

# International Healthcare Terminologies for Morbidity New Era: SNOMED and ICD11

Jean Marie Rodrigues<sup>1</sup>, Constant Kone<sup>2</sup>, Michel Babri<sup>2</sup>, Béatrice Trombert<sup>1</sup>

<sup>1</sup>Université Jean Monnet, Saint Etienne, France

<sup>2</sup>Felix Houphouët Boigny National Polytechnic Institute, Yamoussoukro, Cote d'Ivoire

Email: [rodriguesjeanmarie2@gmail.com](mailto:rodriguesjeanmarie2@gmail.com)

**How to cite this paper:** Rodrigues, J.M., Kone, C., Babri, M. and Trombert, B. (2024) International Healthcare Terminologies for Morbidity New Era: SNOMED and ICD11. *Journal of Biosciences and Medicines*, 12, 357-368.  
<https://doi.org/10.4236/jbm.2024.128028>

**Received:** June 12, 2024

**Accepted:** August 24, 2024

**Published:** August 27, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

The Information Technology (IT) developments have changed the use of Healthcare terminologies from paper-based mortality statistics with the WHO international classifications of diseases (ICD) to the IT-based morbidity implementations for instance for Casemix-based healthcare funding and managing systems. This higher level of granularity is worldwide spread under the umbrella of several national modifications named ICD10 XM. These developments have met the increased use of the International Clinical Reference Terminology named SNOMED. When the updating of WHO ICD10 to WHO ICD11 was decided a merging was envisaged and a WHO SNOMED CT common work proposed a methodology to create a common formal ontology between the 11th version of the WHO International Classification of Diseases and Health Problems (ICD) and the most used in the world clinical terminology named Systematized Nomenclature of Human and Veterinary Medicine Clinical Terms (SCT). The present work follows this unachieved work and aims to develop a SNOMED-based formal ontology for ICD11 chapter 1 using the textual definitions of ICD11 codes which is a completely new character of ICD and the ontology tools provided by SCT in the publicly available SNOMED Browser. There are two key results: the lexical alignment is complete and the ontology alignment is incomplete with the validated SNOMED concept model can be completed with not yet validated attributes and values of the SNOMED Compositional Grammar. The work opens a new era for the seamless use of both international terminologies for morbidity for instance for DRG/Casemix and clinical management use. The main limitation is that it is restricted to 1 out of 26 chapters of ICD11.

## Keywords

Lexical Alignment, Semantic Interoperability, Ontology, SNOMED Concept Model, ICD-11, SNOMED CT

## 1. Introduction

In recent years, information systems have achieved the ability to communicate, exchange data, this is known as functional interoperability. This has been made possible by the use in more developed countries hospitals and healthcare systems of several technologies, including application programming interfaces (APIs) and standardized communications norms and protocols such as TCP/IP, HTTP, FTP, etc.

Semantic interoperability between terminologies, defined as the ability to match the meanings of similar or equivalent terms between different terminologies, to enable clear and precise communication and understanding of information exchanged between IT systems, is still a challenge due to the lack of semantic standards [1] [2].

Each terminology has been designed to meet a particular need, and a particular use (structuring Clinical Terminology of the medical record, coding information, indexing documents, representing entities in expert systems, input for management and funding. When they claim to represent an invariant concept, it is often untrue whenever information entered for one purpose is reused for another [3].

There are several lexical-based alignments between the WHO International Classification of Diseases and Health Problems (ICD-10) and the most used in the world clinical terminology named Systematized Nomenclature of Human and Veterinary Medicine—Clinical Terms (SNOMED CT). The most well-known are from SNOMED CT to ICD-10 [4] [5]. In 2010, SNOMED International, formerly known as the International Health Terminology Standards Development Organization (IHTSDO<sup>®</sup>), signed an institutional agreement with the World Health Organization (WHO). This collaboration aimed to establish a priority set of cross-matches from SNOMED CT to the ICD-10 using a heuristic method to meet the epidemiological, statistical, and administrative needs of SNOMED International members, WHO collaborating centers, and other interested parties. These alignments are available in the SNOMED CT browser under the “refset” tab for certain SNOMED CT concepts [6].

Several methods exist for implementing semantic interoperability. We can group them into lexical-based. and semantic-based alignment methods. Among these methods, those based on semantics can be distinguished between those that model without using formal logic [7] [8] and more recent methods that use the representation of logical definitions, through the use of standardized formal logic, particularly in the semantic web; this is known as applied ontology [9].

Under the umbrella of the WHO SNOMED/IHTSDO 2010 agreement a common work WHO IHTSDO has proposed the development of a formal common ontology ICD-11 SCT [10] [11].

Following these recent developments, we propose an approach for building a formal ontological representation of ICD-11 [12] using first a lexical mapping with UMLS MetaMap between ICD-11 and SNOMED CT codes labels [13] and in a second time a SNOMED formal ontology representation with the SNOMED

concept model [14].

## 2. Methods

The ICD, now in its 11th version, came into effect in January 2022 for mortality statistics and contains 17,000 unique codes and 120,000 codable terms that can be consulted in the ICD-11 browser [15]. It consists of 26 chapters and two additional sections: 1) an additional section for the assessment of functioning and 2) a section for extension codes.

The WHO provides a table (Excel file) with correspondence between ICD-11 codes and titles and those of ICD-10 (Version of 01/2024) which is still predominantly used in health information systems, whereas ICD-11, the subject of our work, has not yet been adopted by all countries for morbidity [16]. Its implementation will follow a timetable specific to each country.

Our work is on ICD-11 and the method we propose cannot be applied to ICD-10.

We have studied only Chapter 01 (Certain infectious or parasitic diseases). This chapter includes certain conditions caused by pathogenic organisms or micro-organisms such as bacteria, viruses, parasites, or fungi.

It excludes:

- Infections of the fetus or newborn child are classified in Chapter 19;
- Human prion diseases are classified in Chapter 8;
- CA40 Pneumonia is classified in Chapter 12;
- Infections resulting from a device, implant, or graft, not elsewhere classified (NE83.1) classified in Chapter 22.

It is subdivided into 22 sub-chapters and a title code for infection without precision (IH0Z).

From the Excel file described above, we extracted 1027 codes and titles corresponding to chapter 01 of the ICD-11. We then filtered it by eliminating titles containing the words “other” and “unspecified”, *i.e.*, codes ending in “Y” and “Z”. We finally obtained 651 lines of codes and titles, of which 474 have detailed textual descriptions and 177 do not and have only the label description.

SCT is a medical terminology for healthcare containing 350,000 concepts organized into 19 hierarchies [5]. It is both a reference terminology based on a poly-hierarchical structure of the type and, for certain hierarchies, a formal ontology representation. This formal representation is implemented using a Concept Model [17] and Compositional Grammar rules [18] based on the Ontology Web Language (OWL) of the Semantic Web [19].

The SCT concept model specifies how to define SCT expressions through semantic categories, definition attributes, and expression constraints according to rules validated in the Machine Readable Concept Model (MRCM) [20].

Compositional Grammar is a lightweight syntax based on the OWL2 EL logic profile [21]. This syntax allows more representations expressions than the ones validated today in the SCT browser in a way that is both human-readable and machine-readable.

The proposed method consists of two stages:

- Lexical stage: Alignment of ICD-11 titles and Fully Specified Name (FSN) or synonyms in SCT.
- Ontological stage: Validate the formal ontological representation of the ICD-11 Chapter 1 textual descriptions or, if absent, of the ICD-11 codes titles by the formal ontology of SCT (Concept Model and Compositional Grammar).

## 2.1. Lexical Stage

This step was to map the ICD-11 titles to the Fully Specified Name (FSN) or synonyms of SCT.

Two tools were used in this lexical step, MetaMap [13] for automatic alignment and the SNOMED browser [6] for manual validation.

MetaMap implements processing based on linguistic and algorithmic principles.

It maps any terminology entered as input to any terminology included in the UMLS Metathesaurus. SCT is one of the terminologies included in the Metathesaurus.

- MetaMap proposes an option to activate the processing of conjunctions during the analysis of the input terminology; in our work, ICD-11 codes titles or labels.
- MetaMap provides two types of alignment results. Either an alignment with an expression corresponding to a single concept, known as a pre-coordinated concept, or an alignment with an expression corresponding to several concepts, known as post-coordinated concepts.

After this automatic alignment phase using MetaMap, all post-coordinated alignments were manually validated in the SCT browser and showed more pre-coordinated alignment.

## 2.2. Ontology Stage

It is defined as the analysis of the representation of ICD-11 textual definitions or descriptions by the ontology representations of the SCT concepts aligned lexically in the previous step.

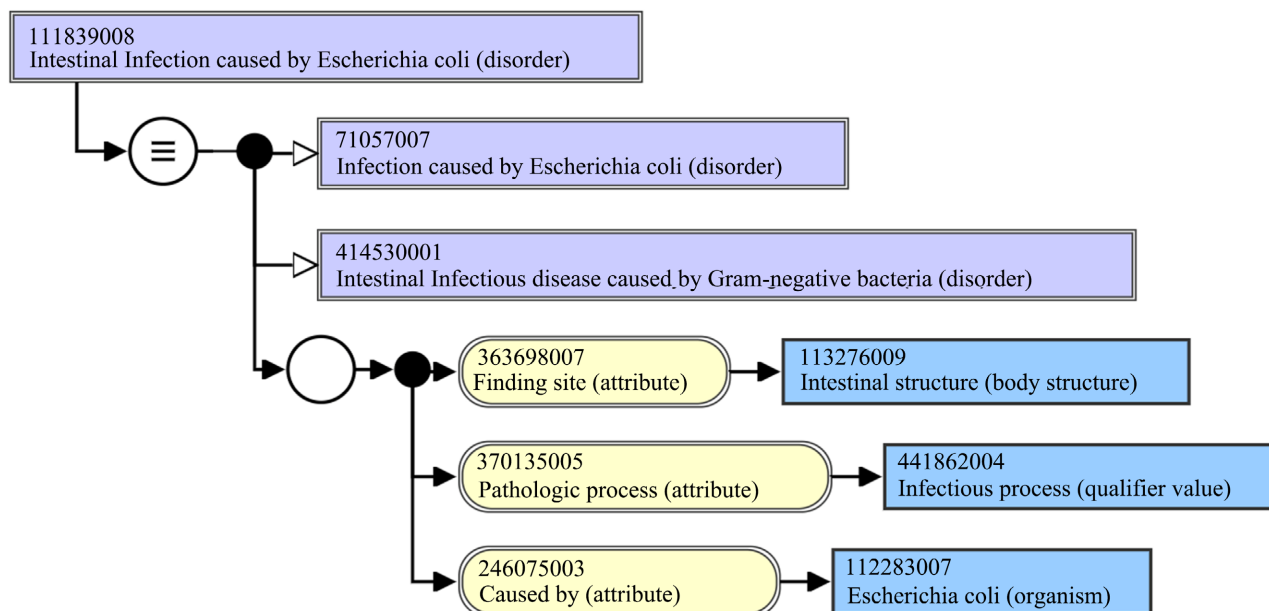
It is important to say that these textual definitions or descriptions are an important try to consider not only the lexical expression of the code labels but also their meaning. This was done by international experts nominated by WHO and based on consensus between them. This explains that it has not been done for all the Chapter 1 codes.

The analysis method compares 2 different artifacts:

On one hand, a text in English describes the health problem by its main characteristics: clinical signs, anatomy localization, transmission and etiology, proof by biological tests identifying the cause which in Chapter 1 is mainly an infectious agent: bacteria, virus, parasite, etc.

On the other hand, a SNOMED concept model representation [14] in EL lan-

guage and by colored diagram available on SCT browser as for example “111839008|Intestinal infection caused by *Escherichia coli* (disorder)” **Figure 1**.



**Figure 1.** SNOMED CT definition by the SNOMED CT concept axiom diagram “111839008|Intestinal infection caused by *Escherichia coli* (disorder)”.

To be fully defined in ontology the representation of a SNOMED concept shall contain all the characteristics necessary and sufficient to represent the code.

This decision is taken by the SNOMED editorial board.

In the case of SCT concepts, there are no textual definitions so if all the characteristics edited in the coded specified name or synonyms are present it is fully defined (necessary and sufficient).

The difficulty of our work is that the ICD-11 textual definition is more detailed than a name and we have had to decide what was necessary and what was sufficient.

For the necessary condition, we have considered that if all the characteristics edited in the ICD-11 code label are present in the conceptual model of the SCT codes lexically aligned with the ICD-11 code the necessary condition was fulfilled

For the sufficient condition, we have decided to consider that the criteria were to assess if these ICD-11 textual definitions characteristics were ALWAYS present.

If Yes these characteristics shall be considered in an ontology representation.











If not for example the characteristics following words as often, sometimes, is possible, MAY BE present, they are not to be considered in an ontology representation.

If the 2 conditions are fulfilled, we consider that the representation of ICD-11 textual definitions or descriptions by the ontological representations of the SCT

concepts aligned lexically in the previous step is COMPLETE

If only the necessary condition is fulfilled, we consider that the representation of ICD-11 textual definitions or descriptions by the ontological representations of the SCT concepts aligned lexically in the previous step is INCOMPLETE.

Firstly, we analyze the ICD-11 textual definitions and colored the words indicating a characteristic always present **Figure 2**.

Attributes	Color codes
363698007 Finding site (attribute) →	
116676008 Associate morphology (attribute) →	
308489006 Pathological process (qualifier value) →	
246075003 Causative agent (attribute) →	
363705008 Has definitional manifestation (attribute) →	
363713009 Has interpretation (attribute) →	
363714003 Interprets (attribute) →	
42752001 Due to (attribute) →	
255234002 After (attribute) →	
263502005 Clinical course (attribute) →	

**Figure 2.** SNOMED CT attributes and textual descriptions color codes.

Secondly, we compare these ICD-11 colored textual definitions words to the concept model of the SNOMED concept lexically aligned in the previous step to check if all the colored characteristics are represented in the SNOMED concept model.

Thirdly, we propose to complete the ontological representation of these ICD-11 codes with the attributes and values needed to represent them according to the SNOMED information model and the SNOMED compositional grammar authorized or not.

These three steps are described in detail in the following.

### 2.2.1. Identification of Characteristics Always Present in ICD-11 Textual Descriptions

We have decided to color the characteristics words in accordance with SNOMED CT attributes as such.

### 2.2.2. Comparing ICD-11 Colored Textual Definitions to the Concept Model of the SNOMED Concept Lexically Aligned in the Previous Step

The example is an incomplete ontology representation of an ICD-11 code and textual definition by the concept model of a SCT concept lexically aligned in the previous stage.

The ICD-11 code and title “1E50.4 Acute hepatitis E” has the textual definition “Liver disease caused by acute infection with the hepatitis E virus. This disease is characterized by **nausea**. **Transmission** is generally via the **fecal-oral route**. Confirmation is by detection of anti-hepatitis E virus IgM antibodies in an individual’s serum” is lexically aligned to the concept “SCT 235867002|Acute hepatitis E (disorder)|”.

Some ICD-11 textual definition characteristics are present in the fully defined concept model of the SCT concept lexically aligned: liver, acute, infection and hepatitis E virus.

Other are missing: Nausea, Transmission Confirmation is by detection of anti-hepatitis E virus IgM antibodies as shown in **Figure 3**.

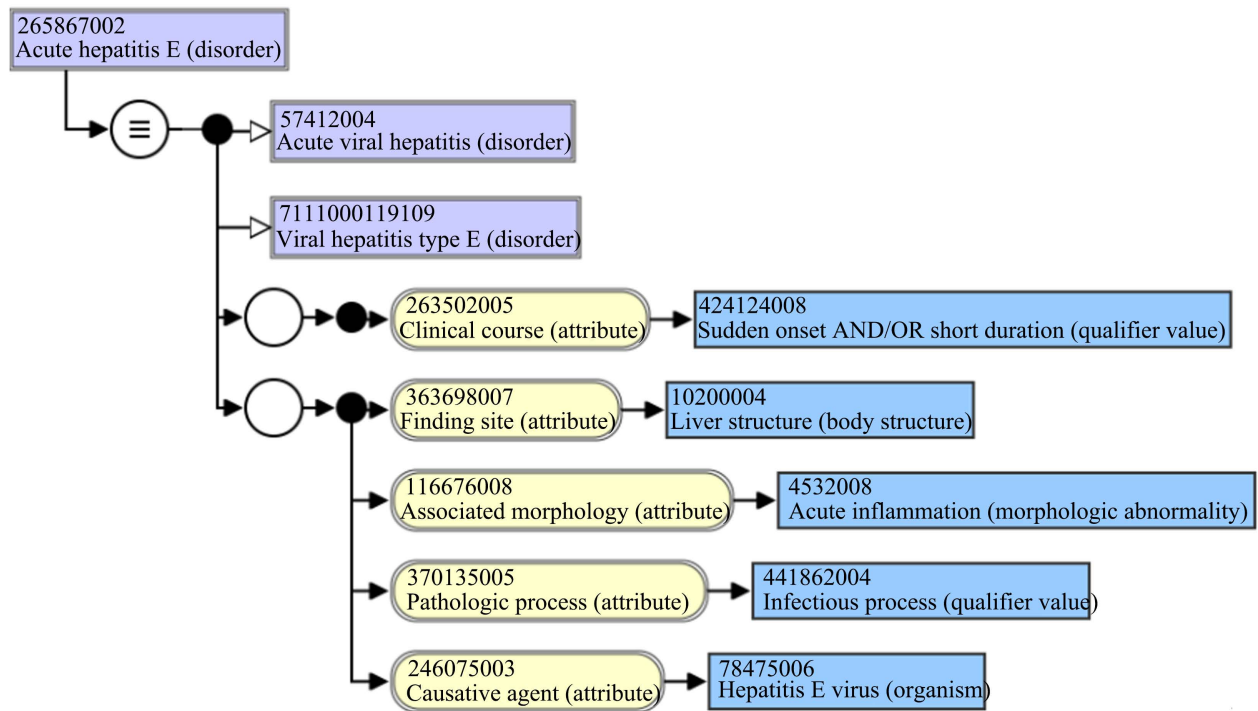
There is an incomplete ontology representation of the ICD11 code and textual definition by the fully defined concept model of the SCT concept lexically aligned with the ICD 11 code (see Chapter 2.2.3).

### 2.2.3. How to Complete an Ontology Representation of ICD-11 Codes When the SCT Concept Model of the SCT Concept Lexically Aligned Is Not Sufficient

There are 2 situations:

1) It is possible to represent the missing characteristics with the authorized attributes and values.

2) It is not possible but possible with unauthorized attributes and values available on the SNOMED browser Information Model and Compositional Grammar.



**Figure 3.** SNOMED CT definition by SNOMED CT concept model “235867002|Acute hepatitis E (disorder)”.

The appropriate attributes are first searched among the fifty attributes found in the conceptual model for clinical findings [17] and if not present, in the group of the so-called non-approved attributes (about 1000) [22]. Those are attributes that can be used to model definitions of SCT concepts or expressions, but that have not yet been authorized to model SCT concepts in the SNOMED Concept Model.

Result of the search for attributes and values in the example E50.4 Acute hepatitis E: **Figure 4.**

We have drawn a blue arrow between the attributes and values authorized by the SNOMED CT concept model today browser (over the blue arrow) and the attributes and values authorized or not authorized by the SNOMED CT concept model rules and not used in the present version of the SNOMED browser (under the blue arrow). We have isolated by a dotted line the attribute and value neither authorized nor validated by the present version of the SNOMED browser.

Step 1 authorized attributes and values

- For the symptom nausea, the attribute relation 363705008|Has definitional manifestation|does exist and is authorized with the concept value 422587007|Nausea (finding)|.
- The notion of confirmation is by detection of anti-hepatitis E virus IgM antibodies

Can be represented by the attribute-value pairs (authorized): Attribute: Interpret 363714003|Value: 710654003|Immunoglobulin M antibody to Hepatitis E virus (substance); Attribute: 363713009|Has interpretation (attribute)|Value: 260373001|Detected (qualifier value)|.

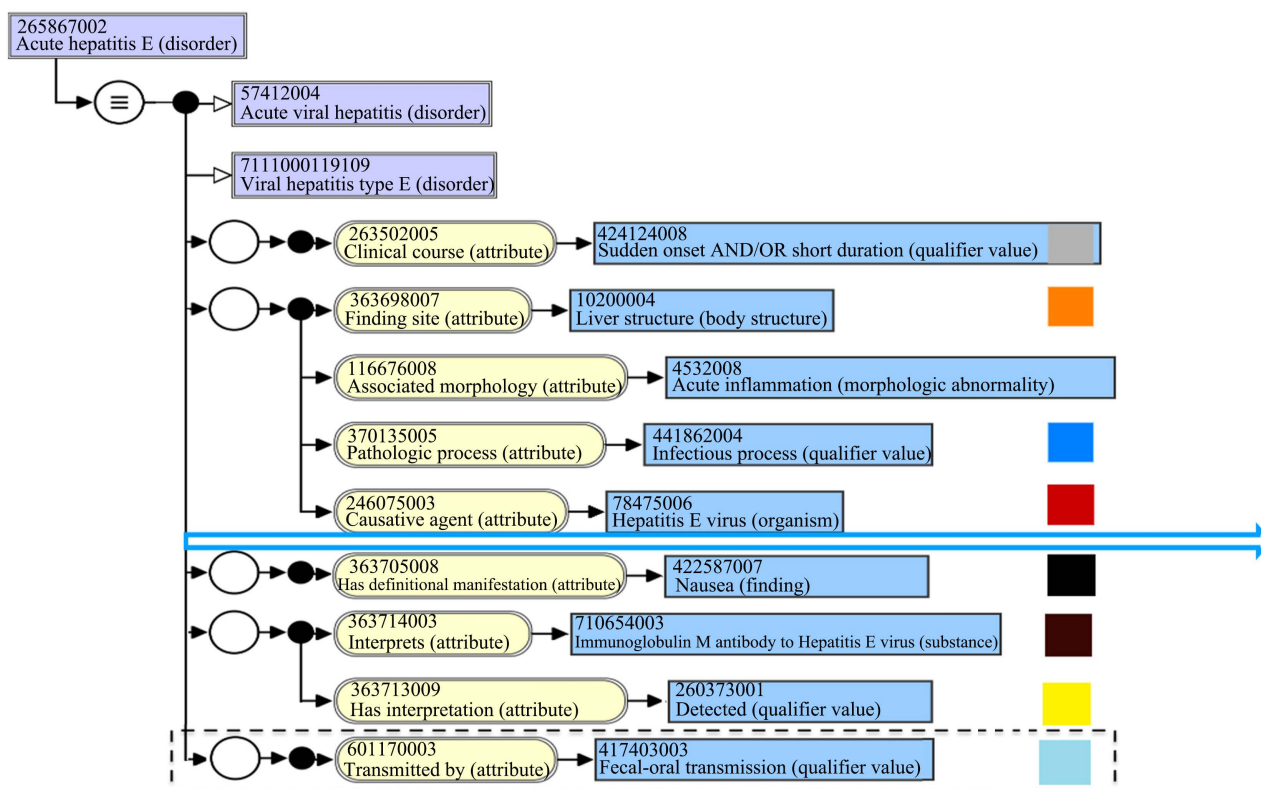


Figure 4. Represents the complete ontology representation of the ICD-11 code textual definition 1E50.4 Acute hepatitis E.

Step 2 unauthorized attributes

- The notion of transmission by the fecal-oral way is materialized by the attribute 60117003|Transmitted by|(not authorized) with the concept value 417403003|Fecal-oral transmission (qualifier value).

### 3. Results

#### 3.1. Lexical Stage

The final results are shown in Table 1. The alignment is total.

Table 1. ICD-11 SNOMED CT lexical alignment.

Lexical Alignments	Number	%
Pre-coordinated SCT concepts	615	94
Post-coordinated SCT Concepts	36	6
Total	651	100

#### 3.2. Ontological Stage

Table 2 and Table 3 present the results of the ontological analysis. When the SNOMED concept model is Not Fully Defined (FD) the ontological representation of ICD-11 codes and textual description is always incomplete but can be completed with validated or unvalidated attributes and values.

**Table 2.** Ontology representation of ICD11 codes by SCT FD concepts models.

Ontology Representation	Number	%
Complete with validated attributes and values	260	53
Incomplete and complete with validated or unvalidated attributes and values	231	47
Incomplete	0	0
Total	491	100

**Table 3.** Ontology representation of ICD11 codes by SCT Not FD concepts models.

Ontology Representation	Number	%
Complete with validated attributes and values	0	0
Incomplete and complete with validated or unvalidated attributes and values	160	100
Incomplete	0	0
Total	160	100

#### 4. Discussion

The method proposed in this work involves lexical alignment with MetaMap [13] and ontological representation with SNOMED concept model representation [17].

First, we found a 100% lexical match between ICD-11 titles and fully specified names or synonyms of SCT concepts, 94% for a pre-coordinated concept, and 6% for post-coordinated concepts. This result is not surprising given that most healthcare terminologies on a specific subject and in the same language share a large number of terms. During the development of ICD11 SCT was one healthcare terminology taken into account.

The second so-called ontology representation showed a reduced correspondence (53% for SCT concepts FD and 260/651 around 40% for the total) between ICD-11 textual definition or description and SCT concept model for SCT concepts lexically aligned with ICD-11 labels.

This result has two main consequences for the future:

1) It explains the limitations of the mapping between labels or fully defined names of different healthcare terminologies. It is an argument for developing rather a common ontology between them as for ICD11 and SNOMED CT.

2) It shows that the SCT ontology based on an information model and a compositional grammar easily available on the SNOMED browser is under-utilized today for the SCT concept model.

Our work has some limitations:

- Some ICD-11 titles do not yet have a description. We have considered that the ICD-11 code label took its place.
- This work focuses on only Chapter 1 of ICD-11 and shall be tested on the whole ICD-11.

- We excluded from our work the ICD-11 codes “other” and “unspecified” which cannot have any formal logical meaning and needs queries to be represented which limit the possibility to re-use aggregative data as ICD 11 outside mortality statistics. This is important for Epidemiology and Economic re-use as input for example in DRG/Casemix financing system.

## 5. Conclusions

1) We have shown in this work that a formal ontological representation of a health terminology proposing a detailed textual description of its titles as ICD-11 is possible by using the lexical alignment and ontology representation tools of other terminology systems (UMLS and SCT).

2) The work measures the gap between a lexical alignment and a meaning alignment between different health terminologies.

- In this example, the ICD11 textual definition is either fully represented by or sometimes more precise than the SNOMED concept model representation.
- As a consequence, there is a need to extend the use and validation of SNOMED attributes and values presently not authorized by SNOMED for the SNOMED codes concept model.
- The work opens the way to a SNOMED concept model-like formal Ontology of ICD-11 critical point before a joint development of a Common Ontology between ICD-11 and SNOMED CTs final goal for a seamless connection between the 2 worldwide health terminology systems and their multiple reuse.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Schulz, S., Stegwee, R. and Chronaki, C. (2018) Standards in Healthcare Data. In: Kubben, P., Dumontier, M. and Dekker, A., Eds., *Fundamentals of Clinical Data Science*, Springer International Publishing, 19-36. [https://doi.org/10.1007/978-3-319-99713-1\\_3](https://doi.org/10.1007/978-3-319-99713-1_3)
- [2] Benson, T. and Grieve, G. (2016) Principles of Health Interoperability. Springer, 451. <https://doi.org/10.1007/978-3-319-30370-3>
- [3] Venot, A., Burgun, A. and Quantin, C., (2012) Medical Informatics, e-Health: Fundamentals and Applications. Springer, 11-41. <https://doi.org/10.1007/978-2-8178-0478-1>
- [4] Giannangelo, K. and Millar, J. (2012) Mapping SNOMED CT to ICD-10. *Studies in Health Technology and Informatics*, **180**, 83-87.
- [5] Campbell, J.R., Brear, H., Scichilone, R., White, S., Giannangelo, K., Carlsen, B., Solbrig, H. and Fung, K.W. (2013) Semantic Interoperation and Electronic Health Records: Context Sensitive Mapping from SNOMED CT to ICD-10. *Studies in Health Technology and Informatics*, **192**, 603-607.
- [6] SNOMED International. SNOMED CT Browser 2021. <https://browser.ihtsdotools.org/?perspective=full&conceptId1=404684003&edition=>

[MAIN/2022-10-31&release=&languages=en](#)

- [7] Névéol, A., Shooshan, S.E., Humphrey, S.M., Mork, J.G. and Aronson, A.R. (2009) A Recent Advance in the Automatic Indexing of the Biomedical Literature. *Journal of Biomedical Informatics*, **42**, 814-823. <https://doi.org/10.1016/j.jbi.2008.12.007>
- [8] Rocha, R.A., Rocha, B.H. and Huff, S.M. (1993) Automated Translation between Medical Vocabularies Using a Frame-Based Interlingua. *Proceedings Symposium on Computer Applications in Medical Care*, Washington, DC, 30 October-3 November 1993, 690-694.
- [9] Smith, B. (1998) Applied Ontology: A New Discipline Is Born. *Philosophy Today*, **12**, 5-6.
- [10] Rodrigues, J.-M., Robinson, D., Della Mea, V., Campbell, J., Rector, A., Schulz, S., et al. (2015) Semantic Alignment between ICD-11 and SNOMED CT. *Studies in Health Technology and Informatics*, **216**, 790-794.
- [11] Rodrigues, J.-M., Schulz, S., Mizén, B., Trombert, B. and Rector, A. (2018) Scrutinizing SNOMED CT's Ability to Reconcile Clinical Language Ambiguities with an Ontology Representation. *Studies in Health Technology and Informatics*, **247**, 910-914.
- [12] Rector, A., Schulz, S., Rodrigues, J.M., Chute, C.G. and Solbrig, H. (2019) On Beyond Gruber: "Ontologies" in Today's Biomedical Information Systems and the Limits of Owl. *Journal of Biomedical Informatics*, **100**, Article 100002. <https://doi.org/10.1016/j.yjbinx.2019.100002>
- [13] Aronson, A.R. and Lang, F. (2010) An Overview of MetaMap: Historical Perspective and Recent Advances. *Journal of the American Medical Informatics Association*, **17**, 229-236. <https://doi.org/10.1136/jamia.2009.002733>
- [14] Koné, C.J., Babri, M. and Rodrigues, J.M. (2023) SNOMED CT: A Clinical Terminology but also a Formal Ontology. *Journal of Biosciences and Medicines*, **11**, 326-333. <https://doi.org/10.4236/jbm.2023.1111027>
- [15] WHO (2018) Eleventh Revision of the International Classification of Diseases. [https://apps.who.int/gb/ebwha/pdf\\_files/EB144/B144\\_22-en.pdf](https://apps.who.int/gb/ebwha/pdf_files/EB144/B144_22-en.pdf)
- [16] WHO (2022) ICD-11 for Mortality and Morbidity Statistics. <https://icd.who.int/browse11/l-m/en>
- [17] SNOMED CT Concept Model 2021.
- [18] Group SCLP (2021) SNOMED CT Compositional Grammar.
- [19] Bodenreider, O., Cornet, R. and Vreeman, D. (2018) Recent Developments in Clinical Terminologies—SNOMED CT, LOINC, and RxNorm. *Yearbook of Medical Informatics*, **27**, 129-139. <https://doi.org/10.1055/s-0038-1667077>
- [20] SNOMED International (2021) SNOMED CT MRCM Maintenance Tool 2022. <https://browser.ihtsdotools.org/mrcm/?branch=MAIN%2F2022-11-30>
- [21] Heiyanthuduwage, S.R., Schwitter, R. and Orgun, M.A. (2016) OWL 2 Learn Profile: An Ontology Sublanguage for the Learning Domain. *SpringerPlus*, **5**, Article No. 291. <https://doi.org/10.1186/s40064-016-1826-0>
- [22] National Health Service (2021) SNOMED CT Browser-UK SNOMED CT Clinical Edition 2022.