

Formalisation de la terminologie LOINC® et évaluation de ses avantages pour la classification des tests de laboratoire

Mélissa Mary^{1,2}, Lina F. Soualmia^{2,3} et Xavier Gansel¹

¹ bioMérieux SA, Dépt. Développement et Intégration, 38390 La Balme Les Grottes,
{melissa.mary ;xavier.gansel}@biomerieux.com

² Normandie Universités, UNIVROUEN - LITIS EA 4108, 76000 Rouen
Lina.Soualmia@chu-rouen.fr

³ LIMICS INSERM UMR_1142, 75000 Paris,

Résumé : La numérisation croissante des données médicales a fait émerger de nouveaux standards pour représenter l'information au sein des dossiers patients informatisés. Dans le domaine du diagnostic *in vitro*, deux standards sémantiques sont recommandés pour coder les descriptions des tests de laboratoire (la terminologie LOINC®) et les résultats obtenus (l'ontologie SNOMED CT®). La structuration non formelle des tests et l'absence de relations hiérarchiques au sein de LOINC® limitent les capacités d'interrogation et de réutilisation des données du diagnostic *in vitro* au sein des dossiers patients. L'objectif de ce travail est d'évaluer les avantages d'une représentation formelle permettant de décrire et de classer automatiquement les tests de laboratoire. Dans cet article, nous présentons ontoLOINC, une représentation ontologique de LOINC® et les processus de génération que nous avons mis en place. Dans un second temps, nous comparons les classifications des tests dans ontoLOINC et dans un formalisme fondé sur SNOMED CT® (LOINC—SNOMED CT). La formalisation des tests et la classification des éléments utilisés dans leur définition permet d'obtenir des classifications cohérentes avec la terminologie.

Mots-clés : modélisation d'ontologie – peuplement d'ontologie – classification automatique – évaluation, LOINC® – SNOMED CT®

1 Introduction

L'informatisation croissante des données médicales au sein de dossiers médicaux partagés a pour objectif d'améliorer la prise en charge du patient par l'établissement d'échanges d'informations interopérables entre acteurs et systèmes de la chaîne de soins (Macary, 2007; Stroetmann, 2009). Une des clés de l'interopérabilité repose sur l'utilisation des vocabulaires standards pour coder l'information au sein des dossiers patients. Ces vocabulaires spécialisés dans un domaine peuvent être représentés sous forme de terminologie, thesaurus, ou ontologie. Dans le domaine du diagnostic *in vitro*¹ deux ressources sont recommandées par les instances de standardisation nationales et internationales pour coder les informations au sein des comptes-rendus d'analyses (Blumenthal, 2010; Stroetmann, 2009). La terminologie LOINC®² (*Logical Observation Identifiers Names and Codes*) est préconisée pour décrire les tests de laboratoire, et l'ontologie SNOMED CT®³ (*Systematized Nomenclature of MEDicine – Clinical Terms*) pour coder les résultats obtenus. Les informations exprimées par LOINC® et SNOMED CT® dans les dossiers patients sont interdépendantes : l'interprétation d'un résultat varie en fonction du test qui a permis de l'obtenir. LOINC® et SNOMED CT®

¹ Le diagnostic *in vitro* regroupe l'ensemble des analyses biologiques réalisées sur un échantillon clinique (sang, urine) qui permettent de détecter et caractériser une pathologie

² <http://loinc.org/>

³ <http://www.snomed.org/>

doivent donc être utilisées conjointement pour permettre l'interrogation et l'agrégation des données issues d'un compte-rendu d'analyse. L'absence de représentation formelle et l'absence de relations hiérarchiques entre les tests LOINC® limitent les capacités d'interrogation et de réutilisation des données du diagnostic *in vitro* au sein des dossiers patients. Pour répondre à cette problématique, le *Regenstrief Institute* et l'IHTSDO (maintenant appelé SNOMED International) ont mis en place une collaboration en 2013 (IHTSDO & Regenstrief Institute, 2013; Vreeman, 2015) qui a pour vocation de rendre interopérable les données du compte-rendu de laboratoire en proposant une ressource alignant les tests LOINC® avec l'ontologie SNOMED CT® (ressource LOINC—SCT).

L'objectif de ce travail est d'étudier l'impact d'une formalisation logique de la représentation des tests de diagnostic *in vitro*, initialement décrit par la terminologie LOINC. Dans un premier temps, nous avons développé ontoLOINC une ontologie représentant la terminologie LOINC® ; elle pour pallier aux contraintes inhérentes à la ressource LOINC—SCT. Dans une seconde partie nous nous comparons ontoLOINC et LOINC—SCT en étudiant les inférences obtenues à partir des deux modèles. Cette étude permet notamment d'illustrer les bénéfices d'une formalisation pour la classification des LOINC®.

Cet article est organisé comme suit : dans la section 2 nous décrivons les ressources LOINC®, SNOMED CT® et l'alignement LOINC—SCT qui ont été utilisés pour cette étude. La section 3 présente ontoLOINC, son processus de génération et le protocole de classification automatique des tests. La section 4 présente les résultats de classification automatique sur les représentations ontoLOINC et LOINC—SCT. Nous discutons les résultats obtenus en section 5 et concluons ce travail en section 6.

2 Ressources utilisées

2.1 Logical Observation Identifiers Names and Codes

La terminologie LOINC® a été construite en 1994 pour décrire les tests cliniques ou réalisés par le laboratoire d'analyse sur un prélèvement (sang, urine, *etc.*). Elle est mise à jour deux fois par an par le Regenstrief Institute et est couverte par une licence d'utilisation gratuite. La terminologie est accessible en ligne ou via le logiciel RELMA. (McDonald *et al.*, 2003; Sheide & Wilson, 2013).

Dans cette étude nous utilisons la version 2.52 de la terminologie (décembre 2015) ; elle décrit 70 789 tests actifs dans une structure normée par 6 dimensions obligatoires et 3 optionnelles. La description d'un test représente à la fois des informations sur le protocole expérimental (*Composant, Milieu, Technique* et *Temps*) et sur le type de résultat attendu (*Échelle* et *Unité*). Dans cet article nous utilisons le terme générique *partie* pour parler de la valeur d'une dimension utilisée pour décrire un test.

Table 1 Dimensions du modèle LOINC®

Dimension	Description	Exemple
Composant	Objet du test ou caractéristique testée	<i>Bacteria identified</i>
Milieu	Echantillon sur lequel est réalisé le test	<i>Bld (Blood)</i>
Méthode	Technique d'analyse	<i>IF (Immunofluorescence)</i>
Temps	Temporalité du test (cinétique ou ponctuelle)	<i>Pt (ponctuel)</i>
Échelle	Type de résultats	<i>Qn pour quantitatif</i>
Unité	Description précise de l'unité du résultat	<i>MCnc (Mass/volume)</i>
Challenge	Contexte de la réalisation du test	
Diviseur	Composant utilisé dans le cas où le résultat du test correspond un ratio	
Ajustement	Paramètre d'ajustement du résultat de la mesure.	

Un test est également contextualisé par des informations complémentaires comme la *classe* qui spécifie un sous type de test (MICRO pour microbiologie, SERO pour la serologie, *etc.*).

2.2 Systematized Nomenclature Of MEDicine Clinical Terms

SNOMED CT® est une ontologie sous licence créée en 2002 pour décrire l'ensemble du domaine clinique (Bhattacharyya, 2016; Cornet & de Keizer, 2008). Elle est maintenue à jour bi-annuellement par SNOMED International (anciennement appelé IHTSDO) et est diffusée dans un système de fichiers plats nommé RF2. SNOMED International fournit également un script permettant de transformer les données du format RF2 en ontologie OWL EL (Bodenreider *et al.*, 2007; Schulz *et al.*, 2009).

L'ontologie est composée de 350 000 concepts répartis en 19 axes. Les axes permettent d'organiser l'ontologie en sous-domaine de connaissance. Par exemple, l'axe *123037004|Body Structure|* regroupe l'ensemble des concepts relatifs à la description anatomique et cellulaire du corps humain. Les concepts sont organisés en poly-hiérarchies par des relations de subsomption. Au sein de l'ontologie, on distingue les concepts :

- 1 pré-coordonnées : ils sont gérés par SNOMED International et constituent le corps de la ressource ;
- 2 post-coordonnées : ils sont définis pour une utilisation particulière à partir de concepts pré-coordonnées et n'appartiennent pas à la ressource SNOMED CT (SNOMED International, 2017a).

Le concept *Left arm fracture* n'existe pas nativement dans SNOMED CT® ; il peut être créé par post-coordination (Figure 1) à partir des concepts (*Left upper arm structure, Fracture*) et des relations (*Finding site, Associated morphology*) existants dans la ressource.

Class: 'Left Arm Fracture'

EquivalentTo sct:'Disorder of Bone' and

sct:associated Morphology some sct:Fracture and

sct:finding Site some sct:'Left upper arm structure'

Figure 1 Post-coordination du concept *Left arm fracture* en syntaxe Manchester

2.3 Description des tests de laboratoire en SNOMED CT®

Dans cette étude nous utilisons une ressource (LOINC—SCT) qui associe dans un même format LOINC® et SNOMED CT®. Cette dernière est créée par des experts des deux standards et diffusée depuis avril 2016⁴ aux formats RF2 et OWL (IHTSDO & Regenstrief Institute, 2013). Cette ressource se compose de deux éléments. Le premier élément décrit l'alignement entre les parties⁵ LOINC® et concepts SNOMED CT®.

Cet alignement est utilisé par la suite pour définir formellement les tests LOINC® en concepts post-coordonnés SNOMED CT® (second élément de la ressource). Les tests LOINC® sont représentés par des concepts définis ; ils utilisent les patrons de conceptions développés par un groupe de travail de SNOMED International (figure 2, SNOMED International, 2017b).

La troisième version⁴ est une version préliminaire qui couvre 13 756 tests LOINC, soit 20% de la terminologie. La version finale proposera une description de l'ensemble des tests LOINC® utilisés pour représenter les tests de laboratoire et les tests cliniques *Vital Sign*.

⁴<http://loinc.org/news/alpha-phase-3-edition-of-draft-loinc-snomed-ct-mappings-and-expression-associations-now-available.html/>

⁵ valeurs des dimensions d'un test

<p>A Class: <i>IgM:MCnc:Pt:Urine:Qn</i></p> <p><u>Annotation</u> LOINCID 56123-3 LOINCLONGCOMMONNAME "IgM [Mass/volume] in Urine"</p> <p><u>EquivalentTo</u> <i>sct:'Observable Entity'</i> and sct:Component some <i>sct:'Immunoglobulin M'</i> and sct:Property Type some <i>sct:'Mass concentration'</i> and sct:Time aspect some <i>sct:'Single point in time'</i> and sct:Inheres In some <i>sct:Urine</i> and sct:Direct Site some <i>sct:'Spot urine sample'</i> and sct:Scale some <i>sct:Quantitative</i></p>	<p>B Class: <i>IgM:MRat:24H:Urine:Qn</i></p> <p><u>Annotation</u> LOINCID 58764-2 LOINCLONGCOMMONNAME " IgM [Mass/time] in 24 hour Urine "</p> <p><u>EquivalentTo</u> <i>sct:'Observable Entity'</i> and sct:Process Output some <i>sct:'Immunoglobulin M'</i> and sct:Property Type some <i>sct:'Mass Rate'</i> and sct:Process Duration some <i>sct:'24 hours'</i> and sct:Inheres In some <i>sct:Urine</i> and sct:Direct Site some <i>sct:'Spot urine sample'</i> and sct:Scale some <i>sct:Quantitative</i></p>
--	--

Figure 2 Exemple de formalisation d'un test issue de la terminologie LOINC® avec le patron de conception général (A) et le patron de tests processus (B). Les attributs spécifiques au patron processus sont colorés en bleu

3 Méthodes

3.1 ontoLOINC et processus de génération

L'utilisation de la formalisation des tests LOINC® avec SNOMED CT® est limitée. En effet, seul 20% des tests LOINC® sont représentés dans LOINC—SCT. D'autre part, seules les institutions qui possèdent une licence d'utilisation de l'ontologie SNOMED CT® peuvent utiliser cette ressource pour développer de nouvelles applications. C'est pourquoi nous avons créé une nouvelle ressource, ontoLOINC, que nous décrivons dans cette section.

LOINC Test class

Annotation
loinc:longCommonName « test longCommonName »
loinc:shortName « test shortName »
loinc:LOINCCode « LOINC identifier »
rdf:SeeAlso some 'LOINC class'

EquivalentTo
loinc:hasComponent some *loinc:'Component part'* and
loinc:hasProperty some *loinc:'Property part'* and
loinc:hasTime some *loinc:'Time Aspect part'* and
loinc:hasSystem some *loinc:'System part'* and
loinc:hasScale some *loinc:'Scale part'* and
loinc:hasMethod some *loinc:'Method part'*

Figure 3 Patron de formalisation d'un test dans ontoLOINC (LOINC Modèle)

d'ontoLOINC en trois étapes (Figure 4). Tout d'abord, les *parties* et *tests* sont extraits de la base RELMA (version 2.52) à l'aide de requêtes SQL et stockés dans les fichiers PART.TXT et TEST.TXT. L'algorithme de peuplement d'ontoLOINC est implémenté à l'aide d'un programme java en deux étapes. Dans un premier temps, ontoLOINC est peuplé par l'ensemble des concepts *partie* et *classe*. L'URI de ces concepts est construite à partir de l'identifiant stocké dans le champs PART_ID du fichier PART.TXT. Les concepts sont répartis dans les hiérarchies *loinc:'LOINC class'* *loinc:'LOINC part'* grâce aux informations stockée dans le champs TYPE_PART. Les tests sont ensuite formalisés sous forme de

Les concepts sont organisés dans ontoLOINC au sein de trois hiérarchies. *loinc:'LOINC part'* contient l'ensemble des concepts représentant des parties; ils sont classés par dimension. *loinc:'LOINC class'* contient l'ensemble des concepts permettant de contextualiser un test par rapport à son utilisation (classe du test). La hiérarchie *loinc:'LOINC test'* inclut l'ensemble des concepts représentant les tests LOINC. Dans ontoLOINC, nous décrivons un test LOINC® (Figure 3) avec des informations terminologiques (**loinc:longCommonName**, **loinc:shortName**) et une définition formelle. La structuration d'un test LOINC® spécifie la dimension du test à la fois par la relation utilisée (en distinguant une relation par dimension) mais aussi l'espace d'arrivée des relations dans le patron (co-domaine, ou range).

Nous avons développé le processus de génération

concepts définis à partir de des informations décrites dans le fichier TEST.TXT ainsi les concepts *partie* et *classe*.

L'ontologie ainsi créée représente les 70 795 tests actifs de la version 2.52 de la terminologie et 45 004 concepts parties et classes.

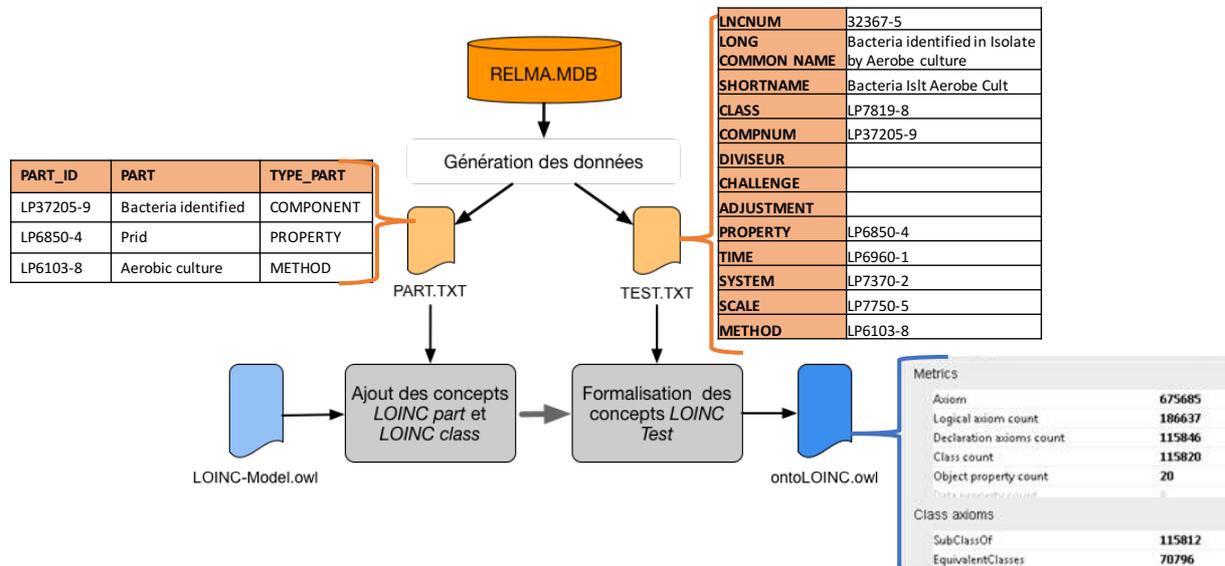


Figure 4 Processus de création d'ontoLOINC

3.2 Inférences sur ontoLOINC et LOINC-SCT

3.2.1 Construction de l'ontologie LOINC—SCT

L'ontologie LOINC—SCT est construite en deux étapes. Dans un premier temps, le système de fichier RF2 de la SNOMED CT (janvier 2016) est transformé en ontologie grâce au script perl distribué par SNOMED International. La ressource LOINC—SCT au format OWL est ensuite fusionnée avec l'ontologie SNOMED CT®.

3.2.2 Protocole d'inférence et analyse des données

L'inférence a été exécutée avec le raisonneur ELK (Kazakov *et al.*, 2012) ; il a été choisi pour sa capacité à raisonner sur des ontologies OWL EL volumineuses. Il nous a permis d'obtenir la liste des tests équivalents (4.1) et la classification automatique des tests.

Durant l'analyse des réorganisations hiérarchiques des tests LOINC, nous pouvons distinguer deux types de reclassifications (Figure 5): la surDéfinition et la sousRestriction (Mary *et al.*, 2017; Mary *et al.*, 2016). La **surDéfinition** spécifie toute classification résultant de l'ajout d'une restriction⁶ dans la définition formelle du test subsumé. Dans l'exemple ci-dessous le test 20789-4 |*Escherichia coli* O157 identified in Isolate by Organism specific culture| (B) est spécifié par une restriction portant sur la dimension *Méthode* non présente chez le test parent 20789-4 |*Escherichia coli* serotype [Identifier] in Isolate (A). La **sousRestriction** représente toute classification résultant d'une spécialisation d'au moins une restriction de la définition du test subsumé. La spécialisation d'une restriction est portée par le concept utilisé ; celui-ci est subsumé de manière directe ou indirecte par le concept utilisé dans la restriction du test parent (Figure 5 C).

⁶ une restriction correspond à une partie de l'axiome définissant un concept test.

L'analyse des classifications est réalisée automatiquement avec un script R. Celui-ci compare entre subsumeur et subsumé la valeur des parties de chaque dimension.

<p>A Class loinc:'<i>Escherichia coli serotype [Identifier] in Isolate</i>'</p> <p><u>EquivalentTo</u> loinc:'<i>LOINC test</i>' and</p> <p>loinc:hasComponent some loinc:'<i>Escherichia coli serotype</i>' and</p> <p>loinc:hasProperty some loinc:'<i>Prid</i>' and</p> <p>loinc:hasTime some loinc:'<i>Pt</i>' and</p> <p>loinc:hasSystem some loinc:'<i>Isolate</i>' and</p> <p>loinc:hasScale some loinc:'<i>Nom</i>'</p>	<p>B Class loinc:'<i>Escherichia coli O157 identified in Isolate by Organism specific culture</i>'</p> <p><u>EquivalentTo</u> loinc:'<i>LOINC test</i>' and</p> <p>loinc:hasComponent some loinc:'<i>Escherichia coli O157 identified</i>' and</p> <p>loinc:hasProperty some loinc:'<i>Prid</i>' and</p> <p>loinc:hasTime some loinc:'<i>Pt</i>' and</p> <p>loinc:hasSystem some loinc:'<i>Isolate</i>' and</p> <p>loinc:hasScale some loinc:'<i>Nom</i>'</p> <p>loinc:hasMethod some loinc:'<i>Organism specific culture</i>'</p>
<p>C $\frac{\text{Escherichia coli O157} \sqsubset \text{Escherichia coli serotype}}{\text{hasComponent 'Escherichia coli O157'} \sqsubset \text{hasComponent 'Escherichia coli serotype'}}$</p>	

Figure 5 Exemple de reclassification par surDéfinition et sousRestriction de tests formalisés avec le ontoLOINC. **A.** Définition du test subsumeur **B.** Définition du test subsumé **C.** Explication de l'inférence par sousRestriction

4 Résultats

4.1 Équivalence de tests

4.1.1 LOINC—SCT

La classification automatique de LOINC—SCT a mis en évidence l'équivalence de 45 tests deux à deux. Après une analyse de chacune des équivalences inférées nous observons que 70% s'expliquent par une différence de granularité entre les parties *Composant* et le concept SNOMED CT® utilisé pour les décrire. Ces résultats ont été présentés aux experts internationaux du domaine impliqués dans la création de cette ressource. Leur analyse met en évidence deux causes racines. Dans 44% des cas, l'équivalence entre tests met en avant un problème de la construction de la ressource LOINC—SCT résultant :

- de mauvais alignements entre partie LOINC® et concept SNOMED CT® (14 cas) ;
- de l'utilisation d'un concept défini pour représenter un test dont au moins une partie est alignée approximativement sur un concept SNOMED CT (6 cas).

Les autres équivalences mettent en exergue des problèmes terminologiques de LOINC. Les experts ont identifié 20 tests équivalents résultant d'une sémantique redondante et 6 tests équivalents causés par une description « non consistante ou floue » dans LOINC.

4.1.2 ontoLOINC

La classification automatique d'ontoLOINC a mis en évidence 3 702 relations d'équivalence (**owl:equivalentClass**) entre 4 516 tests distincts (6,3% des tests totaux). Nous pouvons observer deux causes d'équivalence. Tout d'abord 68% des relations d'équivalence résultent de l'absence de restriction portant sur la dimension *Composant* (2 518) pour les deux tests. Ces équivalences s'expliquent par l'absence de partie *Composant* dans le fichier TEST.TXT pour 2 626 tests. Dans 25% des cas (955), l'équivalence entre deux tests s'explique par la présence d'une information complémentaire à la dimension *Système* (appelée *Super Système*) qui n'est pas modélisée dans ontoLOINC.

Enfin 228 relations d'équivalence ne s'expliquent ni par l'absence de partie *Composant* ni par l'absence de modélisation du *Super Système*, dont 138 correspondent à des tests de la

classe MICRO. L'analyse de ces 138 relations a mis en évidence une cause d'équivalence. La distinction des tests équivalents est portée par la quatrième sous partie du champs *Composant* (McDonald *et al.*, 2017) ; celui-ci spécifie un numéro d'identification unique pour le test. Les tests *630-4 | Bacteria identified in Urine by Culture|* et *17972-1 | Bacteria # 4 identified in Urine by Culture|* représentant un test de culture bactérienne sur un échantillon d'urine. Le test *17972-1* spécifie un numéro d'identification (information portée par #4). Ce sous composant est un artefact utilisé pour la structuration d'un compte rendu avec la syntaxe HL7 et n'apporte aucune information sémantique sur la nature du test.

Pour conclure, parmi les 3 702 relations d'équivalence, 93% montrent les limites de la construction de l'ontologie ontoLOINC que ce soit dans le processus d'extraction des données ou dans le modèle lui-même. Bien que nécessitant un certain nombre de corrections, la première version d'ontoLOINC propose cependant une formalisation correcte de plus de 90% des tests de la terminologie LOINC.

4.2 Réorganisation hiérarchique des tests

Dans cette section nous présentons les résultats de classification automatique des tests. Nous décrivons ces classification d'un point de vue structurel (table 2) et sémantique (figure 6). L'étude structurelle décrit la topologie des classifications. L'étude sémantique permet d'expliquer la classification en fonction des dimensions (éléments sémantiques, table 1) utilisées pour formaliser les tests. La figure 6 mesure la proportion d'inférences expliquées par chacune des 9 dimensions initiales de la terminologie.

4.2.1 LOINC—SCT

Table 2 Métriques des classifications pour ontoLOINC et LOINC—SCT

Information générales	ontoLOINC	LOINC—SCT
Nombre de tests décrits	70 795	13 756
Nombre de relations taxonomiques (non redondante)	25 859	7 961
Test subsumés (unique)	25 082 (35,4%)	6 789 (49,3%)
Topologie des arbres taxonomiques		
Profondeur médiane de reclassification	1 (81,2%)	1 (61,8%)
Profondeur maximale de reclassification	3 (0,7%)	4 (1,2%)
Nombre de tests poly hiérarchisés	777 (3,1%)	1 127 (16,7%)
Raison de classifications		
surDéfinition	25 838	2 580
sousRestriction	0	4 988
surDéfinition + sousRestriction	0	219
Non Expliqué	21	174

Le raisonneur ELK infère 7 961 relations hiérarchiques et permet de classer 49,3% des tests (6 789) représentés dans LOINC—SCT (Table 2). L'étude des raisons de classification montre d'une part que plus de 62% des classifications sont induites par sousRestriction. La figure 6 illustre l'impact de chaque dimension pour les classifications par sousRestriction de LOINC—SCT (violet pointillé). Nous observons que la classification des concepts utilisés pour décrire les parties *Composant* expliquent plus de 92% des classifications. De plus, la classification des concepts des parties *Systèmes*, *Méthode* et *Propriété* explique dans une moindre mesure (total <10%) les inférences dues à la sousRestriction. Nous observons également que 32% des classifications sont inférées par l'ajout d'une restriction dans la définition du test fils (surDéfinition). L'analyse des causes de surDéfinition (figure 6 tirets) montre que plus de 90% des surDéfinitions représentent l'ajout d'une restriction portant sur la

Méthode du test, les 10% restants sont dues à l'ajout d'une restriction représentant une dimension optionnelle (*Challenge*).

4.2.2 ontoLOINC

Le raisonneur ELK infère plus de 300 000 relations hiérarchiques dont 90% sont redondantes. Cette redondance s'explique par les 302 relations d'équivalence ; par la suite l'étude de la classification se fait sur les 25 859 relations non redondantes (Table 2). Tout d'abord, nous observons 39% des tests d'ontoLOINC ont été reclassés et uniquement par ajout d'une sous restriction (*surDéfinition*). Ce résultat s'explique par l'absence de classification hiérarchique des parties dans ontoLOINC.

Parmi les classifications inférées, nous observons que 31,5% résultent de l'absence de parties *Composant* pour 2 626 tests. L'absence de classification par sousRestriction s'explique par l'absence d'organisation hiérarchique des parties (entre elles) dans ontoLOINC. La Figure 6 montre également que les classifications résultent principalement des dimensions *Méthode* (55,8%) et *Challenge* (32,5%). Enfin 1 959 reclassifications de tests sont dues à une *surDéfinition* de la dimension *System* (7,6%).

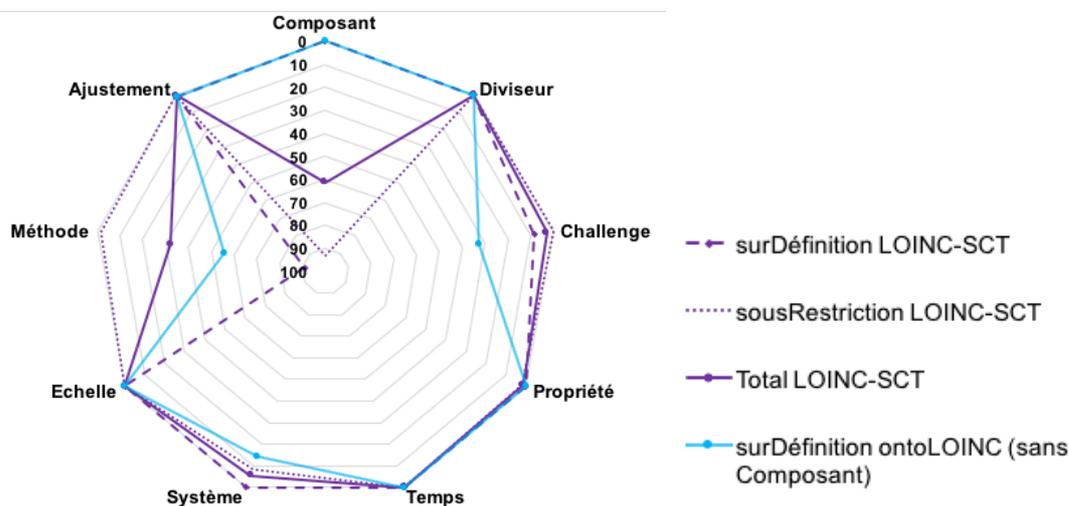


Figure 6 Impact des dimensions pour la reclassification par type de reclassification pour ontoLOINC (bleu) et LOINC—SCT (violet).

5 Discussion : des bénéfices d'une représentation ontologique

Dans cette section, nous présentons, dans un premier temps, l'impact d'une formalisation ontologique de LOINC® pour visualiser la terminologie. Dans une seconde section nous comparons les formalismes ontoLOINC et LOINC-SCT.

5.1 Classification des tests et visualisation

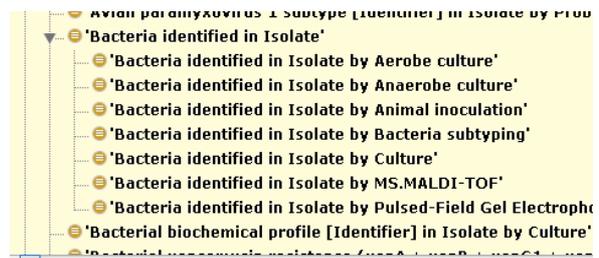
L'organisation des tests au sein d'une hiérarchie propose une visualisation contextuelle des tests les uns par rapport aux autres. La figure 7 montre les différences d'organisation hiérarchique autour du test 42803-7|*Bacteria identified in Isolate*]. La terminologie LOINC® propose une visualisation des tests (hiérarchie Multi-Axiale - figure 7 A) sous forme de catégories construites en fonction des dimensions *Composant* et *Système* ; elle ne propose pas de classification des tests entre eux. ontoLOINC (figure 7 B) propose un premier niveau d'organisation des tests les uns par rapport aux autres. Dans l'exemple, on observe que la classification ontoLOINC affine la visualisation hiérarchique des tests en tenant compte des

parties *Méthode* (*Culture, Aerobic culture, etc.*). La classification obtenue avec LOINC-SCT (figure 7 C) illustre l'impact des sousRestrictions dans l'organisation hiérarchique des tests. On observe d'une part une organisation hiérarchique à trois niveaux, contre deux dans ontoLOINC, mais surtout l'apparition de nouveaux tests dans la hiérarchie. Ces nouveaux tests correspondent à des tests d'identification d'un sous ensemble de bactéries spécifiques, tel que les Streptocoques ou *Escherichia coli*. Cet exemple met en avant l'intérêt de la classification hiérarchique des parties pour enrichir la classification des tests.

A. Hiérarchie Multi-Axiale

B. ontoLOINC

Code	Category or ShortName
LP31755-9	Microbiology
LP14559-6	Microorganism
LP98185-9	Bacteria
LP37205-9	Bacteria identified
LP46585-3	Bacteria Identified Isolate
42803-7	Bacteria identified in Isolate
32367-5	Bacteria identified in Isolate by Aerobe culture
20878-5	Bacteria identified in Isolate by Anaerobe culture
20877-7	Bacteria identified in Isolate by Animal inoculation
612-2	Bacteria identified in Isolate by Bacteria subtyping
43409-2	Bacteria identified in Isolate by Culture
75756-7	Bacteria identified in Isolate by MS.MALDI-TOF
42661-9	Bacteria identified in Isolate by Pulsed-Field Gel...



C. LOINC-SCT

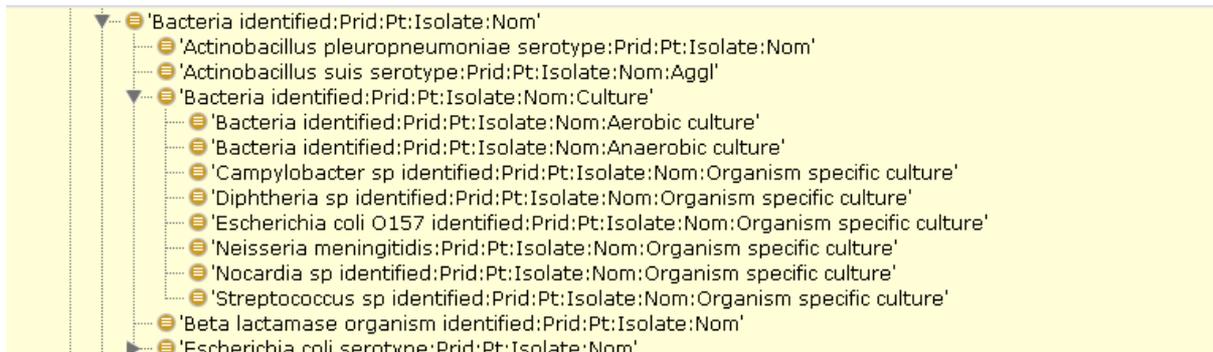


Figure 7 Organisation hiérarchiques autour du test 42803-7|*Bacteria identified in Isolate*| proposé par la hiérarchie Multi-Axiale (A), et obtenus par classification automatique d'ontoLOINC (B) et LOINC-SCT (C).

5.2 Comparaison de l'organisation d'ontoLOINC avec LOINC—SCT

Si ontoLOINC et LOINC—SCT proposent une formalisation de la terminologie LOINC, leurs différences impacte leur utilisabilité.

LOINC—SCT est développée par une collaboration d'experts du *Regenstrief Institute* et de SNOMED International ce qui en fait une ressource de meilleure qualité qu'ontoLOINC. L'étude des équivalences a permis de mettre en évidence certains défaut de conceptions dans la terminologie. Contrairement à ontoLOINC elle propose un panel de classifications plus exhaustif grâce à la représentation hiérarchique des concepts utilisés pour formaliser les tests. Son utilisabilité est cependant limitée par trois inconvénients. D'une part son utilisation est restreinte par la licence d'exploitation de la SNOMED CT, ce qui complexifie le déploiement d'applications basées sur LOINC—SCT. D'autre part la version LOINC—SCT étudiée représente partiellement (20%) des tests de la terminologie LOINC. Si celle-ci est amenée à être incrémentée dans les prochaines version, seul les tests de laboratoires et les tests cliniques

Vital Sign ont vocation à être représentés avec SNOMED CT. Dans une étude précédente (Mary *et al.*, 2016), nous avons montré que l'utilisation de la ressource LOINC—SCT requiert une compréhension fine de l'ontologie SNOMED CT pour exprimer des requêtes sur les tests.

Table 3 Comparaison des formalismes logiques de LOINC.

<i>Caractéristiques Générales</i>	ontoLOINC (version 1)	LOINC—SCT (alpha release version 3)
Libre de droit	Oui	Non
Version de la terminologie	Au choix (ici 2.52)	2.52
Couverture de la terminologie	Au choix (Test actif)	Partielle (20%)
Taille de l'ontologie	115 832 concepts	331 515 concepts
Créateur	Utilisatrice avertie de LOINC	Experts de LOINC® et SNOMED CT
Comparaison par rapport au modèle de la terminologie		
Isomorphisme	Oui	Non
Formalisation d'un test	1 patron explicitement décrit (modèle LOINC)	2 patrons décrits textuellement dans la documentation
Impact sur le développement et l'utilisation de la terminologie		
Identification de tests redondant	Non (trop de bruit)	Oui
Classification des parties	Pas dans la 1 ^{ère} version	Oui

ontoLOINC, décrite dans ce papier, est une première version d'ontologie (mars 2017) qui nécessite encore quelques corrections. Nous sommes aujourd'hui en contact avec les membres du Regenstrief Institute pour identifier les meilleurs protocoles d'extraction des données de la dimension *Composant*. Le processus de construction de cette ontologie offre la possibilité de personnaliser de l'ontologie LOINC®. L'extraction des données permet de sélectionner les tests de n'importe quelle version de la terminologie, mais également sur d'autres critères comme la *classe* du test par exemple. Cependant l'utilisation de RELMA pour extraire les éléments nécessaires à la construction d'ontoLOINC nécessite une autorisation préalable du Regenstrief Institute (cf. Licence LOINC®).

Le patron de formalisation d'un test est strictement identique à la structure d'un test dans la terminologie LOINC® ; ce qui facilite l'assimilation de cette nouvelle ressource par des experts biologistes. Néanmoins l'absence d'organisation hiérarchique des concepts partie limite la reclassification des tests à des surDéfinitions.

6 Conclusion

Dans cet article nous présentons tout d'abord ontoLOINC, une ontologie permettant de représenter la terminologie LOINC® ainsi que son processus de génération. Nous avons ensuite étudié les bénéfices des formalismes logiques pour la classification des tests de laboratoire. L'étude de la classification de la ressource LOINC—SCT⁴ (IHTSDO & Regenstrief Institute, 2013) a permis de démontrer les bénéfices d'une organisation hiérarchique des parties pour la classification automatique des tests et leur visualisation.

Nous travaillons actuellement à une seconde version d'ontoLOINC pour améliorer la formalisation de la terminologie LOINC® en ajoutant la dimension *Super Système* et corriger l'extraction des parties représentant la dimension *Composant*. Une fois corrigé et avec l'aval du *Regenstrief Intitute*, nous mettrons à disposition ontoLOINC sur les répertoires d'ontologies publiques tel que BioPortal⁷.

Nous développons également un module d'intégration d'ontoLOINC avec d'autres ontologies ; celui-ci utilise des alignements simples pour affiner la classification des tests. Ce module d'intégration sera utilisé dans un premier temps pour compléter la comparaison

⁷ <https://bioportal.bioontology.org/>

d'ontoLOINC et LOINC—SCT. Nous réutiliserons les alignements entre parties et concepts décrit dans la ressource LOINC—SCT pour générer une nouvelle ontologie LOINC® intégrée à SNOMED CT®. Celle-ci nous permettra de valider les performances du modèle ontoLOINC pour classer les tests par sousRestriction.

À moyen terme, nous souhaitons étudier les classifications de tests obtenus sur ontoLOINC intégrée avec d'autres ontologies reconnues comme standard en biologie :

- CheBI comme taxonomie des composés chimiques ;
- NCBI Taxonomie pour les tests d'identification ;

Pour conclure, cet article présente deux représentations formelles de la terminologie LOINC®. Ce mode de représentation offre de nouvelles opportunités pour manipuler les données de diagnostic *in vitro*. L'approche de représentation formelle de la terminologie LOINC® a déjà été exploitée pour construire de nouvelles applications, notamment pour sélectionner les données de diagnostic à fort intérêt épidémiologique (Eilbeck, *et al.*, 2013). La formalisation de LOINC® nous permet d'envisager des systèmes d'interprétation des résultats d'analyses de diagnostic *in vitro* comme par exemple l'interprétation des résultats de microbiologie (Bright, *et al.*, 2012).

Remerciement

Nous remercions l'ensemble du groupe de travail des *Observable Entity* de SNOMED International pour leur expertise concernant la ressource LOINC—SCT et plus particulièrement Suzanne Santamaria, Farzaneh Ashrafi, Swampna Abhyankar et Daniel Karlsson.

Référence

- BHATTACHARYYA, S. B. (2016). Using SNOMED CT. Dans S. Bhattacharyya (dir.), *Introduction to SNOMED CT* (p. 157-182). Singapore : Springer Singapore. doi:10.1007/978-981-287-895-3_9
- BLUMENTHAL, D. (2010). Launching HITECH. *New England Journal of Medicine*, 362(5), 382-385. doi:10.1056/NEJMp0912825
- BODENREIDER, O., SMITH, B., KUMAR, A., & BURGUN, A. (2007). Investigating subsumption in SNOMED CT: An exploration into large description logic-based biomedical terminologies. *Artificial intelligence in medicine*, 39(3), 183-195.
- BRIGHT, T. J., FURUYA, E. Y., KUPERMAN, G. J., CIMINO, J. J., & BAKKEN, S. (2012). Development and Evaluation of an Ontology for Guiding Appropriate Antibiotic Prescribing. *Journal of Biomedical Informatics*, 45(1), 120-128. doi:10.1016/j.jbi.2011.10.001
- CORNET, R., & DE KEIZER, N. (2008). Forty years of SNOMED: a literature review. *BMC Medical Informatics and Decision Making*, 8(Suppl 1), S2. doi:10.1186/1472-6947-8-S1-S2
- EILBECK, K., JACOBS, J., MCGARVEY, S., VINION, C., & STAES, C. J. (2013). *Exploring the use of ontologies and automated reasoning to manage selection of reportable condition lab tests from LOINC*. Communication présentée au ICBO (p. 12-15).
- IHTSDO, & REGENSTRIEF INSTITUTE. (2013). Regenstrief and the IHTSDO are working together to link LOINC and SNOMED CT. Repéré à <https://loinc.org/collaboration/ihtsdo>
- KAZAKOV, Y., KRÖTZSCH, M., & SIMANCIK, F.. (2012). *ELK Reasoner: Architecture and Evaluation*., Communication présentée au OWL Reasoner Evaluation Workshop (p. 12).
- MACARY, F. (2007). IHDE, CDA et LOINC : des composants d'interopérabilité au service du partage des résultats de biologie médicale. *Spectra biologie*, 26(158), 51-57.
- MARY, M., SOUALMIA, L. F., & GANSEL, X. (2016). *Projection des propriétés d'une ontologie pour la classification d'une ressource terminologique*. Communication présentée au Journée Francophones sur les Ontologies (p. 12), Bordeaux, France.
- MARY, M., SOUALMIA, L. F., GANSEL, X., DARMONI, S. J., KARLSSON, D., & SCHULZ, S. (2017). *Ontological Representation of Laboratory Test Observables: Challenges and Perspectives in the SNOMED CT Observable Entity Model Adoption*. Communication présentée au AIME 2017 Proceedings (p. 10), Communication présentée au 16th Conference on Artificial Intelligence in Medicine, Vienna, Austria.

- MCDONALD, C. J., HUFF, S., DECKARD, J., ARMSON, S., ABHYANKAR, S., & VREEMAN, D. J. (2017). *Loinc User Guide*. Indianapolis,USA : Regenstrief Institute.
- MCDONALD, C. J., HUFF, S. M., SUICO, J. G., HILL, G., LEAVELLE, D., ALLER, R., FORREY, A., MERCER, KA., DEMOOR, G., HOOK, J., CASE, J. & MALONEY, P. (2003). LOINC, a universal standard for identifying laboratory observations: a 5-year update. *Clinical chemistry*, 49(4), 624-633.
- SCHULZ, S., SUNTISRIVARAPORN, B., BAADER, F., & BOEKER, M. (2009). SNOMED reaching its adolescence: Ontologists' and logicians' health check. *International Journal of Medical Informatics*, 78, Supplement 1, S86-S94. doi:10.1016/j.ijmedinf.2008.06.004
- SHEIDE, A., & WILSON, P. S. (2013). Reading up on LOINC. *Journal of AHIMA/American Health Information Management Association*, 84(4), 58-60.
- SNOMED INTERNATIONAL. (2017a). SNOMED CT Document Library. Repéré à <https://confluence.ihtsdotools.org/display/DOC/SNOMED+CT+Document+Library>
- SNOMED INTERNATIONAL. (2017b). Observable and Investigation Model Project Home. Repéré à <https://confluence.ihtsdotools.org/display/OBSERVABLE/Observable+and+Investigation+Model+Project+Home>
- STROETMAN, V., D. KALRA, P. LEWALLE, A. RECTOR, J. RODRIGUES, K. STROETMAN, G. SURJAN, B. USTUN, M. VIRTANEN, & P. ZANSTRA. (2009). *Semantic Interoperability for Better Health and Safer Healthcare* (Report Publication no KK-80-09-453-EN-C)(p1-34). European Commission.
- VREEMAN, D. (2015). Guidelines for using LOINC and SNOMED CT Together. *Daniel Vreeman*. Repéré à <https://danielvreeman.com/guidelines-for-using-loinc-and-snomed-ct-together-without-overlap/>