

# Un Serveur Multi-Terminologies de Santé

Michel Joubert<sup>1</sup>, Pierre-Yves Vandebussche<sup>2</sup>, Badisse Dahamna<sup>3</sup>,  
Jean Delahousse<sup>2</sup>, Stefan Darmoni<sup>3</sup> et Marius Fieschi<sup>1</sup>

<sup>1</sup>LERTIM, EA3283, Faculté de Médecine, Université d'Aix-Marseille II  
<http://www.lertim.org>

<sup>2</sup>MONDECA, Paris

<http://www.mondeca.com>

<sup>3</sup>CISMeF, Centre Hospitalier de Rouen

<http://www.chu-rouen.fr/cismef/>

**Résumé :** Il existe dans le domaine de la santé pratiquement autant de terminologies que de champs d'application. De ce fait, et devant le besoin grandissant de coopération entre les acteurs de la santé, apparaît la nécessité de concevoir un modèle apte à rendre ces terminologies « interopérables », ce qui nécessite un référentiel sémantique commun. Deux équipes hospitalo-universitaires, LERTIM (Marseille) et CISMeF (Rouen), et une société d'informatique, MONDECA, se sont associées pour intégrer une douzaine de terminologies de santé dans un système unique apte à les exploiter simultanément. Elles ont tout d'abord conçu un méta-modèle des terminologies. Ce méta-modèle a ensuite été confronté au modèle d'ITM®, la plateforme de MONDECA, pour leur intégration. Le résultat est un serveur opérationnel multi-terminologies intégrant de manière homogène les terminologies le plus fréquemment utilisées dans le domaine de la santé. L'étape suivante consistera à intégrer d'autres terminologies moins usitées mais tout aussi importantes.

**Mots-clés :** terminologies de santé, intégration de terminologies, alignements de terminologies, Unified Medical Language System, interopérabilité sémantique.

## 1 Introduction

### 1.1 Contexte

Il existe dans le domaine de la santé pratiquement autant de terminologies que de champs d'application. Ce sont, pour les plus utilisées d'entre elles :

- SNOMED pour le codage des données cliniques (la France a acquis SNOMED International [1] et SNOMED Clinical Terms [2] est maintenant largement acceptée dans le monde anglophone, mais non encore traduite en français
- CIM10 [3] et CCAM [4] sont réglementairement utilisées à des fins épidémiologiques et médico-économiques dans tous les établissements français de soins publics ou privés

- CISP [5] est une classification conçue pour la médecine de famille et les soins primaires à l'initiative de la WONCA
- MeSH [6] est le thesaurus d'indexation et de recherche d'information privilégié dans le domaine de la santé dans le monde entier
- les domaines de la pharmacologie et de la pharmacovigilance utilisent la classification ATC [7], MedDRA [8] et WHO-ART [9]
- LOINC [10] est un standard pour enregistrer électroniquement les résultats de laboratoire (l'Assistance Publique – Hôpitaux de Paris collabore à sa francisation)

Devant ce constat et la nécessité grandissante de permettre la coopération de différents acteurs de la santé, il apparaît nécessaire de rendre ces terminologies « interopérables ». Nous entendons par là, la mise en œuvre d'un référentiel sémantique commun permettant l'interaction efficace, avec un minimum de perte de sens. L'interopérabilité sémantique nécessite à nos yeux un modèle commun de représentation des termes, quelle que soient leurs terminologie ou référentiel d'origine, ainsi que les moyens de mettre en relation les termes d'une terminologie avec ses équivalents, directs ou indirects, dans d'autres référentiels.

Le projet UMLS [11] de la « National Library of Medicine » (NLM) des Etats-Unis intègre les terminologies de santé dans un thesaurus unique appelé Metathesaurus. Plus d'une centaine de terminologies à ce jour, pour plus de 1.400.000 concepts. Malgré tout, le Metathesaurus ne rend pas les terminologies intégrées interopérables au sens où nous l'avons énoncé ci-dessus puisqu'il intègre les différentes terminologies telles qu'elles se présentent sans établir de liens entre les termes de celles-ci autrement que par le rattachement de termes équivalents à un même identifiant du Metathesaurus. Dans le monde anglophone, des chercheurs en science de l'information et en informatique médicales se sont penchés sur le problème de la mise à disposition de serveurs de terminologies dans le domaine de la santé [12, 13]. De ces études sont nés un certain nombre de systèmes, dont le plus célèbre est sans doute en Europe le système GALEN [14]. Depuis, des projets et initiatives ont suivi cette voie :

- SYMBIOmatics dans le domaine des « Information & Communication Technologies for Health » [15]
- L'objectif du réseau d'excellence intitulé « Semantic Interoperability and Data Mining in Biomedicine » (NoE) [16] est de faire de l'Europe le leader scientifique international en informatique médicale et biomédicale
- Le projet SemanticHEALTH [17] a pour ambition de fédérer au niveau européen les différentes actions en faveur de l'interopérabilité sémantique des systèmes de santé et de leurs infrastructures
- Le CEN [18] pour la normalisation en informatique de santé en Europe

- Health Language, Inc. [19] qui délivre une plateforme offrant des services facilitant l'accès aux terminologies standard et le « mapping » entre elles

L'intérêt est donc grand aujourd'hui de non seulement développer et maintenir des terminologies de santé, mais aussi de les rendre interoperables dans des systèmes informatiques délivrant des services à des applications.

## **1.2 Un serveur multi-terminologies de santé**

Le LERTIM, CISMef et l'industriel MONDECA, tous trois spécialisés dans la représentation et le traitement des terminologies et des connaissances se sont associés pour concevoir et réaliser un « serveur multi-terminologies de santé » francophone (SMTS) [20]. Ces trois partenaires ont réalisé l'intégration d'une douzaine de terminologies francophones, dont celles citées ci-dessus (à l'exception de LOINC actuellement), dans un système informatique unique capable de les exploiter simultanément. Ces exploitations sont multiples et correspondent à des services particuliers en rapport avec des tâches utiles dans les systèmes d'information en santé : aide au codage de l'activité des professionnels de santé, communication entre applicatifs utilisant des nomenclatures différentes, intégration dans des entrepôts de données de bases de données hétérogènes, analyse de données dont le codage a été réalisé avec des référentiels ayant évolué dans le temps, etc. Une vision de l'architecture « trois-tiers » du système est présentée par le schéma de la Figure 1. Le niveau le plus haut représente l'outil de gestion des terminologies. Le niveau intermédiaire représente les outils d'exploitation des terminologies et en particulier les services utiles aux applications. Le niveau inférieur illustre un certain nombre d'applications qui peuvent utiliser les services offerts.

L'objet du présent travail était de :

- Concevoir un méta-modèle de terminologies dans lequel chaque modèle de terminologie puisse s'intégrer [21]
- Concevoir et développer un processus d'intégration des terminologies dans un serveur multi-terminologies conforme au méta-modèle ci-dessus
- Construire et intégrer des alignements entre terminologies au sein dudit serveur

Cette approche présente l'avantage de combiner le respect des structures originelles de chacune des terminologies avec un regroupement des méta-données inhérentes à chacune d'elles.

Suite à cette première réalisation ces trois mêmes partenaires se sont associés à deux sociétés industrielles, VIDAL SA et Memodata, deux équipes hospitalo-universitaires, DSPIM de Saint Etienne et LabSTIC de Nice, l'équipe LIMSI du CNRS Orsay, la fondation Health On the Net de Genève, dans un projet pré-

industriel, InterSTIS<sup>1</sup>, soutenu par l'Agence Nationale de la Recherche en réponse à l'appel à projets TecSan (Technologies pour la Santé).

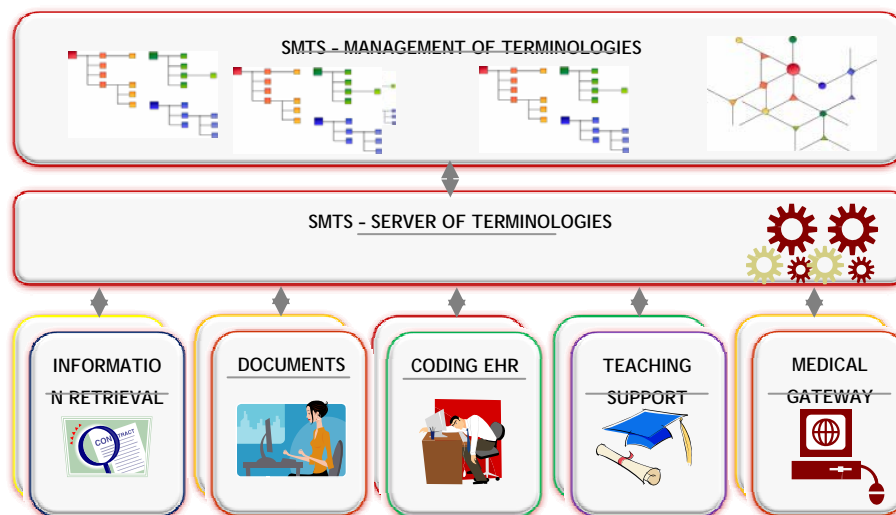


Fig. 1 –Architecture et services délivrés par SMTS

## 2 Matériel et méthodes

### 2.1 « Unified Medical Language System » (UMLS)

L'un des composants de l'UMLS est son Metathesaurus [22]. De manière plus particulière, nous en utiliserons :

- Une table nommée MRCONSO qui recense de manière unique chaque concept répertorié dans UMLS et auquel un identificateur est attribué (CUI)
- Une table nommée MRREL qui décrit les relations, lorsqu'elles existent, entre les concepts dans leurs terminologies d'origine
- Une table nommée MRMAP qui décrit des alignements explicites entre les terminologies intégrées dans UMLS

Quand un alignement explicite existe entre deux concepts  $CUI_1$  et  $CUI_2$  de deux terminologies (ex. ICD-9-CM vers SNOMED CT [23]), on peut en déduire que ces deux concepts sont alignés dans les différentes terminologies dans lesquelles ils sont représentés, et quel que soit le langage dans lequel ils sont désignés. En d'autres termes, des alignements explicites entre terminologies peuvent être « réutilisés » pour

<sup>1</sup> Interopérabilité Sémantique des Terminologies dans les Systèmes d'Information de Santé français. Projet ANR-07-TecSan-010. <http://www.interstis.org>.

d'autres terminologies du fait de l'organisation conceptuelle de UMLS [24]. Ce travail a été réalisé avec la version 2009 AA de UMLS.

## 2.2 Alignements de terminologies

La méthode d'alignement que nous avons adoptée s'inspire de celles décrites dans les définitions des propriétés de SKOS [25]:

- *ExactMatch* est utilisé pour lier deux concepts possédant un haut degré de similitude qui fait qu'ils sont synonymes (cette propriété est transitive)
- *CloseMatch* est utilisé pour lier deux concepts possédant un degré de similitude suffisant pour qu'ils puissent être échangés dans certains cas (cette propriété n'est pas transitive)
- *BroadMatch* and *NarrowMatch* sont utilisés pour établir des liens hiérarchiques entre concepts plus généraux et plus spécifiques
- *RelatedMatch* est utilisé pour décrire des liens d'association entre concepts

La méthode d'alignement que nous appliquons est la suivante : supposons deux termes  $t_1$  et  $t_2$  de deux terminologies  $T_1$  et  $T_2$  respectivement; supposons  $CUI_1$  et  $CUI_2$  les projections respectives de  $t_1$  et  $t_2$  dans le Metathesaurus, alors  $t_1$  et  $t_2$  sont alignés si :

- $CUI_1=CUI_2$  (dans MRCONSO), ceci correspond à la propriété SKOS *ExactMatch*, ou
- Un parent dans la hiérarchie de  $t_1$  est aligné avec  $t_2$  ou inversement (grâce à MRREL), ceci correspond aux propriétés SKOS *BroadMatch* et *NarrowMatch*, ou
- Il existe un alignement explicite entre  $CUI_1$  and  $CUI_2$  (dans MRMAP), ceci correspond à la propriété SKOS *CloseMatch*

L'algorithme est déroulé séquentiellement et s'arrête dès qu'une tentative d'alignement est fructueuse.

Pour illustrer la dernière étape de l'algorithme ci-dessus, supposons qu'un alignement explicite entre deux concepts provienne de terminologies  $T_1$  et  $T_2$  qui ne sont pas intégrées dans SMTS (ex. ICD-9-CM et SNOMED CT en anglais), nous l'appliquons à  $t_1$  et  $t_2$  puisqu'il a été établi entre  $CUI_1$  et  $CUI_2$  que  $t_1$  et  $t_2$  désignent en français. Nous discuterons plus loin la propriété SKOS *RelatedMatch* que nous envisageons d'implémenter ultérieurement.

## 3 Résultats

La conception et la réalisation du SMTS se sont faites en trois étapes :

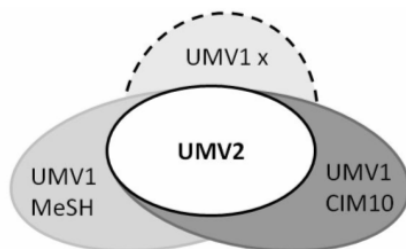
- Conception d'un modèle général capable de représenter chacune des terminologies qui soit compatible avec le mode de représentation des

terminologies du logiciel ITM<sup>®</sup> (Intelligent Topic Manager) de MONDECA [26]

- Développement pour chaque terminologie d'un traducteur pour l'intégrer dans le système
- Réalisation et intégration d'alignements inter-terminologies

### 3.1 Modélisation des terminologies

Nous avons conçu un modèle de chacune des terminologies. La confrontation de ces modèles a permis de dégager un méta-modèle dont chacun des modèles initiaux est une spécialisation. L'objectif de ce méta-modèle est de factoriser les artefacts (classes, relations, attributs) communs à l'ensemble des terminologies. Cette factorisation facilite l'intégration de multiples terminologies au sein d'une même plateforme. L'identification des éléments communs des terminologies est subjective : elle se base sur des standards et sur notre expérience concernant la gestion de terminologies pour prendre la forme actuelle. Des artefacts, spécifiques à certaines terminologies, doivent néanmoins être représentés pour ne pas perdre d'information en dehors du méta-modèle unifié. Nous avons donc un équilibre à choisir : représenter fidèlement une terminologie sans perte d'information tout en extrayant des artefacts communs aux terminologies dans le but d'offrir par la suite des services communs indépendants d'une terminologie en particulier. Il advient donc de distinguer le méta-modèle unifié que nous appellerons UMV2 de ses extensions spécifiques à chaque terminologie que nous appellerons UMV1x (où x est une terminologie particulière) comme illustré en Figure2.



**Fig. 2** –Représentation du méta-modèle unifié (UMV2) et de ses extensions spécifiques à chaque terminologie (UMV1x).

La modélisation de l'UMV2 se veut être un condensé de nos besoins pratiques en nous inspirant des standards comme SKOS [25] et ISO 25964-5 [27]. Notre volonté est de rester indépendant de tous standards tout en restant toujours interopérable avec ceux-ci. Nous avons donc préféré utiliser notre propre méta-modèle que d'étendre celui d'un standard. Ce choix se justifie également par la définition de nouvelles fonctionnalités rendues possible par le méta-modèle (comme par exemple la définition en intension des concepts contenus dans un groupe). Tout élément du modèle peut-être étendu.

*Concept* est l'élément central de l'UMV2 ; il possède des attributs: un *label* représentant le terme préféré de ce concept, des *synonymes*, une *notation* (le code du concept servant par exemple à l'indexation d'un document), etc. Un *Concept* possède également des relations telles que *BroaderNarrowerRelation* qui permet de l'inscrire dans une hiérarchie ou encore *RelatedRelation* pour indiquer des concepts connexes.

*Group* est l'entité qui représente un ensemble de *Concepts*. Par exemple nous utiliserons la sous-classe *Terminologie* pour représenter : une terminologie (ex : CIM10, MeSH) ou encore des sous-ensemble de terminologies (ex : SMQ [*Standardised MedDRA Queries*] de MedDRA). Les groupes peuvent être organisés hiérarchiquement avec la relation *SuperSubGroup*. L'appartenance de *Concepts* à un *Group* peut se représenter de deux manières :

- Par extension : l'appartenance d'un *Concept* à un *Group* est déclarée explicitement au travers de la propriété *inGroup*
- Par intention : l'appartenance d'un *Concept* à un *Group* est déclaré implicitement comme étant l'ensemble des *Concepts* répondant à une *query*.

La figure 3 présente le diagramme de classes UML [28] de notre méta-modèle de terminologies.

Nous illustrons l'extension UMV1 avec la terminologie CIM10 en Figure 4. Nous avons défini une classe abstraite *CIM10Concept* permettant de spécifier la relation particulière à CIM10 : *CIM10Exclusion* mais également de définir des attributs particuliers à cette terminologie. Nous avons ensuite défini une sous-classe de *CIM10Concept* pour chaque niveau de la hiérarchie de concepts, ceci permet de définir des attributs spécifiques à un niveau particulier.

### 3.2 Intégration des terminologies

Les terminologies sont présentées à ITM<sup>®</sup> dans le langage OWL-DL [29] au format RDF/XML. Les alignements entre terminologies sont également présentés dans le format RDF/XML [30]. Nous utilisons la logique de description que fournit le langage OWL pour représenter l'extension du modèle général pour chaque terminologie (définition de classes, de restrictions, de datatypes et d'objects properties). Le contenu de la terminologie est ensuite représenté comme instance OWL des classes du modèle.

La méthode que nous avons adoptée est la suivante :

- Création d'une représentation en OWL de chaque terminologie
- Intégration de chaque terminologie dans la base de données d'ITM<sup>®</sup>
- Insertion des alignements entre les terminologies

Chaque terminologie et alignement sont gérés dans un environnement de travail différent au sein de la plateforme. Ceci permet de faciliter la maintenance et les évolutions de ces ressources.

Cette approche offre une grande flexibilité puisqu'une terminologie peut être intégrée à tout moment. Ses alignements avec les terminologies déjà présentes, s'ils existent, sont intégrés séparément, encore une fois à n'importe quel moment.

Le modèle général validé a été confronté à celui d'ITM<sup>®</sup> afin de l'implémenter. Ceci ayant été fait, nous avons réalisé pour chaque terminologie le traducteur adapté afin de l'intégrer dans la base de données des terminologies d'ITM<sup>®</sup>.

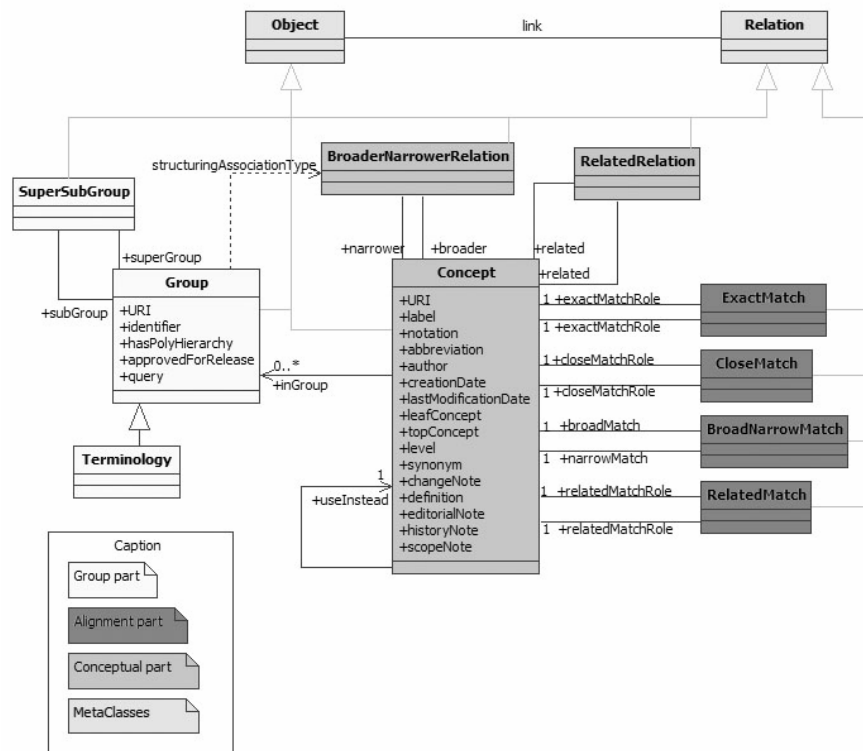


Fig. 3 – Diagramme de classes UML du méta-modèle de SMTS

### 3.3 Production du format RDF/XML d'une terminologie

La production du format RDF/XML d'une terminologie se fait en deux étapes :

- Un pré-traitement pour modéliser la terminologie et la présenter dans un format standard
- L'écriture d'un traducteur pour produire au format RDF/XML le contenu de la terminologie décrite par son modèle

Le modèle d'une terminologie est réalisé en UML. Son contenu est formaté en SQL ou en XML. Une opération manuelle est nécessaire pour assurer la conformité entre le modèle UML avec celui produit en SQL ou XML. Ceci est illustré par le schéma de la Figure 5. La seconde étape consiste à traduire le modèle UML d'une terminologie en un ensemble de classes Java que le traducteur exploitera ensuite. Pour cela nous utilisons le logiciel SWOOP [31]. Le traducteur produit ensuite une



représentation de la terminologie conforme au langage OWL-DL. Un dernier traitement (sérialisation) instancie les différentes classes du modèle OWL avec les données d'origine de la terminologies en tant qu'objets permanents dans la base de données d' ITM®.

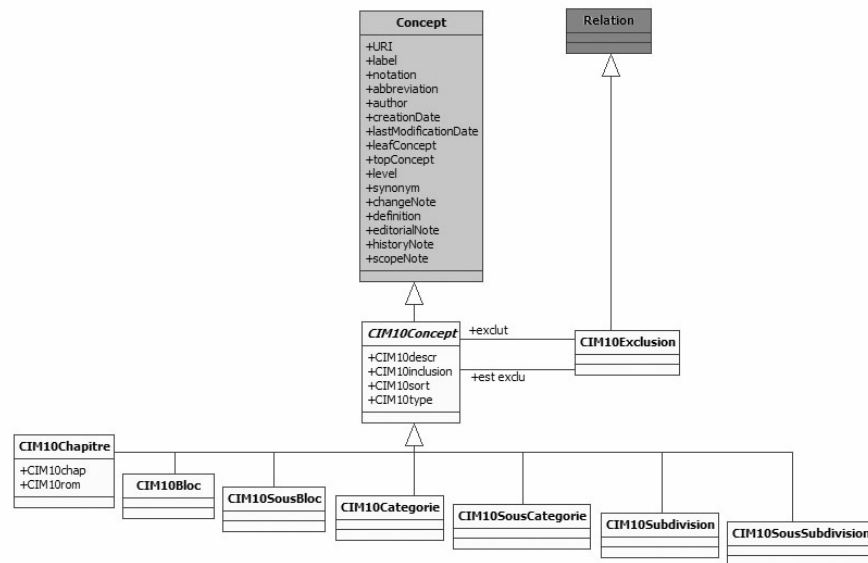


Fig. 4 –Diagramme de classes UML de CIM10

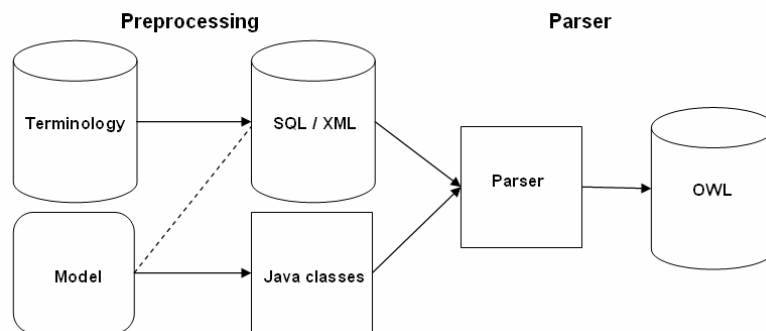


Fig. 5 –Processus de production dans le langage OWL d'une terminologie

### 3.4 Alignements entre terminologies

Les alignements entre terminologies sont opérés deux à deux comme décrit précédemment. Les résultats sont stockés dans une base de données relationnelle afin d'en extraire trois fichiers au format RDF pour chaque paire de terminologies traitées : un fichier d'*ExactMatch*, un fichier décrivant les *BroadNarrowMatch*, et un dernier fichier décrivant les *CloseMatch*. Ces fichiers sont ensuite fournis à ITM<sup>®</sup> pour intégration dans le système. Ce travail ne porte que sur les termes préférentiels (PTs) de chacune des terminologies.

La table 1 montre les alignements opérés entre trois terminologies : CIM10 (9.306 PTs), SNOMED Int. (SNMI, 107.900 PTs) et MeSH (25.186 PTs). Une cellule à l'intersection d'une ligne et d'une colonne indique le nombre de termes préférentiels de la terminologie en ligne alignés avec ceux de la terminologie en colonne. Par exemple, 6.891 (74 %) des termes de CIM10 sont alignés avec des termes de SNOMED. Alors que 19.757 (18 %) des termes de SNOMED sont alignés avec des termes de CIM10.

	CIM10	SNMI	MeSH
CIM10		6.891	4.617
SNMI	19.757		48.390
MeSH	3.605	15.839	

**Table 1.** Alignements entre CIM10, SNOMED et MeSH.

## 4 Discussion

Le SMTS n'est pas une simple intégration statique de différentes terminologies dans un même modèle. Il propose une interopérabilité de ces terminologies enrichie par l'exploitation de leurs modèles de construction et de leurs structures. En effet, le terme « infarctus aigu du myocarde » de SNOMED, qui n'a pas d'équivalent direct dans MeSH est mis en relation avec le terme « infarctus du myocarde » de MeSH. Ce mécanisme appliqué à SNOMED et CIM10 offre, par exemple, la possibilité de proposer automatiquement un codage CIM-10 médico-économique au vu d'une information clinique codée en SNOMED.

Ceci met en évidence un des aspects que nous citons en introduction : toutes les terminologies de santé n'ont pas été conçues avec le même objectif. De ce fait les résultats obtenus en tentant de les mettre en relation terme à terme et deux à deux peuvent paraître contestables. Cependant, si l'on se penche de plus près sur chacune des terminologies à mettre en relation, on peut se rendre compte que ces résultats peuvent être considérablement améliorés. Par exemple, si on aligne CIM10 sans les chapitres XX (causes externes de morbidité et de mortalité) et XXI (facteurs influant sur l'état de santé et motifs de recours aux services de santé) avec l'axe D (maladies)

de SNOMED, on obtient des résultats présentés par la table 2, soit 86% des termes préférentiels de CIM10 sont alignés avec ceux de SNOMED, au lieu de 74% cités précédemment. Nous pensons donc qu'il faudra tenir compte des spécificités de chaque terminologie avant de les exploiter de cette manière.

CIM10 → SNMI	Nb de PTs	% de PTs
Exact match	2.837	36
BroadNarrow match	1.314	17
Close match	2.538	33

**Table 2.** Alignement de CIM10 sans les chapitres XX et XXI avec l'axe D de SNOMED.

C'est ce type de relations que nous comptons développer encore pour rapprocher la SNOMED et la CCAM, par exemple. Mais aussi tous les rapprochements possibles entre les terminologies intégrées dans la base de terminologies du SMTS. L'intégration de nouvelles terminologies francophones doit se faire dans les prochains mois, à commencer par LOINC au fur et à mesure de sa traduction en français, mais aussi le thesaurus Orphanet des maladies rares [32], et le thesaurus unifié que la société VIDAL SA [33] est en cours de construction. Ce dernier doit unifier l'ensemble des terminologies utilisées par VIDAL SA dans sa gestion des informations sur les médicaments : les codes ATC, les indications et contre-indications, les interactions médicamenteuses, etc.

Le diagramme de la Figure 3 ne fait pas apparaître explicitement les notions généralement attendues pour la représentation d'une terminologie. En particulier la notion de *Term* n'y est pas présente. Ni la notion de synonymie entre terme préférentiel et termes synonymes. Ces notions, nous avons choisi de les synthétiser dans la notion de *Concept* et ses attributs par souci de généralité et d'intégration des terminologies qui ne sont pas toutes construites de la même manière en fonction de leurs finalités. La Figure 6 présente une alternative à la représentation d'une terminologie dans notre méta-modèle. Les notions de terme préférentiel et de synonymes y sont introduites au travers de classes reliées au *Concept*. Nous n'avons toutefois pas retenu cette représentation qui est trop complexe pour les terminologies que nous gérons actuellement et a un coût sur la volumétrie de stockage de ces informations.

Toutes les terminologies de santé francophones, ou leurs équivalents anglophones existantes, n'ont pas été intégrées dans UMLS. C'est le cas de CCAM qui est purement française. C'est une classification qui, lorsque l'on décompose l'un de ses codes, décrit un acte médical selon trois axes : action, morphologie et technique. Nous avons décidé d'aligner chaque élément résultant de la décomposition d'un code CCAM avec des concepts du Metathesaurus de UMLS. Pour cela nous avons utilisé une table de correspondance entre termes français et termes anglais venant d'une étude précédente [34] et Metamap [35] pour aligner les termes anglais avec des noms de concepts du Metathesaurus. Pour évaluer cette approche, nous avons décomposé les libellés de SNOMED International en anglais et nous les avons alignés avec les concepts du Metathesaurus à l'aide de Metamap. De cette façon, nous espérons que les

concepts alignés dans le Metathesaurus après décomposition des libellés CCAM pourront être alignés avec ceux obtenus de l'alignement des termes SNOMED. Ceci constitue un mode particulier de *RelatedMatch*, au sens de SKOS, entre deux concepts de terminologies différentes : si les éléments résultant de leurs décompositions sont alignés de manière identique avec des concepts du Metathesaurus, on peut penser qu'ils sont suffisamment proches d'un point de vue sémantique pour être alignés. A la manière des « related articles » qui sont indexés dans une base de données documentaire à l'aide des mêmes mot-clés. Les premiers résultats sont encourageants puisque 45% des décompositions de codes CCAM sont totalement couvertes par des décompositions de concepts SNOMED. Il faut préciser que seuls sont considérés les concepts SNOMED dont tous les éléments résultant de leur décomposition sont alignés avec leurs correspondants issus de CCAM.

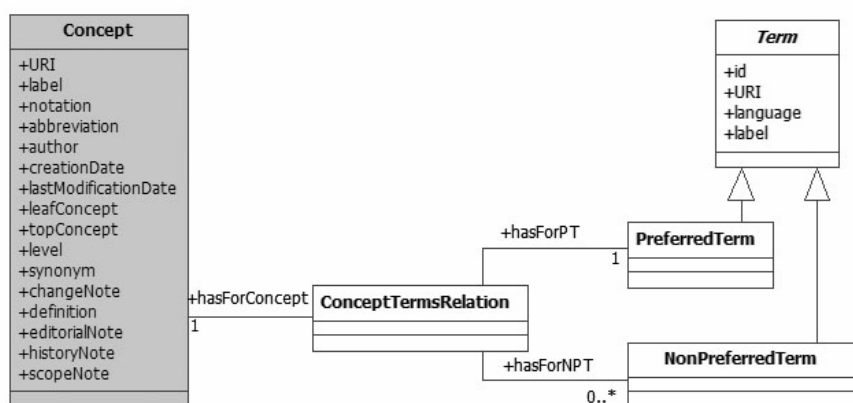


Fig. 6 – Représentation alternative de la partie terminologique de l'UMV2

## 5 Conclusion

Le projet InterSTIS soutenu par l'appel à propositions TecSan de 2007 par l'Agence Nationale pour la Recherche doit nous permettre d'établir définitivement les relations souhaitées entre les termes des terminologies de santé intégrées dans le STMS. L'objectif du projet est de rendre interopérables les terminologies de santé francophones au sein d'un même système informatique. Il est dans la continuité du SMTS présenté ici et ses résultats en seront l'aboutissement. Pour ce faire, nous exploiterons toutes les sources de connaissances génériques possibles dans le domaine biomédical, dont l'UMLS, la SNOMED Clinical Terms et les thesaurus et metathesaurus du « National Cancer Institute » des Etats-Unis [36]. Qui sont des compléments de ceux de UMLS.

Les recherches réalisées pour rendre le contenu des textes médicaux plus facilement compréhensibles par les patients et leurs familles sont d'un grand intérêt. Citons les réflexions de A.T. McCray [37] à propos de la capacité de lecture et de

compréhension de documents relatifs à la santé des personnes concernées. Ces personnes ont besoin d'interagir dans une grande variété d'environnements incluant les cabinets médicaux, les cliniques et hôpitaux. Elles sont aussi confrontées à une quantité importante d'informations relatives à la santé, comme les instructions relatives aux traitements, la formation à l'appareillage, les prescriptions, et l'assurance maladie sous toutes ses formes. Progressivement, ils seront de plus en plus concernés par leur propre prise en charge de leurs états de santé. Face à cette situation, plusieurs projets se sont intéressés à une meilleure compréhension du langage médical par les citoyens. On peut citer : le « Terminology Server » de la NLM [38] dont l'objectif est de réduire l'écart de vocabulaire entre celui des utilisateurs des systèmes de santé et celui des systèmes d'information cliniques et des systèmes de recherche; le projet WRAPIN<sup>2</sup> [39] dont l'objectif était de mettre à disposition des citoyens des sites Internet de santé de qualité accréditée.

Ces actions visent à l'intermédiation sémantique qui rejoint le problème abordé avec le SMTS. Elles visent à bénéficier de l'usage des terminologies standard, sans les faire dévier de leurs objectifs premiers, à des utilisations particulières. C'est sans doute dans cette voie que s'orienteront les travaux autour du SMTS pour l'adapter à l'usage qu'en font ses utilisateurs. Dans le cadre du projet InterSTIS, le laboratoire LabSTIC en expérimente l'usage à des fins d'indexation et de recherche de documents pédagogiques, la Fondation HON intègre ses fonctionnalités dans le moteur de recherche WRAPIN.

La mise en place du SMTS doit déboucher sur la définition de Web services qui exploiteront cette plate-forme d'intermédiation sémantique entre nomenclatures pour favoriser l'activité des utilisateurs et l'exploitation des données recueillies à des fins éventuellement différentes de celle de la finalité qui a présidé à leur recueil.

## **Références**

- [1] Lussier YA, Rothwell DJ, Côté RA. The SNOMED model: a knowledge source for the controlled terminology of the computerized patient record. *Methods Inf Med.* 1998; 37(2): 161-4.
- [2] Stearns MQ, Price C, Spackman KA, Wang AY. SNOMED clinical terms: overview of the development process and project status. *Proc AMIA Symp.* 2001: 662-6.
- [3] World Health Organization. International Classification of Diseases, 10th revision. <http://www.who.int/classifications/icd/en/index.html>
- [4] Assurance Maladie. CCAM. <http://www.ameli.fr/professionnels-de-sante/gestionnaires-de-centres-de-sante/exercer-au-quotidien/codage-actes-medicaux-c.c.a.m./index.php>
- [5] World Organisation of National Colleges, Academies, and Academic Associations of General Practitioners/Family Physicians. <http://www.globalfamilydoctor.com/wicc/icpcstory.html>

---

<sup>2</sup> Projet européen IST2001-33260.

- [6] National Library of Medicine. Medical Subject Headings. <http://www.nlm.nih.gov/mesh/>
- [7] World Health Organization. Collaborating Centre for Drug Statistics Methodology. <http://www.whocc.no/atcddd/>
- [8] Northrop Grumman. MedDRA and the Maintenance and Support Services Organization. <http://www.meddrasso.com/MSSOWeb/index.htm>
- [9] World Health Organization Adverse Reactions Terminology. <http://www.unc-products.com/DynPage.aspx?id=4918>
- [10] Forrey AW, McDonald CJ, DeMoor G et al. Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. *Clin Chem.* 1996; 42(1): 81-90.
- [11] Lindberg DA, Humphreys BL, McCray AT. The Unified Medical Language System. *Methods Inf Med.* 1993; 32(4): 281-91.
- [12] Rector AL, Bechhofer S, Goble CA, Horrocks I, Nowlan WA, Solomon WD. The GRAIL concept modelling language for medical terminology. *Artificial Intelligence in Medicine* 1997, 9: 139-71.
- [13] Chute CG, Elkin PL, Sheretz DD, Tuttle MS. Desiderata for a clinical terminology server. *Proc. AMIA Symp.* 1999: 42-46.
- [14] Rector AL, Solomon WD, Nowlan WA, Rush TW, Zanstra PE, Claassen WM. A Terminology Server for medical language and medical information systems. *Methods Inf Med.* 1995; 34(1-2): 147-57.
- [15] Synergies in Medical Informatics and Bioinformatics. <http://www.symbiomatics.org/>
- [16] Semantic Interoperability and Data Mining in Biomedicine. <http://www.semanticmining.org>
- [17] SemanticHEALTH. <http://www.semantichealth.org>
- [18] European Committee for Standardization. <http://www.cen.eu/cenorm/homepage.htm>
- [19] Health Language, Inc. <http://www.healthlanguage.com/>
- [20] Darmoni S, Joubert M, Dahamna B, Delahousse J, Fieschi M. SMTS: a French Health Multi-terminology Server. *Proc. AMIA 2009*: 808.
- [21] Vandebussche PY, Charlet J. Méta-modèle général de description de ressources terminologiques et ontologiques. *Proc. Ingénierie de la Connaissance 2009*: 193-204.
- [22] National Library of Medicine, UMLS Metathesaurus. <http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html>
- [23] Imel M. A Closer Look: the SNOMED Clinical Terms to ICD-9-CM Mapping. *Journal of AHIMA* 2002; 73: 66-69.
- [24] Fung KW, Bodenreider O. Utilizing the UMLS for semantic mapping between terminologies. *Proc AMIA Annu Symp.* 2005: 266-270.
- [25] World Wide Web Consortium. Simple Knowledge Organization System. <http://www.w3.org/2004/02/skos/>

- [26] MONDECA. Intelligent Topic Manager.  
[http://www.mondeca.com/index.php/en/applications/itm\\_t3](http://www.mondeca.com/index.php/en/applications/itm_t3)
- [27] ISO Technical Committee 46. Thesauri and Interoperability with other Vocabularies. <http://www.niso.org/workrooms/iso25964>
- [28] Object Management Group. Unified Modeling Language.  
<http://www.uml.org/>
- [29] World Wide Web Consortium. Web Ontology Language.  
<http://www.w3.org/TR/owl-features/>
- [30] World Wide Web Consortium. Resource Description Framework.  
<http://www.w3.org/RDF/>
- [31] Kalyanpur A, Parsia B, Hendler J. A Tool for Working with Web Ontologies. *International Journal on Semantic Web and Information Systems* 2005; 1: 36-49.
- [32] Orphanet. <http://www.orpha.net/consor/cgi-bin/index.php?lng=EN>
- [33] Vidal SA. <http://www.vidal.fr/societe/vidal>
- [34] Rodrigues JM, Trombert-Pavot B, Baud R, Wagner J, Meusnier-Carriot F. Galen-In-Use: using artificial intelligence terminology tools to improve the linguistic coherence of a national coding system for surgical procedures. *Stud Health Technol Inform.* 1998: 623-7.
- [35] National Library of Medicine. Metamap.  
<http://skr.nlm.nih.gov/papers/index.shtml#MetaMap>
- [36] National Cancer Institute. <http://www.cancer.gov/cancertopics/terminologyresources>
- [37] McCray A. Promoting Health Literacy. *J Am Med Inform Assoc.* 2005; 12(2): 152-63.
- [38] Cognitive Science Branch. Terminology Server.  
[http://www.lhncbc.nlm.nih.gov/cgsb\\_site/servlet/Turbine/template/research%2Cmedterm%2CTermServer.vm](http://www.lhncbc.nlm.nih.gov/cgsb_site/servlet/Turbine/template/research%2Cmedterm%2CTermServer.vm)
- [39] Joubert M, Gaudinat A, Boyer C, Fieschi M, HON Foundation Council members. WRAPIN : a tool for patient empowerment within EHR. *Stud Health Technol Inform.* 2007; 129: 147-51.