

# Mapping de terminologies diagnostiques en cancérologie par l'intermédiaire du NCI Metathesaurus

## *Mapping of diagnostic terminologies in oncology using the NCI Metathesaurus*

Bérénice Brechat<sup>1,2</sup> Fleur Mouglin<sup>1</sup> Frantz Thiessard<sup>1,2</sup> Vianney Jouhet<sup>1,2</sup>

<sup>1</sup> *Equipe de Recherche en Informatique Appliquée à la Santé, ISPED, Université de Bordeaux*

<sup>2</sup> *Service d'Information Médicale, Pôle de Santé Publique, CHU de Bordeaux*

### Résumé

En France, plusieurs terminologies sont utilisées en routine pour coder des diagnostics de cancer (CIM10 pour le PMSI, CIM-O3 au sein des registres, ADICAP pour l'anatomo-pathologie). Cette hétérogénéité est un frein majeur pour l'utilisation secondaire des données diagnostiques recueillies. Il est nécessaire de mettre en correspondance les différentes terminologies diagnostiques de cancérologie au sein d'un système cohérent et hiérarchisé. Le NCI Thesaurus, qui est une terminologie internationale visant à représenter l'ensemble des connaissances en cancérologie, semble être une excellente base pour constituer ce système unifié. Dans ce travail, le mapping existant entre la CIM-O3 et le NCI Thesaurus au sein du NCI Metathesaurus est considéré. Plus de deux tiers des codes CIM-O3 sont associés à un code NCI Thesaurus unique. Une analyse qualitative des mappings multiples et des cas où un code CIM-O3 n'a pas de correspondant NCI Thesaurus a été réalisée. Différents types de causes ont été identifiés et des pistes de solution sont proposées en discussion.

### Abstract

*Many terminologies are used in France for coding cancer diagnoses (ICD-10 for Diagnosis Related Groups, ICDO3 in cancer registries, ADICAP for pathological anatomy). This heterogeneity largely hinders the use of registries' data. It is thus necessary to integrate the diagnostic terminologies in oncology into a unified and structured system. The NCI Thesaurus is an international terminology, which aims at representing the whole knowledge about oncology. This resource seems to be an excellent basis for constituting this integrated system. In this work, we are investigating the mapping existing between the ICDO3 and the NCI Thesaurus through the NCI Metathesaurus. More than two-third of ICDO3 codes are mapped to a unique NCI Thesaurus code. A qualitative analysis is performed for the multiple mappings and for cases where an ICDO3 code had no corresponding code in the NCI Thesaurus. Some categories of causes were identified and potential solutions are mentioned in the discussion.*

**Mots-clés :** CIM-O3, NCI Thesaurus, mapping de terminologies biomédicales

**Keywords:** ICDO3, NCI Thesaurus, Terminology mapping

*Articles longs des 15<sup>es</sup> Journées francophones d'informatique médicale, JFIM 2014, pages 34–43  
Fès, Maroc, 12–13 juin 2014*

# 1 Introduction

Les activités liées à la santé produisent aujourd'hui une masse importante d'information qu'il faut trier, traiter, recouper pour qu'elle soit exploitable, mais aussi stockée en vue d'une utilisation ultérieure. Face à ces problématiques, les acteurs du domaine médical ont pris conscience de l'importance de structurer l'information qu'ils génèrent [1], mais ces données sont produites par une multitude d'acteurs différents et leur mise en correspondance reste complexe.

Le domaine de la cancérologie illustre bien cette production considérable d'information par des sources multiples. On dispose par exemple de données issues de la clinique, de la recherche ou encore de l'épidémiologie. Pour représenter ces données, un certain nombre de classifications ont été créées et l'enjeu est aujourd'hui de trouver des correspondances entre elles. Les registres des cancers, qui visent à recueillir de façon exhaustive de nombreux éléments relatifs aux cancers incidents dans la population, sont organisés en réseaux avec une volonté de standardiser ce recueil. Le choix s'est porté sur la Classification Internationale des Maladies pour l'Oncologie (CIM-O3) comme nomenclature commune pour la codification des cancers [2]. Pour assurer leur fonction, les registres collectent de l'information issue de sources, et donc de terminologies multiples telles que la classification internationale des maladies ou l'ADICAP (codification des lésions élaborée par l'Association pour le Développement de l'Informatique en Cytologie et en Anatomie Pathologique). Ces registres sont également exploités dans de nombreux travaux et études relatifs au cancer, où leurs données sont croisées avec d'autres bases de données codées grâce à des nomenclatures diverses. Disposer d'un mapping exhaustif entre la CIM-O3 et d'autres terminologies présente donc un intérêt majeur.

Le NCI Thesaurus est une terminologie internationale de référence développée par le National Cancer Institute américain, visant à représenter l'ensemble des connaissances largement utilisée en cancérologie, que ce soit en clinique, en recherche translationnelle et fondamentale, ou encore en épidémiologie. De ce fait, elle semble être un excellent pivot de mapping entre les terminologies du spectre de la cancérologie. On retrouve le NCI Thesaurus, ainsi que la CIM-O3, au sein d'un système unifié de terminologies biomédicales qui contient la plupart des terminologies utilisées dans le domaine de la cancérologie et qui a été développé sur la base de l'UMLS [3] : le NCI Metathesaurus.

L'objectif principal de cette étude est de représenter l'ensemble des concepts de la CIM-O3 au sein du NCI Thesaurus via le NCI Metathesaurus. Pour cela, une analyse de la correspondance entre les concepts CIM-O3/ NCI Thesaurus a été réalisée pour identifier d'éventuelles inexactitudes, erreurs ou absences de mapping.

## 2 État de l'art

La question de l'interopérabilité entre diverses ontologies a été largement étudiée ces dernières années, avec en particulier la volonté de proposer des méthodes automatiques [4,5]. Dans le domaine biomédical également, cette problématique a généré de très nombreux travaux et l'utilisation de l'UMLS comme pivot a été souvent investiguée [6,7,8].

En revanche, il existe peu de travaux où le mapping est basé sur le NCI Metathesaurus [9]. Disposer d'un mapping de qualité entre la CIM-O3 et une terminologie internationale de référence tel le NCI Thesaurus présente un intérêt majeur pour le partage des données épidémiologiques des registres des cancers. On retrouve des travaux portant sur la recherche d'inconsistances d'un tel mapping, réalisés à partir de versions antérieures du NCI Metathesaurus [9]. Cependant, aucune version corrigée et validée de ce type de mapping n'est actuellement disponible. L'actualisation des concepts portés par ce Metathesaurus justifie également un nouveau travail de mapping.

### 3 Matériel et méthodes

#### 3.1 Terminologies utilisées pour le mapping

##### 3.1.1 La CIM-O3

La Classification Internationale des Maladies pour l'Oncologie (CIM-O3 ou ICDO3 en version anglaise), est une classification [10], principalement utilisée par les registres de tumeurs cancéreuses pour le codage des caractéristiques de tumeurs. La CIM-O3 possède un axe morphologique pour décrire la morphologie de la tumeur et un axe topographique qui précise une localisation.

Un code CIM-O3 de tumeur est composé de 10 caractères correspondant à l'agrégation des quatre caractères du code topographique, suivis des cinq caractères du code morphologique, éventuellement suivis d'un caractère facultatif (Figure 1). A chaque code CIM-O3 correspond un terme préféré et des inclusions décrivant généralement des notions plus précises. Par exemple, le code 9440/3 a pour terme préféré « Glioblastome, SAI » et pour inclusions « Glioblastome multiforme », « Spongioblastome multiforme » et « Gliome à cellules indifférenciées ».

La CIM-O3 nécessite donc une post-coordination pour le codage des tumeurs car le diagnostic n'est pas codé tel-que dans la terminologie. Il faut, pour coder un diagnostic précis de tumeur, associer un élément morphologique de la terminologie à un élément topographique de la terminologie.

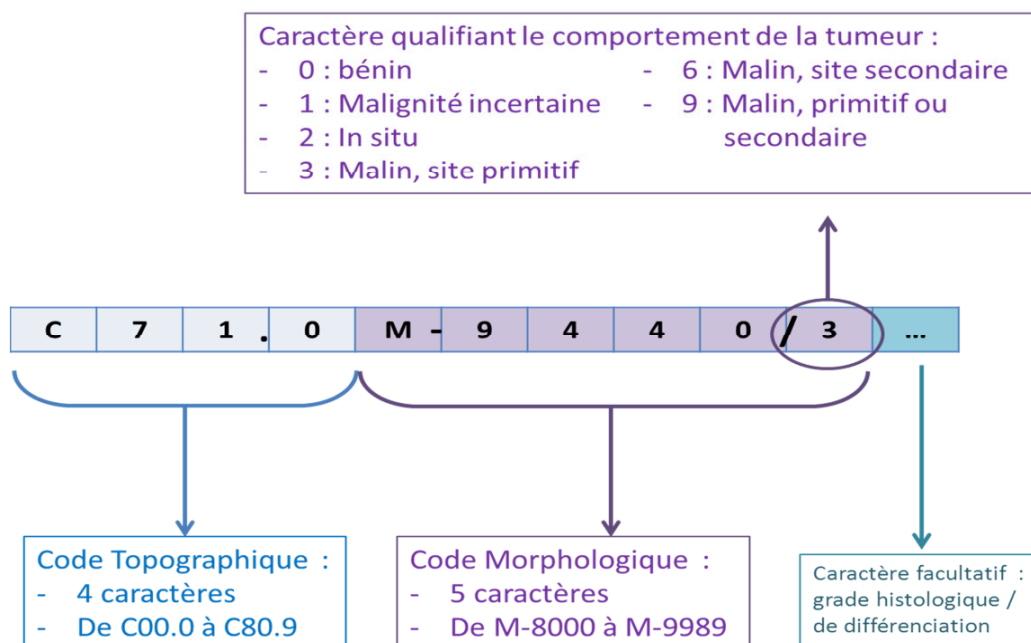


Figure 1 : Eléments permettant la construction d'un code CIM-O3, exemple du glioblastome (code morphologique 9440/3) d'un hémisphère cérébral (code topographique C71.0)

##### 3.1.2 Le NCI thesaurus

Le NCI Thesaurus (NCI) est une terminologie de référence développée par le National Cancer Institute américain, qui couvre le vocabulaire anglais utilisé en cancérologie au sens large [11]. La version utilisée dans ce travail est celle qui est intégrée dans le NCI Metathesaurus (13.06D).

Le NCI comprend plus de 43 000 concepts, organisés selon 20 sous-domaines appelés Kind, tel « Abnormal Cell Kind » qui regroupe les éléments d'anatomo-cyto-pathologie. Associées à ces

concepts, on retrouve des précisions comme le terme préféré (preferred term), des synonymes, des définitions, et d'autres informations complémentaires.

Le NCI possède plus de 90 types de relations définies entre les concepts comme par exemple *disease\_has\_abnormal\_cell* pour qualifier un rapport entre une pathologie et un type de cellule. Aujourd'hui, le NCI possède plus de 200 000 relations entre ses concepts.

Ainsi, le NCI comporte à la fois des entités pré-coordonnées (c'est-à-dire une entité composée d'éléments associés lors de la conception du thésaurus, par exemple « Breast Carcinoma » qui est codé C4872) et des notions pouvant être post-coordonnées grâce aux relations (le terme « Breast Carcinoma » est par exemple relié par *disease\_has\_primary\_anatomic\_site* à « Breast »).

Cette prise en charge des termes pré- et post-coordonnés fait du NCI fait du NCI une ressource pour la construction d'un pivot permettant d'intégrer différentes terminologies diagnostiques en cancérologie. Il permet en effet de représenter aussi bien des terminologies utilisant des termes pré-coordonnés (comme par exemple la CIM-10) que des terminologies nécessitant une post-coordination (comme par exemple la CIM-O3), et donc de mettre en relation ces différents éléments.

### 3.1.3 Le NCI Metathesaurus

Le NCI Metathesaurus est un système utilisé dans le spectre de la cancérologie qui regroupe plus de 75 terminologies biomédicales différentes [12]. Il a été élaboré par le National Cancer Institute américain à partir du Metathesaurus de l'UMLS enrichi par d'autres terminologies biomédicales spécifiques à la cancérologie.

On y retrouve 4 millions de termes représentant 2 millions de concepts biomédicaux.

Ces termes sont reliés entre eux par plus de 22 millions de relations, présentes dans les différentes terminologies. Les différents termes et codes des différentes terminologies représentant un même concept sont codés dans le NCI Metathesaurus sous un code unique : le Concept Unique Identifier (CUI).

A chaque concept est également rattaché l'ensemble des informations complémentaires (définitions, synonymes...) présents initialement dans les terminologies. Enfin, les concepts sont catégorisés par au moins un type sémantique parmi les 133 possibles.

## 3.2 Mapping entre la CIM-O3 et le NCI par le CUI

Les codes de la CIM-O3 ont été mappés aux codes du NCI par correspondance exacte de leur(s) CUI(s). Ainsi, le mapping a été effectué à partir de la liste des CUIs des codes de la CIM-O3. Pour chacun de ces codes, on a recherché le ou les codes du NCI appartenant au même CUI (Figure 2).

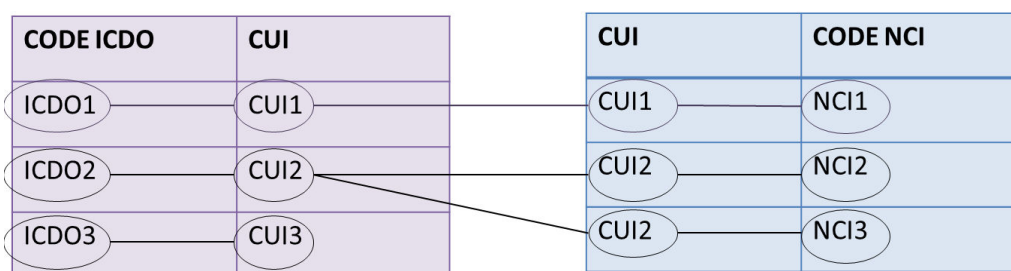


Figure 2 : Principe du mapping des codes de la CIM-O3 avec les codes du NCI par l'intermédiaire des CUI dans le NCI Metathesaurus

Différentes cardinalités de mapping vont émerger :

- Mapping 1-1 : le code CIM-O3 est mappé à un et un seul code du NCI
- Mapping 1-n : le code CIM-O3 est mappé à plusieurs codes du NCI
- Mapping 1-0 : le code CIM-O3 ne retrouve pas de correspondance dans le NCI

L'objectif de l'étude étant de réaliser un mapping de la CIM-O3 vers le NCI, les mappings n-1 n'ont pas été étudiés.

### 3.3 Analyse du mapping code à code entre la CIM-O3 et le NCI

Une analyse qualitative des mappings 1-n et 1-0 a été effectuée pour identifier des motifs de mapping multiple ou d'absence de mapping et tenter de les grouper par type d'erreur.

Cette analyse a été réalisée manuellement afin de mettre en évidence des inexactitudes, des erreurs et des carences de mapping pouvant être résolues, afin de disposer d'un meilleur mapping de la CIM-O3 vers le NCI.

## 4 Résultats

### 4.1 Mapping entre la CIM-O3 et le NCI par le CUI

Le NCI Metathesaurus contient 3634 termes distincts issus de la CIM-O3, codés par 1500 codes CIM-O3 distincts, dont 1091 codes correspondant à des termes de morphologie et 409 codes correspondant à des termes de topographie.

L'utilisation des CUI du NCI Metathesaurus permet de mettre en relation 1261 codes CIM-O3 avec 1670 codes du NCI (Figure 3).

En distinguant le type de code CIM-O3, 963 des 1091 codes morphologiques (soit 88,3%) et 298 des 409 codes topographiques (soit 72,9%) trouvent une correspondance avec un terme du NCI par l'intermédiaire du NCI Metathesaurus. Ainsi, 239 codes de la CIM-O3 (sur les 1500 présents dans le NCI Metathesaurus) ne sont pas mappés à un code NCI (soit 15,9%), correspondant aux mappings de cardinalité 1-0 (Tableau 1). Un exemple de ce type de cardinalité est le code C02.0 codant « Dorsal surface of tongue ».

Tableau 1 : *Récapitulatif du mapping des codes CIM-O3 aux codes NCI par cardinalité du mapping obtenu*

Type de mapping	Nombre (pourcentage)
Mapping 1-1	1008 (67,2%)
Mapping 1-n	253 (16,9%)
Mapping 1-0	239 (15,9%)

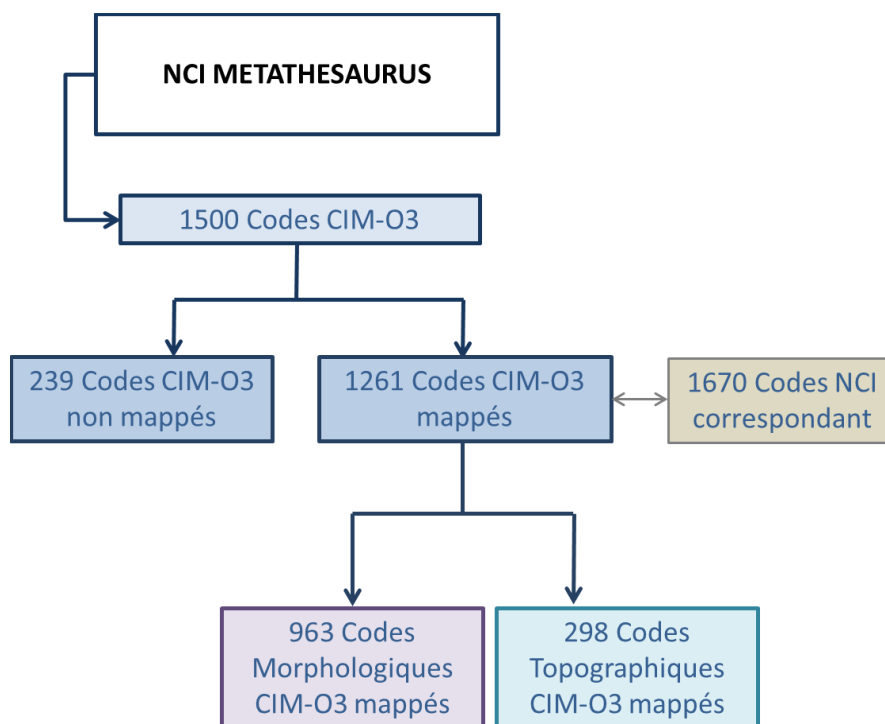


Figure 3 : Répartition et mapping des codes de la CIM-O3 présents dans le NCI Metathesaurus

Par ailleurs, 1008 mappings 1-1 ont été obtenus. Par exemple, le code CIM-O3 8123/3 - « Basaloid carcinoma » est mappé exclusivement au code NCI C4121 - « Basaloid Carcinoma ». Enfin, 253 mappings 1-n ont été identifiés. Par exemple, le code CIM-O3 8000/1 (correspondant au concept de tumeur sans précision) est quant à lui mappé aux codes NCI C65157 et C3262 (Figure 4).

CIM-O3			NCI		
Code CIM-O3	Terme	CUI	CUI	Terme	Code NCI
8000/1	Unclassified tumor, borderline malignancy	C0677041	C0677041	Neoplasm, Uncertain Whether Benign or Malignant	C65157
8000/1	Neoplasm, NOS	C0375111	C0375111	Neoplastic Growth	C3262
8000/1	Tumor, NOS	C0375111	C0375111	Neoplasia	C3262
...	...	...	...	...	...
C30.1	Eustachian tube	C0015183	C0015183	Pharyngotympanic Tube	C12500
C30.1	Auditory tube	C0015183	C0013455	Middle Ear	C12274
C30.1	Middle ear	C0013455			

Code CIM-O3	Code NCI
8000/1	C65157
8000/1	C3262
...	...
C30.1	C12500
C30.1	C12274

Figure 4 : Exemples de mappings 1-n de codes CIM-O3 avec le NCI

## 4.2 Analyse du mapping code à code entre la CIM-O3 et le NCI

### 4.2.1 Différence de granularité

Il existe des mappings multiples dus à la différence de granularité entre les deux terminologies étudiées. Dans ce cas de figure, un même code CIM-O3 est mappé à la fois au code NCI qui lui correspond exactement, mais aussi à ses codes fils plus précis, qui n'ont en fait pas de correspondant dans la CIM-O3. Par exemple, le code CIM-O3 8210/3 - « Adenocarcinoma in adenomatous polyp » est mappé au code NCI C7682 « Carcinoma in a Polyp » mais aussi à son code fils C7676 – « Adenocarcinoma in Adenomatous Polyp » (Figure 5).

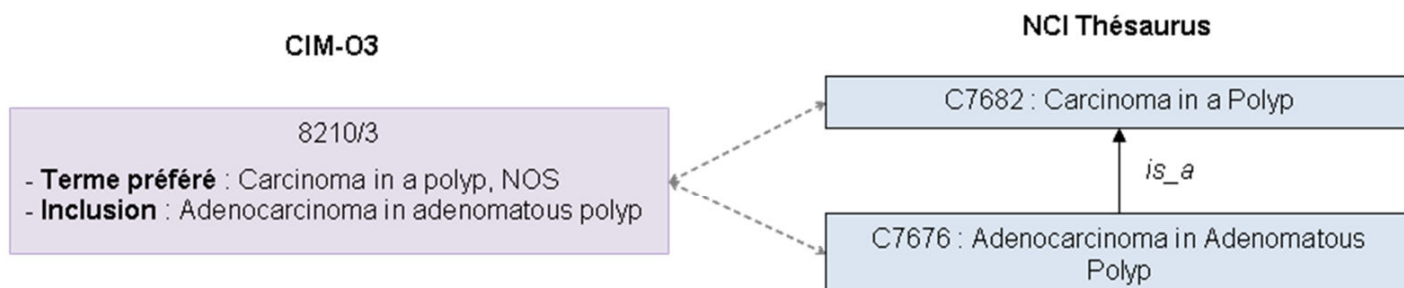


Figure 5 : Exemple de différence de granularité entre la CIM-O3 et le NCI

### 4.2.2 Pré-coordination au sein du NCI

On retrouve des mappings multiples dus à la présence de concepts pré-coordonnés dans le NCI. Les codes NCI associés à ces concepts sont mappés aux codes morphologiques de la CIM-O3 qui sont eux-mêmes également mappés aux codes NCI décrivant uniquement la morphologie correspondante.

Par exemple, le code CIM-O3 8260/3 - « Papillary adenocarcinoma, NOS » est mappé à la fois au concept morphologique « Papillary Adenocarcinoma » codé C2853 dans le NCI, mais aussi au code C6975 du concept pré-coordonné « Papillary Renal Cell Carcinoma » (Figure 6).

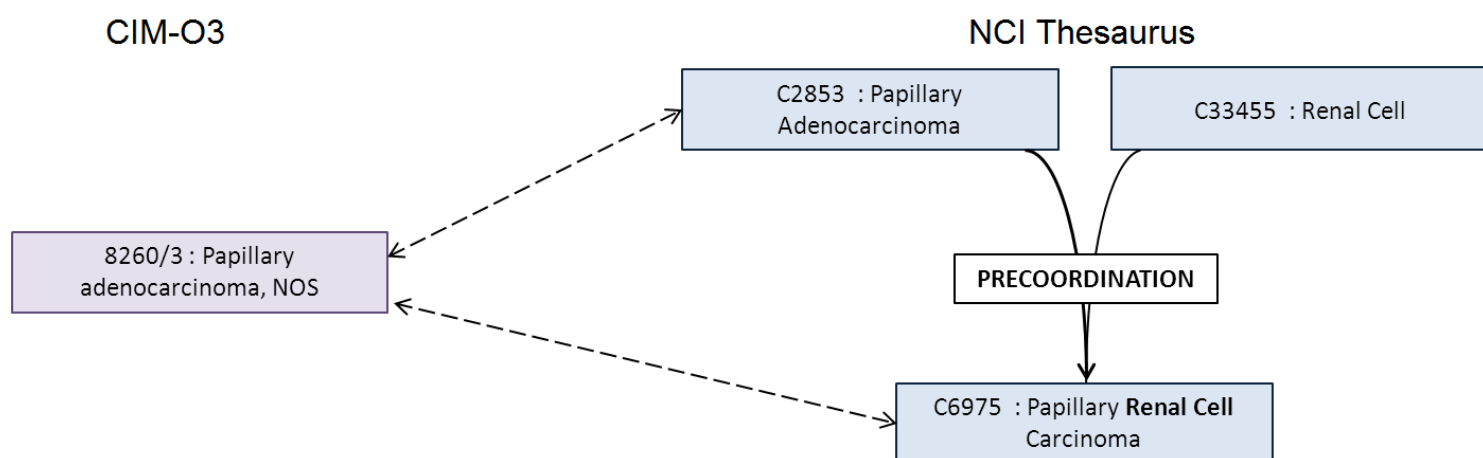


Figure 6 : Exemple de mapping multiple due à la pré-coordination dans le NCI

### 4.2.3 Les mappings 1-0

L'analyse des mappings 1-0 a montré des lacunes dues à la différence de granularité entre les deux terminologies. Ainsi, on retrouve des concepts fins au sein de la CIM-O3 qui ne retrouvent pas de correspondance aussi spécifique au sein du NCI. Par exemple, le code CIM-O3 C02.0 - « Dorsal surface of tongue, NOS » - codant la surface dorsale de la langue ne retrouve pas le même degré de précision au sein du NCI. Dans ce dernier, le concept de langue (C12422 - « Tongue ») est une feuille, sans concept fils.

Par ailleurs, certains mappings 1-0 ne devraient pas l'être car des concepts apparemment synonymes n'appartiennent pas à un même CUI dans le NCI Metathesaurus (problème de « missed synonymy » fréquent dans ce type de large graphe [13]). Par exemple, le code CIM-O3 C42.0 ayant pour terme préféré « Blood » n'est pas mappé au code NCI C12434 ayant pour terme préféré « Blood » car ils sont dans deux CUIs distincts : respectivement, CL429748 et C0005767.

## 5 Discussion

L'utilisation du NCI Metathesaurus a permis de retrouver une majorité de correspondances 1-1 entre la CIM-O3 et le NCI. Cependant, plus de 30% des mappings obtenus sont soit 1-n soit 1-0. Le NCI Metathesaurus ne peut donc pas être utilisé directement pour obtenir un mapping code à code exact de la CIM-O3 vers le NCI. Un travail complémentaire doit ainsi être mené pour parvenir à cet objectif. L'analyse qualitative a permis de mettre en évidence les causes de mappings multiples ou d'absence de mapping dans le but de proposer des solutions ultérieurement. En présence de mappings 1-n, des règles de correction automatiques pourront être définies afin de déterminer la correspondance correcte et d'exclure les autres. Dans cette optique, une approche simple sera de comparer le terme préféré du code CIM-O3 avec les termes des codes NCI associés. Dans la figure 6, on voit que le mapping entre le code 8260/3 « Papillary adenocarcinoma, NOS » et le code NCI C2853 « Papillary Adenocarcinoma » est correct puisque les deux termes sont équivalents tandis que celui entre 8260/3 et C6975 « Papillary **Renal Cell** Carcinoma » est faux. En effet, le terme NCI étant pré-coordonné, il décrit l'organe où se situe le carcinome alors que cette information n'est pas présente dans le code CIM-O3 auquel il est mappé. En considérant les termes associés aux codes NCI, on déterminera automatiquement que le mapping correct est vers C2853. Des méthodes de désambiguïsation par l'intermédiaire de la catégorisation sémantique, comme dans [14], pourraient également être appliquées. Comme l'illustre la figure 7, cela permettrait de résoudre des cas comme celui du code CIM-O3 C38.0 (« Heart » catégorisé par le type sémantique « Body Part, Organ, or Organ Component ») qui est mappé aux codes NCI C12727 (« Heart », également catégorisé par « Body Part, Organ, or Organ Component ») et C4570 (« Malignant Endocardial Neoplasm », catégorisé par le type sémantique « Neoplastic Process »). En filtrant les mappings suivant la catégorisation sémantique, on pourrait écarter automatiquement le deuxième mapping.

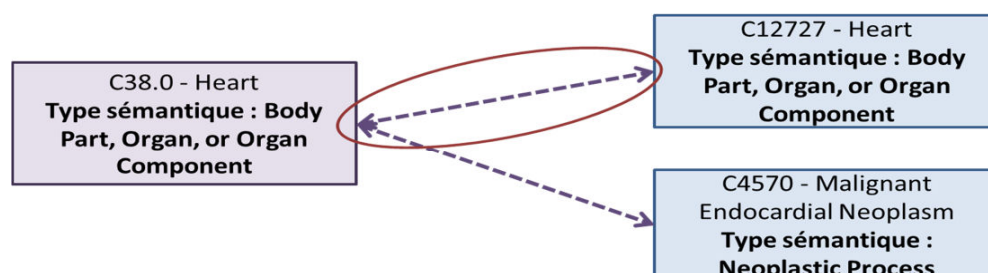


Figure 7 : Exemple de désambiguïsation grâce aux types sémantiques



Pour la plupart des mappings 1-0, la granularité du NCI est insuffisante. Il faudra donc envisager l'ajout de nouveaux codes pour pouvoir représenter les codes CIM-O3 plus fins. Il faudra également traiter les cas dus à la « missed-synonymy ». Par ailleurs, il reste quelques cas isolés de défaut de mapping ne pouvant pas être corrigés automatiquement et pour lesquels une correction manuelle devra être réalisée.

Ce travail de mapping des concepts de la CIM-O3 vers le NCI s'inscrit dans le cadre d'un projet plus large, dont l'objectif est de permettre l'intégration de différentes terminologies diagnostiques en cancérologie. Afin de garantir l'exactitude des mappings obtenus, une évaluation de l'ensemble des résultats sera réalisée par des médecins. Ces mappings validés serviront de jeux de données pour l'évaluation des règles de sélection des mapping multiples.

D'autre part, nous avons identifié un certain nombre de mappings de la forme n-1, à savoir la convergence de plusieurs codes CIM-O3 vers un même code NCI. Par exemple, les codes CIM-O3 8041/3 - « Small cell carcinoma, NOS » et 8042/3 - « Oat cell carcinoma » sont mappés à l'unique code NCI C3915 qui inclut ces deux notions. Ce type de cardinalité n'a pas été pris en compte dans l'analyse qualitative car l'objectif de l'étude était de réaliser un mapping de la CIM-O3 vers le NCI. Cependant, dans un objectif d'intégration des différentes terminologies d'intérêt tout en préservant leur granularité, il est indispensable de traiter ce cas de figure. La prise en compte de ces situations ainsi que le traitement des mappings 1-0 conduiront à l'ajout de termes dans un ensemble basé sur le NCI permettant d'intégrer des terminologies diagnostiques.

Peu de travaux se sont intéressés à l'évaluation des mappings des terminologies diagnostiques à partir du NCI Metathesaurus. Notre travail vise à identifier des mapping uniques et exacts de la CIM-O3 vers le NCI, ce qui diffère de l'objectif principal de l'étude de Burgun et Bodenreider [9] qui cherchait à évaluer la cohérence de la représentation des tumeurs entre ces deux terminologies. Considérant que seuls des termes post-coordonnés existent dans la CIM-O3, nous ne cherchons pas ici à reconstruire les termes pré-coordonnés associant topographie et morphologie pour les mapper vers le NCI. Les résultats du mapping des termes post coordonnés de la CIM-O3 vers le NCI sont cependant similaires à la précédente étude, les différences observées étant dues aux évolutions du NCI Metathesaurus. Le travail de Burgun et Bodenreider [9] s'inscrit dans une perspective d'évaluation de la qualité des terminologies et reste orienté vers l'audit alors que nous cherchons à représenter les connaissances de la CIM-O3 dans une terminologie qui nous permettra, à terme, de mettre en correspondance ses concepts avec d'autres terminologies de cancérologie. Dans cette optique, nous vérifierons que les combinaisons CIM-O3 topographie + morphologie utilisées pour coder des données réelles dans un registre des cancers peuvent être mis en relation avec des termes pré-coordonnés du NCI.

## **6 Conclusion**

Le mapping code à code de la CIM-O3 vers le NCI par l'intermédiaire du NCI Metathesaurus révèle des mappings exacts, des mappings multiples et permet d'identifier les termes absents dans le NCI. L'analyse qualitative a mis en évidence des causes fréquentes aux mappings multiples, ce qui permettra la construction et l'application de règles de correction afin d'obtenir une intégration complète de la CIM-O3 avec le NCI. La confrontation du résultat de cette intégration à des données réelles d'un registre permettra d'identifier l'absence éventuelle de termes pré-coordonnés dans le NCI.

## Références

- [1] Kalra D1, Fernando B, Morrison Z, Sheikh A. *A review of the empirical evidence of the value of structuring and coding of clinical information within electronic health records for direct patient care*. Inform Prim Care. 2012;20(3):171-80.
- [2] Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG, ed. *Cancer registration: principles and methods*. Vol. 95. IARC, 1991
- [3] Lindberg DA, Humphries BL, McCray AT. *The Unified Medical Language System*. *Methods Inf Med*. 1993;32(4):281–291.
- [4] Noy NF. *Tools for mapping and merging ontologies*. In: Staab S, Studer R, editors *Handbook on Ontologies*: Springer-Verlag. 2004;365–384
- [5] Shvaiko P, Euzenat J. *A survey of schema-based matching approaches*. *Journal on Data Semantics*. 2005;4.
- [6] Fung KW, Bodenreider O. *Utilizing the UMLS for semantic mapping between terminologies*. *AMIA Annu Symp Proc*. 2005;266-70
- [7] Mougín F, Burgun A, Bodenreider O. *Comparing drug-class membership in ATC and NDF-RT*. *Proceedings of the 2nd ACM SIGHT International Health Informatics Symposium*, ISBN: 978-1-4503-0781-9, 2012;437-443
- [8] Zhang S, Bodenreider O. *Experience in Aligning Anatomical Ontologies*. *Int J Semant Web Inf Syst*. 2007;3(2):1-26
- [9] Burgun A, Bodenreider O. *Issues in integrating epidemiology and research information in oncology: experience with ICD-O3 and the NCI Thesaurus*. *AMIA Annu Symp Proc*. 2007;85-9.
- [10] Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin M, et al., ed. *International classification of diseases for oncology*. 3<sup>rd</sup> ed. Geneva : World Health Organization. 2000
- [11] Golbeck J, Fragoso G, Hartel FW, Hendler J, Oberthaler J, Parsia B. *The National Cancer Institute's Thesaurus and Ontology*. *J Web Semantics*. 2003;75–80
- [12] NCI Metathesaurus. <http://ncimeta.nci.nih.gov/ncimbrowser/> (consulté le 14/04/2014)
- [13] Hole, W, Srinivasan, M. *Discovering Missed Synonymy in a Large Concept-Oriented Metathesaurus*. *AMIA Annu Symp Proc*. 2003;354-358.
- [14] Mougín F, Dupuch M, Grabar N. *Improving the mapping between MedDRA and SNOMED CT*. In Peleg M, Lavrač N, and Combi C (editors). *Artificial Intelligence in Medicine*. Berlin Heidelberg: Springer. 2011; LNAI 6747:220-224

## Adresse de correspondance

Equipe de Recherche en Informatique Appliquée à la Santé  
ISPED – Université de Bordeaux  
146 Rue Léo Saignat  
33076 Bordeaux  
berenice.brechat@isped.u-bordeaux2.fr