ce dernier est utile pour PNADO et il est représenté dans plusieurs ontologies comme CPRO, IDO, et SNOMED-CT. Dans SNOMED-CT, il est représenté comme une entité indépendante qui inclut les entités *bacteria*, *virus*, *archaea*, *eukarya*, et *prion*. La signification de ce concept peut uniquement être déduite de ses sous-classes, car aucune définition n'est fournie. CPRO représente *organism* comme une sous-classe de *object* alors que IDO le représente comme classe adjacente à *object*. La définition de *object* telle que fournie par BFO est :

"Is a material entity that is: 1) spatially extended in three dimensions; 2) causally unified, meaning its parts are tied together by relations of connection in such a way that if one part of the object is moved in space,

then its other parts will likely be moved also; and 3) maximally self-connected".

Les définitions données par les trois ontologies sont presque similaires, et elles signifient que *organism* est un système individuel vivant qui peut être unicellulaire ou composé de plusieurs milliards de cellules comme les humains. Cela nous amène à conclure que *organism* est définitivement un *object*, contrairement à la représentation fournie par IDO.

### **complication of**

La complication en médecine est un problème médical qui survient pendant ou après une maladie, une procédure, un traitement ou une fonction (la grossesse, par exemple). Ce concept est représenté comme une classe dans plusieurs ontologies comme SNOMED-CT ou Radlex. Dans SNOMED-CT, la complication est une sous-classe de la maladie et représente des complications de la maladie telles que "*complications due to diabetes mellitus, a complication due to Crohn's disease*". Ces complications sont également représentées sous forme de sous-classes de la classe maladie. Une telle représentation crée une chaîne d'héritage *disease* → *complication* → *disease* et, ne couvre pas tous les cas de complications dues à *la maladie, procédure, traitement, et fonction*. Nous proposons de représenter la complication sous forme de relation nommée *complication of* avec quatre sous-propriétés : 1) *complication of the disease* qui relie *disease* à *disease* ; 2) *complication of procedure* relie *procedure* à *disease* ; 3) *complication of treatment* qui relie *treatment* à *disease* ; et 4) *complication of function* qui relie *function* à *disease*. Des axiomes sont progressivement ajoutés lorsque cela s'avère nécessaire pour représenter des complications. Par exemple, dans PNADO, des axiomes sont ajoutés pour illustrer les complications de la pneumonie. Cette façon de représenter les complications améliore l'interopérabilité et la réutilisabilité.

### **has symptom**

RO représente *has symptom* et le définit comme "a relation that holds between a disease or an organism and a phenotype". Nous constatons qu'il s'agit d'une sous-propriété de la relation *has phenotype*. Cette dernière relie les domaines *generically dependent continuant* ou *material anatomical entity* ou *disease* au range *phenotype*. La classe *phenotype*, qui est réutilisée par RO à partir de l'ontologie Combined Phenotype Ontology (UPHENO), est définie comme "a defect or loss of some anatomical structure or a biological process to wild-

type". La classe *phenotype* de OGMS réutilisée dans PNADO est définie comme "a (combination of) quality(ies) of an organism determined by the interaction of its genetic make-up and environment that differentiates specific instances of a species from other instances of the same species". Il est évident que les deux définitions sont différentes, celle fournie par UPHENO ne correspond pas au modèle voulu par PNADO, et par conséquent, nous ne réutilisons pas une relation associant cette classe à une autre classe. Nous créons une nouvelle relation *has a symptom* reliant les deux classes *patient* et *symptom*.

### **is diagnosis of**

*is about* es une nouvelle relation que nous introduisons pour relier les deux classes *diagnosis* et *disease* qui proviennent d'OGMS. En effet, OGMS n'a mis en place aucune relation entre les deux classes et cela a engendré une mauvaise réutilisation de l'ontologie en ajoutant des sous-classes à la classe *diagnosis* du genre *diagnosis of pneumonia*, *diagnosis of bronchitis*, etc., sachant que *pneumonia* et *bronchitis* sont des sous-classes de la classe *disease*. Pour éviter cette pratique, nous proposons la relation *is diagnosis of*, nous lui donnons une définition "a relation that holds between a diagnosis A and a disease B", et attribuons un identifiant.

### **differential diagnosis of**

*differential diagnosis*, ou diagnostic différentiel, sert à distinguer une maladie des autres maladies qui présentent des caractéristiques cliniques similaires. Ces derniers sont les symptômes et les signes cliniques. SNOMED-CT et LOINC le représentent comme une classe, mais le diagnostic différentiel de la pneumonie n'y est pas représenté. En effet, dans ces ontologies, il n'y a pas de représentation de la liste des diagnostics différentiels, ce qui pourrait aider à éviter les erreurs de diagnostic. Dès lors, nous proposons une nouvelle relation qui relie un diagnostic "A" et un diagnostic "B" s'ils présentent des caractéristiques cliniques similaires. Rappelons qu'un diagnostic est la représentation de la conclusion d'un processus de diagnostic qui peut être une maladie ou un syndrome. Voici un exemple d'axiomes ajoutés pour définir cette relation : *infective pneumonia 'differential diagnosis of some bronchitis*.

### **finding of**

OGMS dispose de la classe *physical sign* et de la classe *physical examination*, mais aucune relation n'est définie entre elles. Dans la pratique clinique, le signe clinique est détecté par le clinicien lors de l'examen physique et si OGMS est reprise telle qu'elle, il y aurait une prolifération de concepts. *Physical sign* est défini dans OGMS comme "an abnormal material entity that is part of a patient and hypothesized to be clinically relevant", et *physical examination* est un processus. De ce fait, nous établissons une relation entre les deux classes.

### 5.3 Aspect temporel dans PNADO

L'aspect temporel est pris en compte dans PNADO. Cet aspect n'est pas très présent dans les ontologies médicales existantes, à part BFO. Il concerne des données temporelles, et aussi la séquence des événements ou des processus. BFO prend cet aspect en considération avec des classes et des relations. Nous constatons également qu'aucune ontologie ne s'intéresse à représenter cet aspect alors qu'il serait utile qu'elle soit spécialisée dans la représentation de la temporalité dans le domaine médical et qui serait réutilisée au besoin. Concernant PNADO, nous réutilisons par exemple la relation *precedes* de BFO qui relie *process "A"* au *process "B"*. *season* est une nouvelle classe reliée à l'aspect temporel. La saison où le patient manifeste ses symptômes est très importante et elle est absente dans les ontologies réutilisées ainsi que d'autres qui se spécialisent notamment dans les maladies respiratoires et infectieuses comme l'ontologie COVID-19<sup>76</sup> ou IDO. En effet, parmi les QC définies lors de la définition du domaine et de la portée de l'ontologie celle qui concerne la saison où le patient ressent les symptômes. C'est une information qui aide les médecins dans leur processus de diagnostic. Nous créons la classe *season*, une sous-classe de *patient infection exposure*.

Aux fins de l'interopérabilité et de la réutilisation, tous les concepts PNADO sont annotés avec des identifiants standards formatés selon les recommandations d'OBO Foundry, des synonymes, CUI (Concept Unique Identifier), et des définitions collectées également à partir de l'UMLS et des ontologies réutilisées, lorsque ces informations sont disponibles.

---

<sup>76</sup> <https://bioportal.bioontology.org/ontologies/COVID-19>, dernier accès le 17/03/2021.

# CONCLUSIONS ET PERSPECTIVES

La construction d'une ontologie médicale (ou d'un autre domaine) est en soi un art qui nécessite des techniques et des indications bien précises pour réaliser une belle œuvre. De surcroît, elle requiert une expertise dans l'ingénierie des connaissances et une autre dans le domaine médical (ou d'un autre domaine), ce qui n'est toujours pas évident. Pour construire une ontologie médicale de qualité, une méthodologie munie de l'orientation nécessaire à chaque étape est primordiale, et cela fait défaut aux méthodologies existantes. Cette direction permet, entre autres, de choisir les ressources nécessaires pour la construction et minimiser l'implication des experts du domaine. La question que nous nous sommes posée au début de cette thèse était : comment parvenir, en tant qu'ontologue, ayant des connaissances très minimes dans le domaine médical, à construire une ontologie médicale de qualité, et qui peut être réutilisable et interopérable avec d'autres ontologies, sans trop solliciter les experts du domaine à modéliser ?

**La première contribution** de cette thèse est une réponse à cette question. Elle concerne la conception d'une nouvelle méthodologie détaillée de construction d'ontologies médicales. Cette méthodologie s'inspire de METHONTOLOGY et fournit les directives nécessaires à l'ontologue et permet de minimiser l'implication des experts du domaine. Elle combine l'ingénierie des ontologies à partir de zéro et la réutilisation de celles existantes. Elle est composée de sept phases allant de la définition du domaine et de la portée de l'ontologie jusqu'à sa maintenance et son évolution. Pour minimiser l'intervention des experts du domaine, la méthodologie utilise les GBPC (ou une autre source de connaissance codifiée, structurée) basés sur la médecine factuelle et qui fournit un vocabulaire standard pour aider à intégrer des sources de données biomédicales et hétérogènes. En outre, elle reprend ARCHONTE pour la construction d'une ontologie préliminaire grâce à ses principes différentiels qui éclairent la lanterne du sens de chaque concept et de son emplacement dans l'ontologie. Contrairement aux méthodologies de construction existantes, elle fournit toutes les étapes nécessaires pour construire une ontologie cohérente, y compris la résolution de conflits, lorsqu'un concept est différemment représenté dans plusieurs ontologies. Cela est réalisé grâce à l'utilisation des QRC. Cette méthodologie a aussi mis en lumière sur la manière de construire son corpus de textes, de choisir les ontologies à réutiliser tout en ayant un œil critique. Une ontologie publiée ne veut nullement signifier qu'elle soit de qualité. La

méthodologie propose différentes approches pour l'évaluation et introduit l'utilisation des revues systématiques aux côtés des GBPC.

**La deuxième contribution** est la construction d'une ontologie qui couvre le diagnostic de la pneumonie appelée PNADO, en suivant notre méthodologie, pour ainsi fournir une preuve de concept et valider la méthodologie. Le choix du diagnostic de la pneumonie comme domaine d'application n'est pas fortuit. Les médecins impliqués dans ce projet affirment que les erreurs de diagnostic de cette maladie sont fréquentes. Celles-ci sont dues entre autres aux ambiguïtés au niveau des critères diagnostiques pour l'insuffisance cardiaque et au niveau de l'interprétation de la radiographie thoracique. La mauvaise gestion des connaissances en lien avec cette maladie est aussi problématique. Les médecins affirment que l'organisation de ces connaissances pourrait être bénéfique. Ces derniers n'ont été sollicités que lors de la définition du domaine et de la portée de l'ontologie et lors de son évaluation. Les résultats de l'évaluation ont montré que notre méthodologie est efficace. PNADO est la première ontologie de diagnostic de la pneumonie déclarée, développée pour représenter les différents aspects du diagnostic dans un format logique et formel. PNADO constitue un résultat en soi. Elle pourra être réutilisée et même étendue avec le traitement.

**La troisième contribution** est l'évaluation de plusieurs ontologies médicales même celles qui n'ont pas été réutilisées. En effet, nous avons constaté que plusieurs ontologies consultées dans OBO Foundry ou BioPortal présentent des lacunes. Nous avons détecté des anomalies dans plusieurs ontologies dans ces deux répertoires, même celles qui ne sont pas réutilisées. Cela est dû, entre autres, à l'absence de méthodologie fournissant toutes les étapes détaillées de construction.

**La quatrième contribution** est de proposer quelques améliorations aux ontologies existantes à savoir OGMS, SYMP, IDO et CPRO.

Bien que notre méthodologie ait été destinée à la construction des ontologies médicales, elle peut très bien être appliquée à la construction d'ontologies couvrant les domaines d'autres disciplines. Il faudrait remplacer les GBPC par une autre source de connaissances de référence (meilleure quand elle est codifiée) si une telle ressource existe. Par exemple, des cadres de référence pourraient être utilisés comme source de connaissances codifiées pour représenter un domaine de pronostic.

Les prochaines étapes envisagées sont :

- 1- L'intégration PNADO dans le système d'information SEKMED pour une évaluation et une utilisation dans la pratique clinique ;
- 2- L'extraction des règles décisionnelles à partir des GBPC, pour un système d'aide à la décision clinique. En effet, d'après l'OMS, les erreurs médicales au niveau du processus de diagnostic sont un problème majeur dont souffrent les soins de santé et affectent un nombre de plus en plus grandissant de patients. Leurs causes sont diverses et elles sont reliées entre autres aux difficultés de la gestion des connaissances médicales qui ne cessent d'augmenter et à l'absence d'un outil décisionnel efficace pour soutenir les médecins dans le processus de prise de décision pour réduire ou éviter les erreurs qui peuvent survenir.

# ANNEXES

## ANNEXE A : Matériel pour l'évaluation de PNADO par les médecins

### Partie I : Généralités sur PNADO

Évaluation de PNADO par les médecins Sylvain Croteau (urgentologue à l'hôpital de Gatineau), Serge Chartran (médecin de famille à la clinique de l'UQO), et Yasmine Lisa Rebaine (pneumologue à l'hôpital Charles-Le Moyne à Montréal)

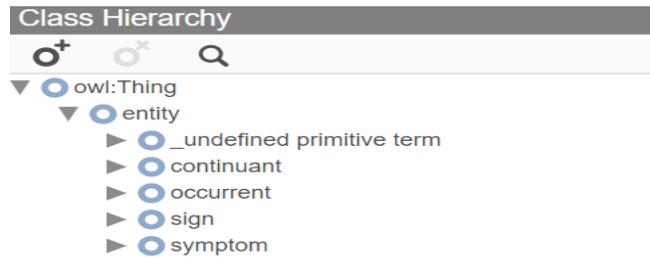
L'ontologie qui fera l'objet de l'évaluation est PNADO: Pneumonia Diagnosis Ontology. L'ontologie couvre la plupart des concepts reliés au diagnostic de la pneumonie : symptômes, signes cliniques, tests de laboratoire et leurs résultats, imagerie, pathogènes, types de pneumonie, diagnostics différentiels.

PNADO est initialement construite à partir de guides de bonnes pratiques cliniques, ensuite, elle est enrichie en réutilisant les ontologies médicales reconnues et validées par la communauté scientifique à savoir SNOMED-CT, HP (Human phenotype Ontology), DDO (Disease Ontology), SYMP(Symptom), etc.

Pour des fins de réutilisabilité et d'interopérabilité avec d'autres ontologies médicales, l'ontologie générique OGMS (Ontology for General Medical Science) a été utilisée. OGMS comprend des termes très généraux qui sont utilisés dans toutes les disciplines médicales, par exemple : "disease", "disorder", "disease course", "diagnosis", "patient", and "healthcare provider". L'OGMS utilise BFO (Basic Formal Ontology) comme ontologie de niveau supérieur. La portée de l'OGMS est limitée aux humains, mais de nombreux termes peuvent être appliqués à une variété d'organismes.

L'ontologie est mise en ligne sur WebProtégé. Veuillez cliquer sur ce lien : <https://webprotege.stanford.edu/>. Ensuite, créez un compte avec un nom utilisateur que vous m'enverrez pour que je puisse partager l'ontologie avec vous. Une fois vous aurez l'accès, veuillez suivre ces quelques indications pour mieux naviguer dans l'ontologie.

A l'ouverture de l'ontologie, vous verrez cet ensemble de concepts :



Veillez cliquer sur chaque concept pour visualiser son arborescence.

### **Quelques définitions :**

***Continuant*** : contient des concepts qui existent à un moment donné, comme une personne, un pathogène. Les continus conservent leur identité à travers le temps.

***Occurent*** : contient des événements qui ont des parties temporelles telles que le processus de diagnostic.

***Symptom*** : contient les symptômes de la maladie de la pneumonie. Ces symptômes sont extraits des guides de bonnes pratiques cliniques et enrichies en réutilisant les ontologies SYMP, HP et SNOMED-CT.

***Sign*** : correspond aux signes cliniques

Dans une ontologie, les termes sont regroupés sous forme de classes (class) qui définissent un groupe d'individus ou d'instances (individual) possédant des propriétés similaires. Exemple : symptom, sign, diagnosis, pathogen etc.

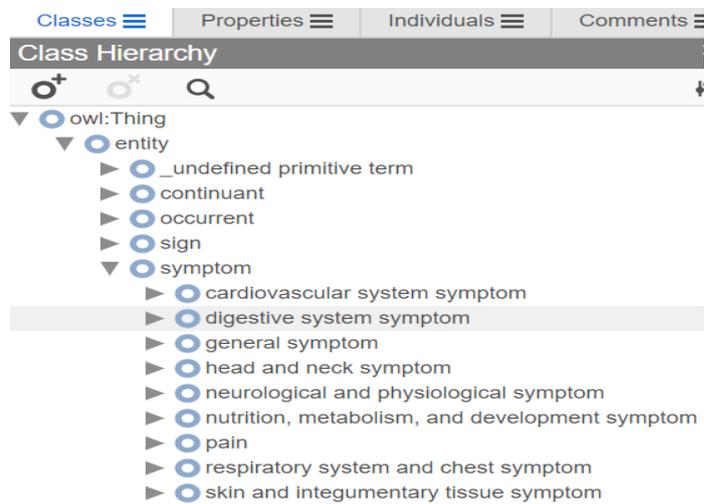
Ces concepts sont reliés entre eux avec des relations (object properties). Exemple : la relation « has diagnosis » relie le concept « patient » et le concept « diagnosis ».

**Important** : Ce que nous souhaitons évaluer dans PNADO sont les classes (class) et les relations (object properties).

### **Accès aux classes (concepts) :**

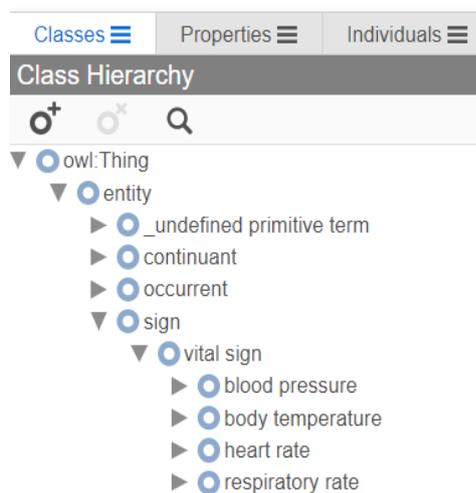
Pour accéder aux concepts qui concernent le diagnostic de la pneumonie et qui nécessitent une évaluation, vous n'aurez qu'à suivre les consignes suivantes :

**1. En cliquant sur **symptom** par exemple, vous aurez les symptômes de la pneumonie :**



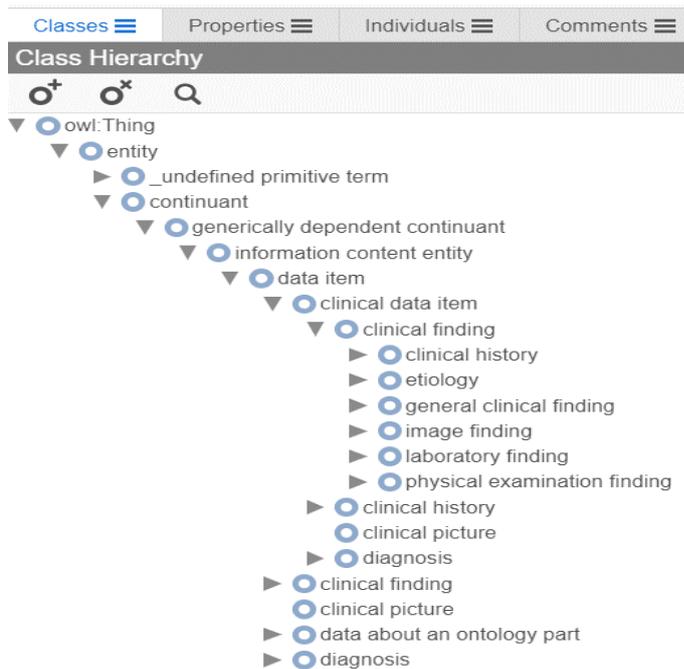
Les symptômes sont classés dans des catégories. Par exemple : les symptômes de la pneumonie qui affectent le système digestif sont mis dans la catégorie « digestive system symptom »

**2. Si vous cliquez sur **sign**, vous obtiendrez les signes cliniques de la maladie de la pneumonie :**

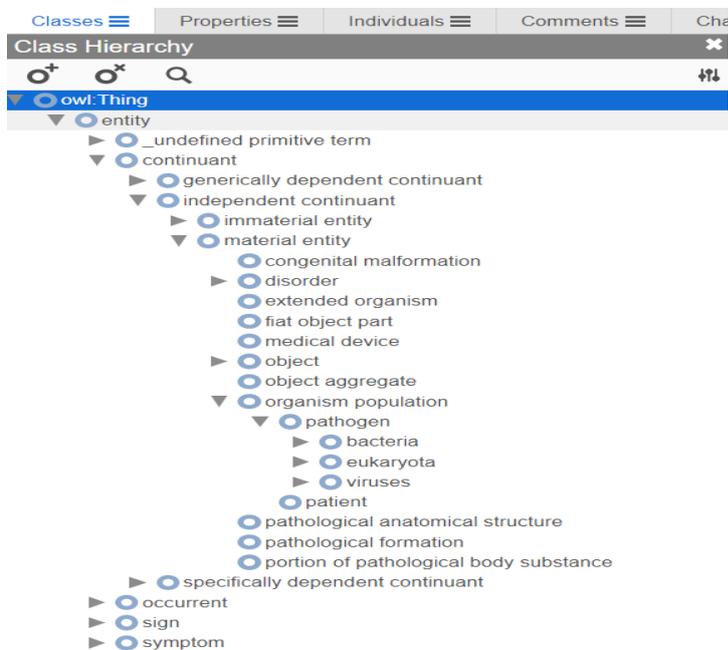


**3. Pour accéder au processus de diagnostic (**diagnostic process**), veuillez suivre ce chemin : **occurrent**→**process**→**planned process**→**health care process**→**diagnostic process****

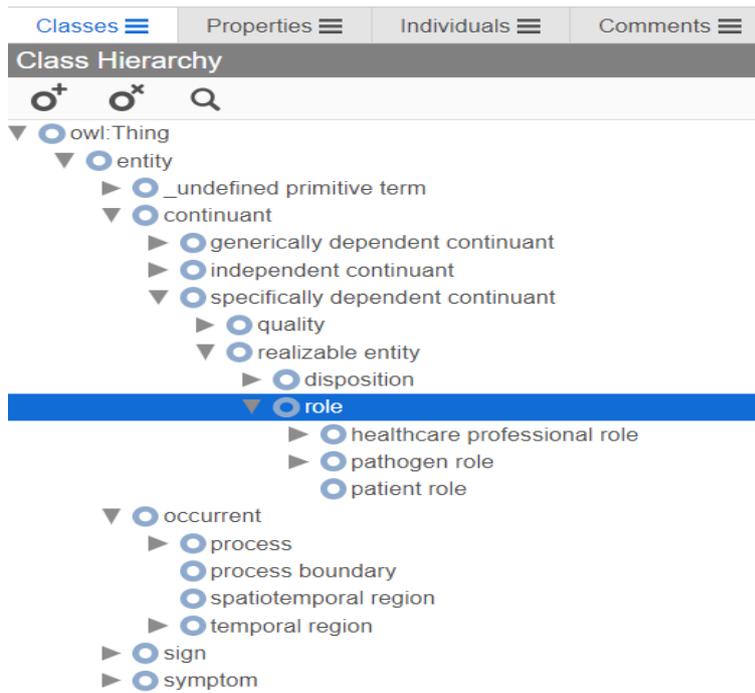




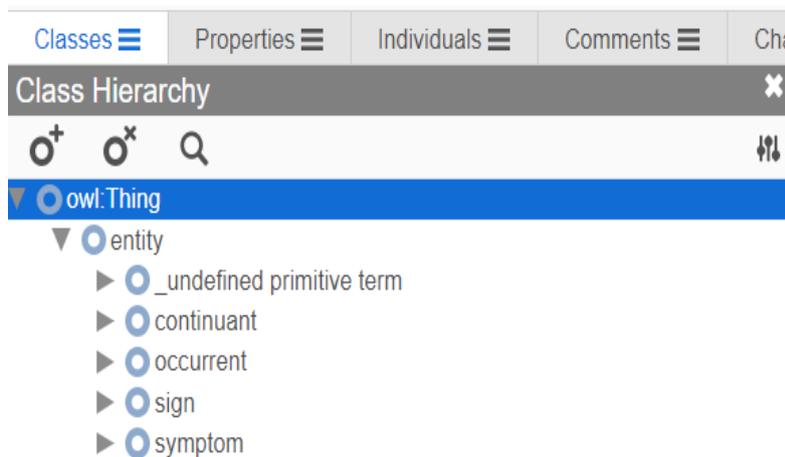
6. Pour visualiser les pathogènes (**pathogen**), veuillez suivre ce chemin :  
 continuant→independent continuant      continuant→material      entity→organism  
 population→pathogen



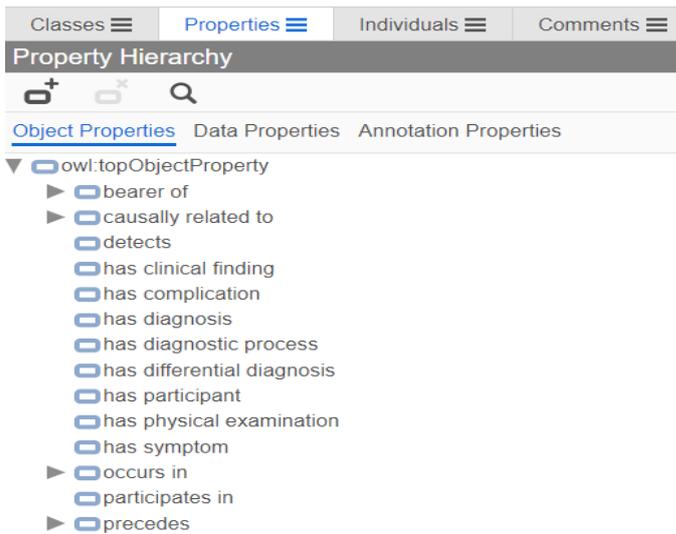
7. Pour accéder aux rôles (**role**), veuillez suivre ce chemin : continuant→ specifically dependent continuant→ realizable →entity role



**8. Pour mieux naviguer dans l'ontologie, vous pouvez faire une recherche en cliquant sur la loupe qui se trouve tout en haut et à gauche de l'interface.**



**9. Accès aux relations (object properties) :**



## Partie II : Évaluation en utilisant la technique de l'échelle (laddering technique)

**Les nouvelles relations à évaluer sont :**

Detects, has clinical finding, complication of, has diagnosis, differential diagnosis of, is diagnosis of, finding of.

### **Questions :**

1. Est-ce qu'il y a des symptômes ou des signes cliniques qui manquent pour diagnostiquer la pneumonie ?
2. Est-ce que la liste des résultats des tests de laboratoire (laboratory finding) est complète ?
3. Est-ce que les tests d'étiologie (etiologic tests) sont bons ?
4. Est-ce que « etiologic tests » doit être à l'intérieur de « laboratory test » ?



5. Est-ce que les définitions données aux relations (objects properties) : detects, has clinical finding, has complication, has diagnosis, has differential diagnosis, has symptom sont bonnes ?

### **Partie III : Évaluation en utilisant les questions de compétence**

To evaluate PNADO, we use competency questions that consist on a set of questions stated by physicians and replied in natural language. PNADO must be able to answer to these questions.

#### **Examples of competency questions:**

Cette partie est rédigée en anglais, car les concepts de PNADO sont dans la langue anglaise.

##### **1) Symptoms:**

- What are the pneumonia symptoms?
- Is the classification of symptoms according to the different systems (digestive system, cardiovascular system, etc.) correct?

- Is palpitation symptom relevant to diagnose pneumonia?
- Is palpitation symptom related to cardiovascular system symptoms?
- Are there any other pneumonia symptoms related to cardiovascular system besides palpitations?
- Is the definition of palpitation symptom correct?
- Does pneumonia have pathognomonic symptoms?
- Is there a relation between phenotype and symptom?

## **2) Signs:**

- What are the pneumonia signs?
- Are the following pneumonia signs in PNADO relevant to pneumonia diagnosis?
- Are there any missing signs to diagnose pneumonia?
- Are the signs definitions correct?

## **3) Diagnosis:**

- The diagnostic process includes etiologic test, medical imaging and procedure.
- Are there any missing concepts related to this process?
- Should we place the etiologic test in the diagnostic process or in the etiological process?

## ANNEXE B : Documentation de PNADO

1- Vue d'ensemble de PNADO (page d'accueil). Elle résume le contenu de l'ontologie en incluant les ontologies importées, OGMS dans notre cas. Il suffit de cliquer sur les composantes affichées pour avoir plus de détails.

The screenshot displays the web interface for the PNADO ontology. It features a top navigation bar with links for 'Ontologies', 'Classes', 'Object Properties', 'Data Properties', 'Annotation Properties', 'Individuals', 'Datatypes', and 'Clouds'. Below this, there are three main sections:

- Contents:** A list of ontology components: All Ontologies, Classes (1473), Object Properties (107), Data Properties (11), Annotation Properties (84), Individuals (18), and Datatypes (7).
- Entities (1696):** A list of specific entities, including 'owl:Thing', 'undefined primitive term', and various medical conditions like 'abnormal breath sounds', 'abnormal homeostasis', 'abnormal pulmonary examination', 'abnormal sputum', 'abnormal tracheal aspiration', 'abscess of lung and mediastinum', 'abscess of lung with pneumonia', 'accidental hypothermia in elderly person', 'achromobacter pneumonia', 'acquired genetic disease', 'acquired genetic disorder', 'actinomycotic meningitis', 'acute bronchitis', 'acute chest pain', 'acute confusion', 'acute confusional state, of cerebrovascular origin', 'acute confusional state, of endocrine origin', 'acute confusional state, of infective origin', 'acute confusional state, of metabolic origin', 'acute confusional state, post-traumatic', 'acute coryza', 'acute diarrhea', 'acute disease course', and 'acute dyspnea'.
- Metadata and Annotations:** This section provides details about the ontology file, including its URL (<http://purl.obolibrary.org/obo/pnado.owl>), the file path (`/C:/Users/Sabrina/Desktop/Ontologies/PNADO24.owl`), and a list of 9 annotations. The annotations include metadata such as collaborators (Sylvain Croteau, Veronique Nabels, Yasmine Lisa Rebaïne), creators (Michal Iglewski, Sabrina Azzi), a date (2019-09-10), a Creative Commons license (CC BY 4.0), and a comment describing the ontology's purpose: "Pneumonia Diagnosis Ontology (PNADO) provides the necessary knowledge to diagnose infective pneumonia. It covers symptoms and clinical signs, clinical history, pathogens, laboratory tests and results, imaging, differential diagnosis and pneumonia complications."

Additional sections include 'References' (Classes (1405), Object Properties (31), Data Properties (11), Annotation Properties (38), Datatypes (6)) and 'Imports (3)' (protege.stanford.edu/plugins/owl/dc/protege-dc.owl, purl.obolibrary.org/obo/ogms.owl, purl.obolibrary.org/obo/iao/dev/ontology-metadata.owl). A small note at the bottom right indicates 'OWL HTML inside'.

2- Une partie des classes et des relations organisées par ordre alphabétique

## All Classes (1473)

- [owl:Thing](#)
- ['\\_undefined primitive term'](#)
- ['abnormal breath sounds'](#)
- ['abnormal homeostasis'](#)
- ['Abnormal pulmonary examination'](#)
- ['abnormal sputum'](#)
- ['abnormal tracheal aspiration'](#)
- ['abscess of lung and mediastinum'](#)
- ['abscess of lung with pneumonia'](#)
- ['accidental hypothermia in elderly person'](#)
- ['achromobacter pneumonia'](#)
- ['acquired genetic disease'](#)
- ['acquired genetic disorder'](#)
- ['actinomycotic meningitis'](#)
- ['acute bronchitis'](#)
- ['acute chest pain'](#)
- ['acute confusion'](#)
- ['acute confusional state, of cerebrovascular origin'](#)
- ['acute confusional state, of endocrine origin'](#)
- ['acute confusional state, of infective origin'](#)
- ['acute confusional state, of metabolic origin'](#)
- ['acute confusional state, post-traumatic'](#)
- ['acute coryza'](#)
- ['acute diarrhea'](#)
- ['acute disease course'](#)
- ['acute dyspnea'](#)
- ['acute headache'](#)
- ['acute idiopathic pulmonary fibrosis'](#)
- ['acute interstitial pneumonia'](#)
- ['acute muscle stiffness of neck'](#)
- ['acute non-psychotic brain syndrome'](#)
- ['acute onset'](#)
- ['acute osteomyelitis'](#)
- ['acute pulmonary African histoplasmosis'](#)
- ['acute purulent meningitis'](#)
- ['acute respiratory distress syndrome'](#)
- ['acute respiratory failure'](#)
- ['acute rhinosinusitis'](#)
- ['acute secondary syphilitic meningitis'](#)
- ['acute sinusitis'](#)
- ['acute syphilitic meningitis'](#)
- ['acute vomiting'](#)
- ['adenoviral pneumonia'](#)
- ['adult failure to thrive'](#)
- ['age at diagnosis'](#)
- ['age at time of procedure'](#)
- ['air alveologram'](#)
- ['air bronchograms'](#)
- ['air space nodules'](#)
- ['alcohol withdrawal delirium'](#)
- ['allergic cough'](#)
- ['allergic rhinitis caused by mold'](#)
- ['allergic rhinitis'](#)

## All Objectproperties (105)

- ['bearer of at all times'](#)
- ['bearer of at some time'](#)
- ['bearer of'](#)
- ['causally related to'](#)
- ['causes condition'](#)
- ['causes or contributes to condition'](#)
- ['complication of disease'](#)
- ['complication of function'](#)
- ['complication of procedure'](#)
- ['complication of treatment'](#)
- ['complication of'](#)
- ['concretized by, at all times'](#)
- ['concretized by, at some time'](#)
- ['concretizes at all times'](#)
- ['concretizes at some time'](#)
- ['contains process'](#)
- ['contributes to condition'](#)
- ['differential diagnosis of'](#)
- ['disposition of at all times'](#)
- ['during which exists'](#)
- ['exists at'](#)
- ['finding of'](#)
- ['function of at all times'](#)
- ['generically depends on at some time'](#)
- ['has clinical finding'](#)
- ['has continuant part at all times that part exists'](#)
- ['has continuant part at all times'](#)
- ['has continuant part at some time'](#)
- ['has diagnosis'](#)
- ['has diagnostic process'](#)
- ['has disposition at all times'](#)
- ['has disposition at some time'](#)
- ['has disposition'](#)
- ['has function at all times'](#)
- ['has function at some time'](#)
- ['has generic dependent at some time'](#)
- ['has history'](#)
- ['has location at all times'](#)
- ['has location at some time'](#)
- ['has material basis at all times'](#)
- ['has member part at all times'](#)
- ['has member part at some time'](#)
- ['has occurrent part'](#)
- ['has participant at all times'](#)
- ['has participant at some time'](#)
- ['has participant'](#)
- ['has profile'](#)
- ['has proper continuant part at all times'](#)
- ['has proper continuant part at some time'](#)
- ['has proper occurrent part'](#)
- ['has proper temporal part'](#)
- ['has quality at all times'](#)
- ['has quality at some time'](#)

## 3- La classe *entity*

Ontologies Classes Object Properties Data Properties Annotation Properties Individuals Datatypes Clouds

Class: *entity*

Annotations (15)

- BFO OWL specification label "entity"
- BFO CLIF specification label "Entity"
- example of usage "Julius Caesar" @en
- example of usage "Verdi's Requiem" @en
- example of usage "the Second World War" @en
- example of usage "your body mass index" @en
- editor note "BFO 2 Reference: In all areas of empirical inquiry we encounter general terms of two sorts. First are general terms which refer to universals or types: animal tuberculosis surgical procedure disease. Second, are general terms used to refer to groups of entities which instantiate a given universal but do not correspond to the extension of any subuniversal of that universal because there is nothing intrinsic to the entities in question by virtue of which they – and only they – are counted as belonging to the given group. Examples are: animal purchased by the Emperor tuberculosis diagnosed on a Wednesday surgical procedure performed on a patient from Stockholm person identified as candidate for clinical trial #2056-555 person who is signatory of Form 656-PPV painting by Leonardo da Vinci. Such terms, which represent what are called 'specializations' in [81] @en
- editor note "Entity doesn't have a closure axiom because the subclasses don't necessarily exhaust all possibilities. For example Werner Ceusters' 'portions of reality' include 4 sorts, entities (as BFO construes them), universals, configurations, and relations. It is an open question as to whether entities as construed in BFO will at some point also include these other portions of reality. See, for example, 'How to track absolutely everything' at [http://www.referent-tracking.com/\\_RTU/papers/CeustersICbookRevised.pdf](http://www.referent-tracking.com/_RTU/papers/CeustersICbookRevised.pdf)" @en
- elucidation "An entity is anything that exists or has existed or will exist. (axiom label in BFO2 Reference: [001-001])" @en
- has axiom label <http://purl.obolibrary.org/obo/bfo/axiom/0000004> Attempt to load owl/rdf
- has axiom label <http://purl.obolibrary.org/obo/bfo/axiom/001-001> Attempt to load owl/rdf
- rdfs:comment "per discussion with Barry Smith"
- rdfs:isDefinedBy <http://purl.obolibrary.org/obo/bfo.owl>
- rdfs:label "entity" @en
- rdfs:seeAlso [http://www.referent-tracking.com/\\_RTU/papers/CeustersICbookRevised.pdf](http://www.referent-tracking.com/_RTU/papers/CeustersICbookRevised.pdf) Attempt to load owl/rdf

Superclasses (1)

- exists at some 'temporal region'

Usage (2)

- exists at **Domain entity**
- has axiom label <http://purl.obolibrary.org/obo/bfo/axiom/101> Attempt to load owl/rdf
- during which exists **Range entity**
- has axiom label <http://purl.obolibrary.org/obo/bfo/axiom/101> Attempt to load owl/rdf

OWL HTML Inside

## 4- La classe *infective pneumonia*

Ontologies Classes Object Properties Data Properties Annotation Properties Individuals Datatypes Clows

Class: 'infective pneumonia'

Annotations (3)

- id "PNAO:0001172"
- database\_cross-reference "UMLS\_CUI:C0729704"
- rdfs:label "infective pneumonia"

Superclasses (16)

- 'infectious disease'
- 'complication of some arrhythmia'
- 'complication of some endocarditis'
- 'complication of some steomyelitis'
- 'complication of some shock'
- 'complication of some 'systemic inflammatory response syndrome''
- 'complication of some pericarditis'
- 'complication of some 'septic arthritis''
- 'complication of some sepsis'
- 'differential diagnosis of some 'pulmonary embolism''
- 'differential diagnosis of some 'acute bronchitis''
- 'differential diagnosis of some 'bronchitis''
- 'differential diagnosis of some 'interstitial lung disease''
- 'differential diagnosis of some 'lung cancer''
- 'differential diagnosis of some 'non infectious pneumonia''
- 'differential diagnosis of some atelectasis'

OWL HTML inside

## 5- La classe *non infectious pneumonia*

Ontologies Classes Object Properties Data Properties Annotation Properties Individuals Datatypes Clows

Class: 'non infectious pneumonia'

Annotations (2)

- id "PNAO:0001163"
- rdfs:label "non infectious pneumonia"

Superclasses (16)

- 'non infectious disease'
- 'complication of some arrhythmia'
- 'complication of some endocarditis'
- 'complication of some steomyelitis'
- 'complication of some shock'
- 'complication of some 'systemic inflammatory response syndrome''
- 'complication of some pericarditis'
- 'complication of some 'septic arthritis''
- 'complication of some sepsis'
- 'differential diagnosis of some 'pulmonary embolism''
- 'differential diagnosis of some 'acute bronchitis''
- 'differential diagnosis of some 'bronchitis''
- 'differential diagnosis of some 'interstitial lung disease''
- 'differential diagnosis of some 'lung cancer''
- 'differential diagnosis of some 'non infectious pneumonia''
- 'differential diagnosis of some atelectasis'

Usage (1)

- 'infective pneumonia' ⊑ 'differential diagnosis of some 'non infectious pneumonia''

OWL HTML inside

## 6- Relation *complication of*

Ontologies Classes Object Properties Data Properties Annotation Properties Individuals Datatypes Clows

Object Property: 'complication of'

Annotations (3)

- definition "a relation that holds between a disease "A" and another disease "B" that develops in the course of the disease "A.""
- id "PNAO:0001289"
- rdfs:label "complication of"

Ranges (1)

- disease

Usage (20)

- 'non infectious pneumonia' ⊑ 'complication of some arrhythmia'
- 'non infectious pneumonia' ⊑ 'complication of some endocarditis'
- 'non infectious pneumonia' ⊑ 'complication of some steomyelitis'
- 'non infectious pneumonia' ⊑ 'complication of some shock'
- 'non infectious pneumonia' ⊑ 'complication of some 'systemic inflammatory response syndrome''
- 'non infectious pneumonia' ⊑ 'complication of some pericarditis'
- 'non infectious pneumonia' ⊑ 'complication of some 'septic arthritis''
- 'non infectious pneumonia' ⊑ 'complication of some sepsis'
- 'infective pneumonia' ⊑ 'complication of some arrhythmia'
- 'infective pneumonia' ⊑ 'complication of some endocarditis'
- 'infective pneumonia' ⊑ 'complication of some steomyelitis'
- 'infective pneumonia' ⊑ 'complication of some shock'
- 'infective pneumonia' ⊑ 'complication of some 'systemic inflammatory response syndrome''
- 'infective pneumonia' ⊑ 'complication of some pericarditis'
- 'infective pneumonia' ⊑ 'complication of some 'septic arthritis''
- 'infective pneumonia' ⊑ 'complication of some sepsis'
- 'complication of function' ⊑ 'complication of'
- 'complication of disease' ⊑ 'complication of'
- 'complication of procedure' ⊑ 'complication of'
- 'complication of treatment' ⊑ 'complication of'

OWL HTML inside

# Bibliographie

1. Corcho, O., A. Gómez-Pérez, and M. Fernández-López, *Ontological engineering*. With examples from the areas of Knowledge Management, e-Commerce and the Semantic Web (Advanced Information and Knowledge Processing), 2004: p. 5.
2. Monteiro, E.J.M., et al. *A recommender system for medical imaging diagnostic*. in *MIE*. 2015.
3. Séroussi, B. and J. Bouaud, *Systèmes informatiques d'aide à la décision en médecine: panorama des approches utilisant les données et les connaissances*. *Pratique Neurologique-FMC*, 2014. 5(4): p. 303-316.
4. Bogner, M.S., *Human error in medicine*. 2018: CRC Press.
5. Barth, J.H., et al., *Why are clinical practice guidelines not followed?* *Clinical Chemistry and Laboratory Medicine (CCLM)*, 2016. 54(7): p. 1133-1139.
6. Birrenbach, T., et al., *Physicians' attitudes toward, use of, and perceived barriers to clinical guidelines: a survey among Swiss physicians*. *Advances in medical education and practice*, 2016. 7: p. 673.
7. Fischer, F., et al. *Barriers and strategies in guideline implementation—a scoping review*. in *Healthcare*. 2016. Multidisciplinary Digital Publishing Institute.
8. Jin, Y., et al., *Barriers and enablers for the implementation of clinical practice guidelines in China: a mixed-method study*. *BMJ open*, 2019. 9(9): p. e026328.
9. Carroll, C., *Qualitative evidence synthesis to improve implementation of clinical guidelines*. *Bmj*, 2017. 356: p. j80.
10. Nendaz, M., et al., *Le raisonnement clinique: données issues de la recherche et implications pour l'enseignement*. *Pédagogie médicale*, 2005. 6(4): p. 236.
11. (CMQ), C.d.m.d.Q., *La pratique médicale en soins de longue durée : guide d'exercice du Collège des médecins du Québec*. . 2007: Montréal, Qc.
12. Balogh, E.P., B.T. Miller, and J. Ball, *Improving Diagnosis in Health Care*. *Committee on Diagnostic Error in Health Care; Board on Health Care Services; Institute of Medicine; The National Academies of Sciences, Engineering, and Medicine*. 2015, Washington (DC): The National Academies Press (US). p. 33.
13. De Vries, T., et al., *Guide to good prescribing: a practical manual*. 1994: p. 11.
14. Zulkarnain, N.Z., F. Meziane, and G. Crofts. *A methodology for biomedical ontology reuse*. in *International conference on applications of natural language to information systems*. 2016. Springer.

15. Shah, T., et al. *A guiding framework for ontology reuse in the biomedical domain*. in *2014 47th Hawaii International Conference on System Sciences*. 2014. IEEE.
16. Ochs, C., et al., *An empirical analysis of ontology reuse in BioPortal*. *Journal of biomedical informatics*, 2017. 71: p. 165-177.
17. El-Sappagh, S., et al., *DMTO: a realistic ontology for standard diabetes mellitus treatment*. *Journal of biomedical semantics*, 2018. 9(1): p. 8.
18. He, Y., et al., *CIDO, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis*. *Scientific Data*, 2020. 7(1): p. 1-5.
19. Lin, F.P., et al., *Cancer Care Treatment Outcome Ontology: A Novel Computable Ontology for Profiling Treatment Outcomes in Patients With Solid Tumors*. *JCO Clinical Cancer Informatics*, 2018. 2: p. 1-14.
20. Barton, A., et al. *The Cardiovascular Disease Ontology*. in *FOIS*. 2014.
21. Farinelli, F., et al., *OntONeo: The Obstetric and Neonatal Ontology*. *Dealing with elements of medical encounters: An approach based on ontological realism.*, 2016. 1747
22. Utecht, J., et al. *OOSTT: A resource for analyzing the organizational structures of trauma centers and trauma systems*. in *CEUR workshop proceedings*. 2016. NIH Public Access.
23. Younesi, E., et al., *PDON: Parkinson's disease ontology for representation and modeling of the Parkinson's disease knowledge domain*. *Theoretical Biology and Medical Modelling*, 2015. 12(1): p. 20.
24. Farquhar, A., R. Fikes, and J. Rice, *The ontolingua server: A tool for collaborative ontology construction*. *International journal of human-computer studies*, 1997. 46(6): p. 707-727.
25. Organisation, W.H., *Diagnostic Errors*. 2016.
26. Brendish, N.J., et al., *Hospitalised adults with pneumonia are frequently misclassified as another diagnosis*. *Respiratory medicine*, 2019. 150: p. 81-84.
27. Centor, R.M. and R. Kumar, *Diagnostic error in community-acquired pneumonia*. *Diagnosis*, 2014. 1(2): p. 145-145.
28. Schiff, G.D., et al., *Diagnostic error in medicine: analysis of 583 physician-reported errors*. *Archives of internal medicine*, 2009. 169(20): p. 1881-1887.
29. Lambert, B.L., et al., *Diagnosing diagnostic errors: Lessons from a multi-institutional collaborative*, in *Advances in Patient Safety: From Research to Implementation*. 2005, Department of Health and Human Services, Agency fo.
30. Schiff, G.D., *Minimizing diagnostic error: the importance of follow-up and feedback*. *The American journal of medicine*, 2008. 121(5): p. S38-S42.

31. Kubilay, Z., et al., *When is pneumonia not pneumonia: a clinicopathologic study of the utility of lung tissue biopsies in determining the suitability of cadaveric tissue for donation*. Cell and tissue banking, 2016. 17(2): p. 205-210.
32. Robert McNutt, M.R.A., MD; Scott Hasler, MD. *Diagnosing Diagnostic Mistakes*. 2005 [cited 2017 november 8]; Available from: <https://psnet.ahrq.gov/webmm/case/95>.
33. Rao, A., et al. *Tabla: An acoustic device designed for low cost pneumonia detection*. in *Healthcare Innovations and Point of Care Technologies (HI-POCT), 2017 IEEE*. 2017. IEEE.
34. Balogh, E.P., B.T. Miller, and J. Ball, *Improving diagnosis in health care*. 2016, Washington: National Academies Press.
35. Dissanayake, P.I., T.K. Colicchio, and J.J. Cimino, *Using clinical reasoning ontologies to make smarter clinical decision support systems: a systematic review and data synthesis*. Journal of the American Medical Informatics Association, 2020. 27(1): p. 159-174.
36. Sherimon, P. and R. Krishnan, *OntoDiabetic: an ontology-based clinical decision support system for diabetic patients*. Arabian Journal for Science and Engineering, 2016. 41(3): p. 1145-1160.
37. Shen, Y., et al., *An ontology-driven clinical decision support system (IDDAP) for infectious disease diagnosis and antibiotic prescription*. Artificial intelligence in medicine, 2018. 86: p. 20-32.
38. Manzoor, U., et al., *Ontology-Based Clinical Decision Support System for Predicting High-Risk Pregnant Woman*. system, 2015. 6(12).
39. Dhondt, U., *Science suprême et ontologie chez Aristote*. Revue philosophique de Louvain, 1961. 59: p. 5-30.
40. Gruber, T.R., *A translation approach to portable ontology specifications*. Knowledge acquisition, 1993. 5(2): p. 199-220.
41. Guarino, N. and C. Welty. *A formal ontology of properties*. in *International Conference on Knowledge Engineering and Knowledge Management*. 2000. Springer.
42. Borst, W.N., *Construction of engineering ontologies for knowledge sharing and reuse*. 1997, Institute for Telematica and Information Technology, University of Twente.
43. Studer, R., V.R. Benjamins, and D. Fensel, *Knowledge engineering: principles and methods*. Data & knowledge engineering, 1998. 25(1-2): p. 161-197.
44. Kalfoglou, Y. and M. Schorlemmer, *Ontology mapping: the state of the art*. The knowledge engineering review, 2003. 18(1): p. 1-31.
45. Predoiu, L., et al., *D4. 2.2 State-of-the-art survey on Ontology Merging and Aligning V2*. EU-IST Integrated Project IST-2003-506826 SEKT, 2005: p. 79.

46. Gruber, T.R., *Toward principles for the design of ontologies used for knowledge sharing?* International journal of human-computer studies, 1995. 43(5-6): p. 907-928.
47. Genesereth, M.R. and N.J. Nilsson, *Logical foundations of artificial intelligence*. 1987: Morgan Kaufmann.
48. Gómez-Pérez, A., *Développement récents en matière de conception, de maintenance et d'utilisation des ontologies*. Terminologies nouvelles, 1999. 19: p. 9-20.
49. Gómez-Pérez, A. and R. Benjamins. *Overview of knowledge sharing and reuse components: Ontologies and problem-solving methods*. 1999. IJCAI and the Scandinavian AI Societies. CEUR Workshop Proceedings.
50. Minsky, M., *A framework for representing knowledge*. 1974.
51. Quillian, M.R., *Word concepts: A theory and simulation of some basic semantic capabilities*. Behavioral science, 1967. 12(5): p. 410-430.
52. Baader, F., et al., *The description logic handbook: Theory, implementation and applications*. 2003: Cambridge university press.
53. Horrocks, I. *The fact system*. in *International Conference on Automated Reasoning with Analytic Tableaux and Related Methods*. 1998. Springer.
54. Sirin, E., et al., *Pellet: A practical owl-dl reasoner*. Journal of Web Semantics, 2007. 5(2): p. 51-53.
55. Haarslev, V., et al., *The RacerPro knowledge representation and reasoning system*. Semantic Web, 2012. 3(3): p. 267-277.
56. Glimm, B., et al., *HermiT: an OWL 2 reasoner*. Journal of Automated Reasoning, 2014. 53(3): p. 245-269.
57. Steigmiller, A., T. Liebig, and B. Glimm, *Konclude: system description*. Journal of Web Semantics, 2014. 27: p. 78-85.
58. Bagosi, T., et al. *The ontop framework for ontology based data access*. in *Chinese Semantic Web and Web Science Conference*. 2014. Springer.
59. Parsia, B., et al., *The OWL reasoner evaluation (ORE) 2015 competition report*. Journal of Automated Reasoning, 2017. 59(4): p. 455-482.
60. Aubert, C., *Diagnostics de consultation en médecine générale établis à partir de la CISP2 sur un échantillon de médecins généralistes en Haute-Normandie*. 2012, Faculté mixte de médecine et de pharmacie de Rouen
61. *CIM-10-CA (Classification statistique internationale des maladies et des problèmes de santé connexes, 10e version, Canada), 2017*. June 29, 2018]; Available from: <https://www.cihi.ca/fr/normes-et-soumission-de-donnees/codification-et-classification/cim-10-ca>.

62. *Overview of SNOMED CT, 2016.* June 29, 2018]; Available from: [https://www.nlm.nih.gov/healthit/snomedct/snomed\\_overview.html](https://www.nlm.nih.gov/healthit/snomedct/snomed_overview.html).
63. Daniel Le Bozec, C., *Gestion des connaissances multi-expertes en imagerie médicale idem: images et diagnostic par l'exemple en médecine.* 2001, Université Paris 6.
64. Dolin, R.H., et al. *The SNOMED RT Procedure Model.* in *Proceedings of the AMIA Symposium.* 2001. American Medical Informatics Association.
65. *SNOMED CT Guide d'initiation, 2014.* International Health Terminology Standards Development Organisation.
66. *SNOMED International SNOMED CT Browser, 2017.* June 29, 2018]; Available from: <http://browser.ihtsdotools.org/?perspective=full&conceptId1=404684003&edition=ca-edition&release=v20161031&server=https://prod-browser%20exten.ihtsdotools.org/api/snomed&langRefset=19491000087109,20581000087109>.
67. Haug, P.J., et al., *An ontology-driven, diagnostic modeling system.* Journal of the American Medical Informatics Association, 2013. 20(e1): p. e102-e110.
68. Grüninger, M. and M.S. Fox, *The role of competency questions in enterprise engineering,* in *Benchmarking—Theory and practice.* 1995, Springer. p. 22-31.
69. Gómez-Pérez, A., N. Juristo, and J. Pazos, *Evaluation and assessment of knowledge sharing technology.* Towards very large knowledge bases, 1995: p. 289-296.
70. Blázquez, J., et al. *Building ontologies at the knowledge level using the ontology design environment.* in *Banff Knowledge Acquisition for Knowledge-Based Systems Workshop.* 1998.
71. Fernández-López, M., A. Gómez-Pérez, and N. Juristo, *Methontology: from ontological art towards ontological engineering.* 1997.
72. Dieng, R., et al., *Méthodes et outils pour la gestion des connaissances: une approche pluridisciplinaire du knowledge management.* 2001: Dunod.
73. Gandon, F., *Ontology Engineering: a survey and a return on experience.* 2002, INRIA.
74. Fernández-López, M. *Overview of methodologies for building ontologies.* in *IJCAI99 workshop on ontologies and problem-solving methods: Lessons learned and future trends.* 1999.
75. IEEE, *1074-2006—IEEE standard for developing a software project life cycle process.* 2006, IEEE.
76. Committee, I.S., *Ieee std 610.12-1990 ieee standard glossary of software engineering terminology.* [online] [http://std-dards.ieee.org/reading/ieee/stdpublic/description/se/610.12-1990\\_desc.html](http://std-dards.ieee.org/reading/ieee/stdpublic/description/se/610.12-1990_desc.html), 1990.

77. Gruninger, M. and M.S. Fox. *The design and evaluation of ontologies for enterprise engineering*. in *Workshop on Implemented Ontologies, European Workshop on Artificial Intelligence, Amsterdam, The Netherlands*. 1994.
78. Uschold, M. and M. King, *Towards a methodology for building ontologies*. 1995: Artificial Intelligence Applications Institute, University of Edinburgh Edinburgh.
79. López, M.F., et al., *Building a chemical ontology using methontology and the ontology design environment*. IEEE Intelligent Systems and their applications, 1999. 14(1): p. 37-46.
80. Benjamins, R., B. Selic, and A. Gangemi, *Law and the semantic web: legal ontologies, methodologies, legal information retrieval, and applications*. Vol. 3369. 2005: Springer.
81. Noy, N.F. and D.L. McGuinness, *Ontology development 101: A guide to creating your first ontology*. 2001, Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, Stanford, CA.
82. Staab, S., et al., *Knowledge processes and ontologies*. IEEE Intelligent systems, 2001. 16(1): p. 26-34.
83. Bachimont, B., A. Isaac, and R. Troncy, *Semantic commitment for designing ontologies: a proposal*. Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web, 2002: p. 211-258.
84. Kassel, G., *Integration of the DOLCE top-level ontology into the OntoSpec methodology*, in *arXiv preprint cs/0510050*. 2005, Université de Picardie Jules Verne.
85. Buitelaar, P., D. Olejnik, and M. Sintek. *A protégé plug-in for ontology extraction from text based on linguistic analysis*. in *European Semantic Web Symposium*. 2004. Springer.
86. Cimiano, P. and J. Völker. *A framework for ontology learning and data-driven change discovery*. in *Proceedings of the 10th International Conference on Applications of Natural Language to Information Systems (NLDB), Lecture Notes in Computer Science, Springer*. 2005. Springer.
87. Aussenac-Gilles, N., S. Despres, and S. Szulman, *The TERMINAE Method and Platform for Ontology Engineering from Texts*. ONTOLOGY LEARNING AND POPULATION: BRIDGING THE GAP BETWEEN TEXT AND KNOWLEDGE, 2008: p. 199-223.
88. Biébow, B. and S. Szulman, *Terminae: une approche terminologique pour la construction d'ontologies du domaine à partir de textes*. Actes de RFIA2000, Reconnaissances des Formes et Intelligence Artificielle, 2000.
89. Cimiano, P. and J. Völker. *text2onto*. in *International conference on application of natural language to information systems*. 2005. Springer.
90. Salton, G., A. Wong, and C.-S. Yang, *A vector space model for automatic indexing*. Communications of the ACM, 1975. 18(11): p. 613-620.

91. Frantzi, K., S. Ananiadou, and J. Tsujii, *The c-value/nc-value method of automatic recognition for multi-word terms*. Research and advanced technology for digital libraries, 1998: p. 520-520.
92. Hearst, M.A. *Automatic acquisition of hyponyms from large text corpora*. in *Proceedings of the 14th conference on Computational linguistics-Volume 2*. 1992. Association for Computational Linguistics.
93. Gómez-Pérez, A. and M.C. Suárez-Figueroa, *NeOn methodology for building ontology networks: a scenario-based methodology*. 2009.
94. Navigli, R., P. Velardi, and A. Gangemi, *Ontology learning and its application to automated terminology translation*. IEEE Intelligent systems, 2003. 18(1): p. 22-31.
95. Fortuna, B., M. Grobelnik, and D. Mladeníć, *Semi-automatic data-driven ontology construction system*. 2006.
96. Hartigan, J.A. and M.A. Wong, *Algorithm AS 136: A k-means clustering algorithm*. Journal of the Royal Statistical Society. Series C (Applied Statistics), 1979. 28(1): p. 100-108.
97. Deerwester, S., et al., *Indexing by latent semantic analysis*. Journal of the American society for information science, 1990. 41(6): p. 391.
98. Goldfain, A., et al. *Vital sign ontology*. in *In Proceedings of the Workshop on Bio-Ontologies*. 2011. Vienna: ISMB.
99. Baclawski, K., et al. *Towards a symptom ontology for semantic web applications*. in *International Semantic Web Conference*. 2004. Springer.
100. Cowell, L.G. and B. Smith, *Infectious disease ontology*, in *Infectious disease informatics*. 2010, Springer. p. 373-395.
101. Sahoo, S.S., et al., *Epilepsy and seizure ontology: towards an epilepsy informatics infrastructure for clinical research and patient care*. Journal of the American Medical Informatics Association, 2014. 21(1): p. 82-89.
102. Scheuermann, R.H., W. Ceusters, and B. Smith, *Toward an ontological treatment of disease and diagnosis*. Summit on translational bioinformatics, 2009. 2009: p. 116.
103. Pathak, J., T.M. Johnson, and C.G. Chute, *Survey of modular ontology techniques and their applications in the biomedical domain*. Integrated computer-aided engineering, 2009. 16(3): p. 225-242.
104. Simperl, E., *Reusing ontologies on the Semantic Web: A feasibility study*. Data & Knowledge Engineering, 2009. 68(10): p. 905-925.
105. Katsumi, M. and M. Grüninger. *What is ontology reuse?* in *FOIS*. 2016.
106. Obrst, L., et al., *Semantic web and big data meets applied ontology*. 2014.

107. Bodenreider, O. and A. Burgun, *Biomedical ontologies*, in *Medical Informatics*. 2005, Springer. p. 211-236.
108. Guarino, N., *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy*. Vol. 46. 1998: IOS press.
109. Borgo, S. and P. Leitão. *The role of foundational ontologies in manufacturing domain applications*. in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. 2004. Springer.
110. Guizzardi, G. *The role of foundational ontologies for conceptual modeling and domain ontology representation*. in *2006 7th International Baltic conference on databases and information systems*. 2006. IEEE.
111. Stenzhorn, H., E. Beibwanger, and S. Schulz. *Towards a top-domain ontology for linking biomedical ontologies*. in *Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems*. 2007. IOS Press.
112. Keet, C.M. *The use of foundational ontologies in ontology development: an empirical assessment*. in *Extended Semantic Web Conference*. 2011. Springer.
113. Smith, B. and W. Ceusters, *Ontological realism: A methodology for coordinated evolution of scientific ontologies*. *Applied ontology*, 2010. 5(3-4): p. 139-188.
114. Elmhadhbi, L., M.-H. Karray, and B. Archimède, *Toward the Use of Upper-Level Ontologies for Semantically Interoperable Systems: An Emergency Management Use Case*, in *Enterprise Interoperability VIII*. 2019, Springer. p. 131-140.
115. Geraci, A., et al., *IEEE standard computer dictionary: Compilation of IEEE standard computer glossaries*. 1991: IEEE Press.
116. Kubicek, H., R. Cimander, and H.J. Scholl, *Organizational interoperability in e-government: lessons from 77 European good-practice cases*. 2011: Springer Science & Business Media.
117. Komatsoulis, G.A., et al., *caCORE version 3: Implementation of a model driven, service-oriented architecture for semantic interoperability*. *Journal of biomedical informatics*, 2008. 41(1): p. 106-123.
118. Arp, R., B. Smith, and A.D. Spear, *Building ontologies with basic formal ontology*. 2015: Mit Press.
119. Heery, R. and S. Anderson, *Digital repositories review*. 2005.
120. Ding, Y. and D. Fensel. *Ontology library systems: The key to successful ontology re-use*. in *Proceedings of the First International Conference on Semantic Web Working*. 2001. CEUR-WS. org.

121. Hartmann, J., R. Palma, and A. Gómez-Pérez, *Ontology repositories*, in *Handbook on Ontologies*. 2009, Springer. p. 551-571.
122. Consortium, G.O., *The gene ontology (GO) project in 2006*. Nucleic acids research, 2006. 34(suppl 1): p. D322-D326.
123. Smith, B., et al., *The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration*. Nature biotechnology, 2007. 25(11): p. 1251-1255.
124. Bodenreider, O., *The unified medical language system (UMLS): integrating biomedical terminology*. Nucleic acids research, 2004. 32(suppl\_1): p. D267-D270.
125. Gu, H., et al., *Quality assurance of umls semantic type assignments using snomed ct hierarchies*. Methods of information in medicine, 2016. 55(02): p. 158-165.
126. Hoehndorf, R., et al., *Aber-OWL: a framework for ontology-based data access in biology*. BMC bioinformatics, 2015. 16(1): p. 26.
127. Sackett, D.L., et al., *Evidence based medicine: what it is and what it isn't*. 1996, British Medical Journal Publishing Group.
128. Lohr, K.N. and M.J. Field, *Clinical practice guidelines: directions for a new program*. Vol. 90. 1990: National Academies Press.
129. Shekelle, P., et al., *When should clinical guidelines be updated?* BMJ: British Medical Journal, 2001. 323(7305): p. 155.
130. Frankel, H.L., et al., *Strategies to improve compliance with evidence-based clinical management guidelines*. Journal of the American College of Surgeons, 1999. 189(6): p. 533-538.
131. Gillois, P., et al., *From paper-based to electronic guidelines: application to French guidelines*. Medinfo, 2001. 10(Pt 1): p. 196-200.
132. Armstrong, D., et al., *Development of clinical guidelines in a health district: an attempt to find consensus*. Quality and Safety in Health Care, 1992. 1(4): p. 241-244.
133. *WHO Handbook for Guideline Development – 2nd Edition*. 2014.
134. Patel, V.L., et al., *Public awareness about depression: The effectiveness of a patient guideline*. The International Journal of Psychiatry in Medicine, 2004. 34(1): p. 1-20.
135. Farquhar, C.M., E.W. Kofa, and J.R. Slutsky, *Clinicians' attitudes to clinical practice guidelines: a systematic review*. The medical journal of Australia, 2002. 177(9): p. 502-506.
136. Croft, P., A. Malmivaara, and M. van Tulder, *The pros and cons of evidence-based medicine*. 2011, LWW.

137. Woolf, S.H., *Shared decision-making: the case for letting patients decide which choice is best*. Journal of Family Practice, 1997. 45(3): p. 205-209.
138. Brezis, M., *Big pharma and health care: unsolvable conflict of interests between private enterprise and public health*. The Israel journal of psychiatry and related sciences, 2007. 45(2): p. 83-9; discussion 90-4.
139. Every-Palmer, S. and J. Howick, *How evidence-based medicine is failing due to biased trials and selective publication*. Journal of evaluation in clinical practice, 2014. 20(6): p. 908-914.
140. Flottorp, S.A., et al., *A checklist for identifying determinants of practice: a systematic review and synthesis of frameworks and taxonomies of factors that prevent or enable improvements in healthcare professional practice*. Implementation Science, 2013. 8(1): p. 35.
141. Venot, A., A. Burgun, and C. Quantin, *Informatique Médicale, e-Santé—Fondements et applications*. 2013: Springer Science & Business Media.
142. Bradley, J.S., et al., *The management of community-acquired pneumonia in infants and children older than 3 months of age: clinical practice guidelines by the Pediatric Infectious Diseases Society and the Infectious Diseases Society of America*. Clinical infectious diseases, 2011. 53(7): p. e25-e76.
143. Peleg, M., *Computer-interpretable clinical guidelines: a methodological review*. Journal of biomedical informatics, 2013. 46(4): p. 744-763.
144. Jenders, R.A., R. Corman, and B. Dasgupta. *Making the standard more standard: a data and query model for knowledge representation in the Arden syntax*. in *AMIA annual symposium proceedings*. 2003. American Medical Informatics Association.
145. Gardner, R.M., T.A. Pryor, and H.R. Warner, *The HELP hospital information system: update 1998*. International journal of medical informatics, 1999. 54(3): p. 169-182.
146. Elkin, P., et al., *Toward standardization of electronic guideline representation*. Md Computing, 2000. 17(6): p. 39-44.
147. Ohno-Machado, L., et al., *The guideline interchange format: a model for representing guidelines*. Journal of the American Medical Informatics Association, 1998. 5(4): p. 357-372.
148. Condamines, A. and N. Aussenac-Gilles, *Corpus et terminologie, AS-34 ASSTICOT-RTP DOC STIC CNRS-Rapport d'activités*. 2003, AS-34 ASSTICOT-RTP DOC STIC CNRS-.
149. Bachimont, B., *Engagement sémantique et engagement ontologique: conception et réalisation d'ontologies en ingénierie des connaissances*. Ingénierie des connaissances: évolutions récentes et nouveaux défis, 2000: p. 305-323.

150. Troncy, R., *Formalisation des connaissances documentaires et des connaissances conceptuelles à l'aide d'ontologies: application à la description de documents audiovisuels*. 2004, Université Joseph-Fourier-Grenoble I.
151. Attal, P., *Sémantique pour l'analyse. De la linguistique à l'informatique (1994)*, François Rastier, Marc Cavazza & Anne Abeille. *Linx*, 1995. 33(2): p. 193-197.
152. Grenon, P., B. Smith, and L. Goldberg, *Biodynamic ontology: applying BFO in the biomedical domain*. *Studies in health technology and informatics*, 2004: p. 20-38.
153. Aronson, A.R., *Metamap: Mapping text to the umls metathesaurus*. Bethesda, MD: NLM, NIH, DHHS, 2006: p. 1-26.
154. Guarino, N., D. Oberle, and S. Staab, *What is an ontology?*, in *Handbook on ontologies*. 2009, Springer. p. 1-17.
155. Katsumi, M. and M. Grüninger, *Choosing ontologies for reuse*. *Applied Ontology*, 2017. 12(3-4): p. 195-221.
156. Grüninger, M. and M.S. Fox, *Methodology for the design and evaluation of ontologies*. 1995.
157. Bezerra, C., F. Freitas, and F. Santana. *Evaluating ontologies with competency questions*. in *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*. 2013. IEEE.
158. Brank, J., M. Grobelnik, and D. Mladenic. *A survey of ontology evaluation techniques*. in *Proceedings of the conference on data mining and data warehouses (SiKDD 2005)*. 2005. Citeseer Ljubljana, Slovenia.
159. Raad, J. and C. Cruz. *A survey on ontology evaluation methods*. in *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*. 2015. Lisbon, Portugal.
160. Hlomani, H. and D. Stacey, *Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey*. *Semantic Web Journal*, 2014. 1(5): p. 1-11.
161. Cimino, J.J., *Desiderata for controlled medical vocabularies in the twenty-first century*. *Methods of information in medicine*, 1998. 37(4-5): p. 394.
162. Corbridge, C., et al., *Laddering: technique and tool use in knowledge acquisition*. *Knowledge Acquisition*, 1994. 6(3): p. 315-341.
163. Hulshof, C.T., *1710d Systematic reviews and evidence-based guidelines, two of a different kind?* 2018, BMJ Publishing Group Ltd.
164. Metlay, J.P., et al., *Diagnosis and treatment of adults with community-acquired pneumonia. An official clinical practice guideline of the American Thoracic Society and Infectious Diseases Society of America*. *American journal of respiratory and critical care medicine*, 2019. 200(7): p. e45-e67.

165. Johnson, A.E., et al., *MIMIC-III, a freely accessible critical care database*. *Scientific data*, 2016. 3(1): p. 1-9.
166. Peroni, S., D. Shotton, and F. Vitali. *The live OWL documentation environment: a tool for the automatic generation of ontology documentation*. in *International Conference on Knowledge Engineering and Knowledge Management*. 2012. Springer.