# SNOMED CT: A Clinical Terminology but Also a Formal Ontology

## Constant Joseph Koné[1*], Michel Babri[1], Jean Marie Rodrigues[2]

[1]Félix Houphouët-Boigny National Polytechnic Institute, Yamoussoukro, Cote d'Ivoire
[2]University Jean Monnet, Saint Etienne, France
Email: *koneconstant@yahoo.fr

## Abstract

**Context and Objective:** Over the past few decades, terminologies developed for clinical descriptions have been increasingly used as key resources for knowledge management, data integration, and decision support to the extent that today they have become essential in the biomedical and health field. Among these clinical terminologies, some may possess the characteristics of one or several types of representation. This is the case for the Systematized Nomenclature of Human and Veterinary Medicine—Clinical Terms (SNOMED CT), which is both a clinical medical terminology and a formal ontology based on the principles of semantic web. **Methods:** We present and discuss, on one hand, the compliance of SNOMED CT with the requirements of a reference clinical terminology and, on the other hand, the specifications of the features and constructions of descriptive of SNOMED CT. **Results:** We demonstrate the consistency of the reference clinical terminology SNOMED CT with the principles stated in James J. Cimino's desiderata and we also show that SNOMED CT contains an ontology based on the EL profile of OWL2 with some simplifications. **Conclusions:** The duality of SNOMED CT shown is crucial for understanding the versatility, depth, and scope in the health field.

## Keywords

SNOMED CT, Clinical Terminology, Ontology, OWL2

## 1. Introduction

There are several types of terminological systems, including thesauruses, terminologies, classifications, taxonomies, vocabularies, nomenclatures, and ontologies [1]. Some terminological resources can possess the characteristics of one or several of these representation artifacts. This is the case with the Systematized

Nomenclature of Human and Veterinary Medicine—Clinical Terms (SNOMED CT) [2], which is both a globally used clinical medical terminology that covers all specialties, disciplines, and clinical requirements and a formal ontology based on the principles of the semantic web, with the Ontology Web Language (OWL) [3] [4] as its reference language.

OWL2, the current version, is a powerful modeling and explication language for general use in some areas of human knowledge [4]. Humans need knowledge and wisdom to derive implications from their understanding. An essential feature of OWL is that it captures the meaning and significance of the knowledge it can represent, surpassing the simple string of characters of terms or words in a language [5] [6] [7].

In this work we present and discuss the compliance of SNOMED CT with the requirements of a clinical reference terminology and with the specifications of the set of features and descriptive logic constructs of SNOMED CT in OWL2 that have an impact on the implementation and maintenance of SNOMED CT.

## 2. Materials and Methods

SNOMED CT (Systematized Nomenclature of Medicine Clinical Terms) is the most comprehensive and internationally recognized medical terminology in the world. It provides a coded language for uniformly and systematically representing clinically meaningful information. SNOMED CT facilitates the sharing and analysis of health data across healthcare systems and applications. Used for the electronic coding of medical data, it enhances the quality, safety, and efficiency of healthcare.

Initially, we discuss the alignment of SNOMED CT with the desiderata articulated by James J. Cimino [8]. After listing Cimino's 12 desideratas, we search for and present characters of SNOMED CT that relate to each desiderata.

Subsequently, in a first part, we briefly present the history of ontology and its implication in digital information sciences. We then present and deliberate on the differences between the logical profile implemented in SNOMED CT in EL (Expression Language) and the original version of OWL2. Specifically, we have identified the constructors specific to each system. We then developed a comparison matrix to highlight the similarities and differences between the two systems.

## 3. Results

### 3.1. SNOMED CT: A Reference Clinical Terminology

SNOMED CT is built on the design requirements for reference clinical terminology outlined in 1998 and expanded in 2001 in a landmark article—"Desiderata for controlled medical vocabularies in the twenty-first century" by James J. Cimino [6]. These requirements can be summarized as follows: 1) content importance, 2) concept consideration, 3) concept permanence, 4) non-significant concept identifier, 5) poly-hierarchy, 6) formal definitions, 7) rejection of the "not

classified elsewhere" notion, 8) multiple granularity, 9) multiple and coherent viewpoints, 10) context representation, 11) capability to evolve, and 12) consideration of redundancy.

In these Desideratas, James J. Cimino talks about the completeness of vocabulary, that is, domain coverage. In this respect, SNOMED CT can boast a broad coverage of the medical and health field by having about 350,000 concepts that cover anatomy, pathologies, observable variables, organisms, pharmaceutical and biological products, interventions, etc. (Desiderata Cimino 1).

According to SNOMED, a concept is defined as "a unit of knowledge created by a unique combination of characters". It has a preferred name and several synonyms in every language available in SNOMED CT, including French (Desiderata Cimino 2 and 3).

Each concept is associated with a non-significant identifier of a positive 64-bit integer type, having a minimum and maximum authorized length of 6 and 18 digits respectively (Desiderata Cimino 4) [9].

The primary hierarchical relation of SNOMED CT is the "is a" relation, forming the base of SNOMED CT hierarchies with a root concept that has no concept above it. A concept with a "is a" relation to a parent concept (more general concept) is called a descendant concept. SNOMED CT allows a concept to have more than one "is a" relation with other concepts, indicating that SNOMED CT has a poly-hierarchical structure (Desiderata Cimino 5).

Unlike aggregations or classifications where all considered domain objects must be classified in a single exclusive and exhaustive place, SNOMED CT is not bound by these requirements. Domain objects are added as needed and from different viewpoints (Desiderata Cimino 9 and 11). There are no concepts like "others, unspecified or not classified elsewhere" (Desiderata Cimino 7).

SNOMED CT has evolved by expanding its coverage of the medical and health field at the contextual level. Initially, SNOMED CT representations used only three hierarchies, but now it has 19, allowing the collection of information not only on specific clinical situations but also on the context (Desiderata Cimino 10) [10].

We often encounter similar terms or concepts at different levels of SNOMED CT, known as redundancy. This redundancy allows more choices when using SNOMED CT (Desiderata Cimino 12). Various mechanisms within SNOMED CT account for this redundancy, including synonymy, concept hierarchy, and concept equivalence.

Lastly, for some hierarchies, SNOMED CT provides a formal ontological representation, which we will address in the following chapter (Desiderata Cimino 6).

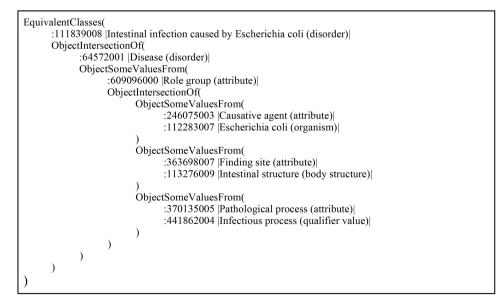### 3.2. SNOMED CT: A Formal Ontology

Ontology is a branch of philosophy that can be defined as the study of what "IS" in the physical world [7]. The application of this approach in digital information sciences, especially in health, began in the 1990s as a representation of informa-

tion entities used in a knowledge base. The most widely cited definition of formal digital ontology is that of Gruber [11]: "An ontology is an explicit specification of a conceptualization." By conceptualization, Gruber means an abstract and simplified view of the world's representation, and by explicit specification, he means that this conceptualization is made unambiguously in a concrete language. Gruber defined five types of components for describing a domain's knowledge: 1) concepts, 2) relations, 3) functions, 4) axioms, and 5) instances. Later, Gruber's definition was supplemented by Studer and his colleagues [12], stating that a formal ontology is a formal and explicit specification of a shared conceptualization. The term "formal" implies that the applied ontology should be machine-readable, meaning machines should be able to interpret the semantics or the meaning of the provided information.

Formal digital ontology has found its place in the biomedical field, confronted with an explosion of knowledge contained in heterogeneous terminological systems [13] [14]. In this regard, SNOMED CT, for certain hierarchies, provides a formal ontological representation that expresses the formal logic of a "Concept Model" following the rules of "compositional grammar" [3] in a subset or profile of OWL 2 named EL for Expression Language, offering application possibilities [15] [16] [17] [18]. For example, the disease "acute intestinal infection due to *Escherichia coli*", represented by the concept 111839008 |Intestinal infection caused by *Escherichia coli* (disorder)|, is equivalent to the logical expression using the following OWL functional syntax, see **Figure 1**.

However, the OWL2 logical profile of SNOMED CT does not cover all the expressive possibilities of EL OWL2 [15] for several reasons, among which complexity issues are found.

We provide a comparative table of the constructors and entities of the EL logical profile of SNOMED CT and the OWL2 specification (see **Table 1**).

```
EquivalentClasses(
    :111839008 |Intestinal infection caused by Escherichia coli (disorder)|
    ObjectIntersectionOf(
        :64572001 |Disease (disorder)|
        ObjectSomeValuesFrom(
            :609096000 |Role group (attribute)|
            ObjectIntersectionOf(
                ObjectSomeValuesFrom(
                    :246075003 |Causative agent (attribute)|
                    :112283007 |Escherichia coli (organism)|
                )
                ObjectSomeValuesFrom(
                    :363698007 |Finding site (attribute)|
                    :113276009 |Intestinal structure (body structure)|
                )
                ObjectSomeValuesFrom(
                    :370135005 |Pathological process (attribute)|
                    :441862004 |Infectious process (qualifier value)|
                )
            )
        )
    )
)
```

**Figure 1.** Definition of "111839008 |Intestinal infection caused by *Escherichia coli* (disorder)|" by the OWL Functional Syntax.

**Table 1.** Comparison of constructors of SNOMED CT's EL logical profile and OWL2.

| Constructors | Logic profile EL SNOMED CT | Logic profile EL OWL2 |
|---|---|---|
| Entities, literal and anonymous | - Class<br>- ObjectProperty<br>- DataProperty<br>- Datatype<br>- Literal<br>- AnnotationProperty | - Class<br>- ObjectProperty<br>- DataProperty<br>- Datatype<br>- Literal<br>- AnnotationProperty<br>- DataRange<br>- Individual<br>- NamedIndividual<br>- AnonymousIndividual |
| Data types | - xsd:decimal<br>- xsd:integer<br>- xsd:string<br>- xsd:dateTime | - rdf:PlainLiteral<br>- rdf:XMLLiteral<br>- rdfs:Literal<br>- owl:real<br>- owl:rational<br>- xsd:decimal<br>- xsd:integer<br>- xsd:nonNegativeInteger<br>- xsd:string<br>- xsd:normalizedString<br>- xsd:token<br>- xsd:Name<br>- xsd:NCName<br>- xsd:NMTOKEN<br>- xsd:hexBinary<br>- xsd:base64Binary<br>- xsd:anyURI<br>- xsd:dateTime<br>- xsd:dateTimeStamp |
| Class restriction | - existential quantification to a class expression (ObjectSomeValuesFrom)<br>- intersection of classes (ObjectIntersectionOf)<br>- existential quantification to a literal (DataHasValue) | - existential quantification to a class expression (ObjectSomeValuesFrom)<br>- existential quantification to a data range (DataSomeValuesFrom)<br>- existential quantification to an individual (ObjectHasValue)<br>- existential quantification to a literal (DataHasValue)<br>- universal quantification to a class expression (ObjectAllValuesFrom)<br>- universal quantification to a range of data (DataAllValuesFrom)<br>- cardinality restrictions (ObjectMaxCardinality, ObjectMinCardinality, ObjectExactCardinality, DataMaxCardinality, DataMinCardinality, and DataExactCardinality)<br>- self-restriction (ObjectHasSelf)<br>- enumerations involving a single individual (ObjectOneOf)<br>- enumerations involving a single literal (DataOneOf)<br>- intersection of classes (ObjectIntersectionOf)<br>- intersection of data ranges (DataIntersectionOf) |

Continued

| Class expressions | - class inclusion (SubClassOf)<br>- class equivalence (EquivalentClasses)<br>- inclusion of object properties (SubObjectPropertyOf)<br>- inclusion of data properties (SubDataPropertyOf)<br>- class disjunction (DisjointClasses)<br>- property equivalence (EquivalentObjectProperties and EquivalentDataProperties)<br>- properties of the transitive object (TransitiveObjectProperty)<br>- reflexive object properties (ReflexiveObjectProperty) | - class inclusion (SubClassOf)<br>- class equivalence (EquivalentClasses)<br>- inclusion of object properties (SubObjectPropertyOf)<br>- inclusion of data properties (SubDataPropertyOf)<br>- class disjunction (DisjointClasses)<br>- property equivalence (EquivalentObjectProperties and EquivalentDataProperties)<br>- properties of the transitive object (TransitiveObjectProperty)<br>- reflexive object properties (ReflexiveObjectProperty)<br>- domain restrictions (ObjectPropertyDomain and DataPropertyDomain)<br>- range restrictions (ObjectPropertyRange and DataPropertyRange)<br>- assertions (SameIndividual, DifferentIndividuals, ClassAssertion, ObjectPropertyAssertion, DataPropertyAssertion, NegativeObjectPropertyAssertion and NegativeDataPropertyAssertion)<br>- functional data properties (FunctionalDataProperty)<br>- keys (HasKey)<br>- disjunction (ObjectUnionOf, DisjointUnion, and DataUnionOf)<br>- class negation (ObjectComplementOf)<br>- disjoint properties (DisjointObjectProperties and DisjointDataProperties)<br>- irreflexive object properties (IrreflexiveObjectProperty)<br>- inverse object properties (InverseObjectProperties) |

Table 2 below illustrates the proportion of usage of the constructors of the EL OWL2 logical profile.

**Table 2.** Comparison of the number of constructors and entities of the EL logical profile in SNOMED CT and OWL2.

| Types of Constructors and entities | Logic profile EL SNOMED CT | Logic Profile EL OWL2 | Percentage of constructors and entities used |
|---|---|---|---|
| Anonymous entities, literals and individuals | 6 | 10 | 60% |
| Data types | 4 | 19 | 21% |
| Class restrictions | 3 | 17 | 18% |
| Class expressions | 9 | 30 | 30% |

## 4. Discussion

SNOMED CT qualifies as a medical terminology if we adhere to the 12 recommendations by J. J. Cimino. Indeed, these 12 outlined criteria have been thoroughly incorporated in the development of this terminology.

When looking at the ontological facet of SNOMED CT, on average, 32% of the constructors available in EL OWL2 are used. The lowest utilization rate is associated with class restrictions, with only three out of 17 available constructors being used, accounting for 18%.

On the other hand, a number of constructors from the EL logical profile of OWL2 are excluded from that of SNOMED CT due to their complexity and/or additional impact on reasoning times with the current hardware and algorithms. The main features excluded from the SNOMED CT EL logical profile are:

- Universal quantification (ObjectAllValuesFrom);
- Disjunction (ObjectUnionOf, DisjointUnion, and DataUnionOf);
- Class negation (ObjectComplementOf);
- Domain and value constraints which are supported via the SNOMED CT Machine Readable Concept Model;
- Anonymous entities because SNOMED CT does not see their utility in its context of use.

There is a natural tension between the desire to use features so that the content can be more expressive and precise and the cost and complexity of these features. In other words, it allows for sufficient expressiveness for the content while remaining manageable for implementation.

The constructors excluded from the EL profile of SNOMED CT have too high a cost for the implementation of SNOMED CT to be bearable at the moment. For example, in his study, S. Schulz showed that the use of negation is not recommended in SNOMED CT expressions [18]. Extensions of SNOMED CT can choose to implement even more features than the SNOMED CT logical profile offers, with the consequence of not being able to function properly and having reasoning times that increase significantly.

## 5. Conclusion

In this text, we highlight the consistency of the reference clinical terminology SNOMED CT with the principles outlined in J. J. Cimino's desiderata for the 21st-century reference clinical terminologies. A balanced approach that takes into account complexity, cost, and the added value of features is essential to optimize the utility and efficiency of SNOMED CT in the future.

## Conflicts of Interest

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

## References

[1] Duclos, C., *et al.* (2014) Medical Vocabulary, Terminological Resources and Information Coding in the Health Domain. In: Venot, A., Burgun, A. and Quantin, C., Eds., *Medical Informatics, e-Health, Health Informatics*, Springer, Paris. https://doi.org/10.1007/978-2-8178-0478-1_2

[2] SNOMED International (2023) SNOMED CT Starter Guide. https://confluence.ihtsdotools.org/display/docstart/snomed+ct+starter+guide

[3] SNOMED Compositional Grammar (2023) Specification and Guide. https://confluence.ihtsdotools.org/display/DOCSCG

[4] Group WCOW (2023) OWL 2 Web Ontology Language Primer (Second Edition).

https://www.w3.org/TR/owl2-overview/

[5] Roche, C. (2005) Terminologie et ontologie. *Langages*, **157**, 48-62.
https://doi.org/10.3917/lang.157.0048

[6] Geller, J. (2003) John Sowa, Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks/Cole, 2000, 512 pp., $70.95 (Cloth), ISBN 0-534-94965-7. *Minds and Machines*, **13**, 441-444.
https://doi.org/10.1023/A:1024132024747

[7] Monnin, A. and Félix, E. (2009) Essai de comparaison des ontologies informatiques et philosophiques: Entre être et artefacts: Ontologie et dynamique des systèmes complexes, perspectives interdisciplinaires. Megève, France, p. 14.

[8] Cimino, J.J. (1998) Desiderata for Controlled Medical Vocabularies in the Twenty-First Century. *Methods of Information in Medicine*, **37**, 394-403.
https://doi.org/10.1055/s-0038-1634558

[9] SNOMED CT Release File Specifications (2023) SNOMED CT Identifiers.
https://confluence.ihtsdotools.org/display/DOCRELFMT/6+SNOMED+CT+Identif
iers

[10] SNOMED International, SNOMED (2023) SNOMED CT Concept Model.
https://confluence.ihtsdotools.org/display/DOCSTART/6.+SNOMED+CT+Concept
+Model

[11] Genesereth, M., Fikes, R., Brachman, R., Gruber, T., Hayes, P., Letsinger, R., *et al.* (1992) Knowledge Interchange Format Version 3.0 Reference Manual.
http://cs.auckland.ac.nz/courses/compsci367s2c/resources/kif.pdf

[12] Studer, R., Benjamins, R. and Fensel, D. (1998) Knowledge Engineering: Principles and Methods. *Data & Knowledge Engineering*, **25**, 161-197.
https://doi.org/10.1016/S0169-023X(97)00056-6

[13] Bodenreider, O., Cornet, R. and Vreeman, D.J. (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

[14] 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 ID: 100002.
https://doi.org/10.1016/j.yjbinx.2019.100002

[15] W3C Recommendation (2023) OWL2 Web Ontology Language Document Overview (Second Edition). http://www.w3.org/TR/2012/REC-owl-overview-20121211/

[16] 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

[17] El-Sappagh, S., Franda, F., Ali, F. and Kwak, K.-S. (2018) SNOMED CT Standard Ontology Based on the Ontology for General Medical Science. *BMC Medical Informatics and Decision Making*, **18**, Article No. 76.
https://doi.org/10.1186/s12911-018-0651-5

[18] Schulz, S., Markó, K. and Suntisrivaraporn, B. (2008) Formal Representation of Complex SNOMED CT Expressions. *BMC Medical Informatics and Decision Making*, **8**, Article No. S9. https://doi.org/10.1186/1472-6947-8-S1-S9