Definitions in SNOMED CT through the lens of Terminology: from formal to textual

Sara Carvalho [1,2], Cornelia Wermuth [3] and Rute Costa [1]

[1] *NOVA-CLUNL, NOVA FCSH, Avenida de Berna, 26-C, Lisboa, 1069-061, Portugal*
[2] *CLLC-UA, Departamento de Línguas e Culturas da Universidade de Aveiro, Aveiro, 3810-164, Portugal*
[3] *KU Leuven, Sint-Jacobsmarkt 49-51, Antwerp, 2000, Belgium*

Abstract

This paper aims to show how Terminology can help foster interoperability and more effective knowledge representation, organisation and sharing in the biomedical field, and on the other hand, support specialised communication among various stakeholders. SNOMED CT will be used to illustrate this, with the focus being on formal and textual (or natural language) definitions – the latter currently underrepresented in this resource - and on how a double-dimensional terminological approach can benefit textual definition drafting, thereby assisting the work carried out by SNOMED CT national translation teams.

Keywords

terminology, formal definition, natural language definition, SNOMED CT

1. Introduction

Over the last decades, healthcare has been impacted by critical changes resulting mainly from societal, economic, and technological factors. Healthcare-related technology has grown exponentially, backed by more advanced eHealth and mHealth solutions [1], [2], as well as by a recent trend towards the use of the Internet of Things (IoT) in this field [3], [4]. This, in turn, has led to an unprecedented collection, storage, and transfer of massive amounts of data. Big data and machine learning have revolutionised healthcare by supporting more effective decision-making processes, namely in disease prediction [5], despite the ongoing ethical and data privacy concerns [6]. Interoperability, i.e. the ability to effectively exchange and use information across systems or applications [7], and its underlying challenges have also been at the heart of the current healthcare landscape. While technological innovation has been gaining significant traction, the human side of healthcare has additionally become paramount in this sector. Care is increasingly people-/patient-centred, with tailored approaches designed to respond to the patient's needs. In the present digital healthcare ecosystem, the patient is taking on an active and more empowered role concerning both the collection of his/her personal data and the way that data is stored and shared [8].

Within this framework, this paper aims to show how Terminology can help address the challenges that characterise both the technological and the human spheres: on the one hand, by fostering interoperability and more effective knowledge representation, organisation, and sharing in this field, and on the other hand, by supporting specialised communication among various stakeholders, especially patients. The international biomedical reference terminology SNOMED CT will be used to illustrate the preceding points, with the focus being on the formal and textual (or natural language) definitions in this resource. Since the latter are currently underrepresented in this reference terminology, it will be demonstrated how Terminology can support textual definition drafting, thereby assisting the translation of SNOMED CT into different languages.

2nd International Conference on "Multilingual digital terminology today. Design, representation formats and management systems" (MDTT) 2023, June 29-30, 2023, Lisbon, Portugal.
EMAIL: sara.carvalho@ua.pt (S. Carvalho); cornelia.wermuth@kuleuven.be (C. Wermuth); rute.costa@fcsh.unl.pt (R. Costa)
ORCID: 0000-0002-7501-5405 (S. Carvalho); 0000-0001-5144-8508 (C. Wermuth); 0000-0002-3452-7228 (R. Costa)

The paper is thus structured as follows: after this introduction, we present the theoretical and methodological background to our knowledge-organisation approach to terminology work (Section 2); next, a description of SNOMED CT's principles and structure is provided, plus information about the translation of this resource and types of definitions (Section 3); based on a case study from SNOMED CT, we then analyse a formal definition and propose a methodology to draft a natural language definition anchored in terminological principles (Section 4); the fifth and final section is dedicated to concluding remarks and future work.

2. Terminology and Knowledge Organization: some prolegomena

The advent of the Semantic Web [9] and, more recently, of Linked Data [10], has paved the way for new perspectives and opportunities for Terminology, namely by making terminological resources (inter)operable via ontologies (computational artefacts in the sense of Knowledge Engineering [11]). Terminology's latest theoretical frameworks have been aiming at a more consistent interrelation between what is linguistic and what is conceptual. This paper follows this paradigm and, thus, a double-dimensional approach to Terminology [12], [13], which considers, within a given domain, not only the systematic and semi-automatic processing and analysis of specialised texts and their terms[2] but also a (semi)formal representation of the core concepts[3] and their respective organisation into systems. These two dimensions, autonomous yet interconnected, are indispensable to terminology work. Recent research [14] supported by this double-dimensional framework has shown how a mixed methodology departing from an onomasiological approach can enrich the various stages of the process (**Annex 1**) and provide added value to the subsequent results, e.g. definitions – both formal and textual.

Definitions are one of the key forms of concept representation in Terminology and are strongly based on an Aristotelian perspective (*definiendum* = *genus proximum* + *differentia specifica*) [15]. According to [16] and [17], the intensional definition, comprising the "immediate generic concept and the delimiting characteristics" [16], is regarded as preferential, as it allows a given concept to be identified within a concept system and enables its differentiation from other related concepts. While intensional definitions play a central role in Terminology, the actual process of drafting such definitions is not exempt from challenges: firstly, it is still a rather manual process and hence, very time-consuming, despite the various guidelines that have been published throughout the years [18], [19], [20], [21]; secondly, and if stemming exclusively from lexical-semantic relations, the definitions may be "prone to all kinds of inconsistencies" [22, p. 24], since they are language-dependent and therefore, potentially influenced by polysemy, ambiguity, and ellipses; thirdly, even if supported by conceptual relations, intensional definitions, when exclusively based on hierarchical concept systems, may be unable to fully depict the multidimensionality that characterises some concepts [23].

When addressing the topic of definitions and their application to terminology work, reference must also be made to the notion of definitional templates, initially proposed by [24] and further developed within the scope of Frame-based Terminology (FBT) [25], [26]. Comprising a slot-filler core structure, in which the slots correspond to domain-specific conceptual relations (both hierarchical and non-hierarchical) and the fillers (values) are replaced by the concepts linked to the *definiendum* via the preceding conceptual relations, this template framework provides a stable foundation that guides the drafting of natural language definitions, which then become "mini-knowledge representations" [25, p. 345]. In FBT, this definitional template approach has been applied to various domains, namely healthcare [27], the environment, with EcoLexicon [28], and tourism [29].

The work presented in this paper aims to take this definitional template approach further by incorporating ontologies and using formal SNOMED CT definitions as a use case. It is believed that providing a language-independent, logically consistent, and interoperable conceptual foundation can help optimise the drafting process without jeopardising the added value supplied by linguistic diversity.

[2] Our definition of *term* follows the one proposed by [16]: "designation that represents a general concept by linguistic means".

[3] In this work, *concept* is defined as a "unit of knowledge created by a combination of characteristics" (*ibid.*)

3. SNOMED CT: core principles, definition(s) and the translation process

SNOMED CT is a multilingual reference terminology in healthcare that covers all the areas of medical activity[4]. Maintained by SNOMED International, it has been developed as a knowledge resource for information retrieval and exchange for use in various applications (e.g. to record and process medical data in electronic health records). As regards its core structure, the system consists of a relationship-based representation of concepts using the formal ontology language Description Logic (e.g. Common Cold <causative agent> Virus). The content of SNOMED CT is represented using three types of components: concepts, descriptions, and relationships, further detailed in **Annex 2**.

SNOMED CT contains more than 350,000 concepts from clinical and non-clinical concept fields or domains. Each SNOMED CT concept represents a "clinical idea" (i.e. the meaning of a medical term) and is represented by a machine-readable unique numerical concept identifier. The concepts are arranged in 19 top-level hierarchies according to the domain to which they belong. Within these hierarchies, the concepts are ordered from general to specific.

The descriptions in SNOMED CT link human-understandable medical terms to concepts. Descriptions, which also have a machine-readable unique numeric description identifier, are divided into the Fully Specified Name (FSN) and Synonyms (SYN), one of which is labelled as Preferred Term (PT). This fulfils the requirements of a standardised medical term, which is commonly used in clinical practice. Other synonyms are labelled as "acceptable". The provision of additional synonyms next to the PT is optional and useful for retrieval purposes. Each concept is represented by at least the FSN and the PT. Each SNOMED CT translation contains an additional set of descriptions that link medical terms in another language to the same SNOMED CT concepts.

In SNOMED CT, two types of relationships are used: hierarchical relationships (Is-a-Relationships) and associative relationships (Attribute Relationships). The |is-a|-relationship links a concept to more general concepts in a hierarchy and forms the concept hierarchies in SNOMED CT. At the top of the SNOMED CT hierarchy is the root concept (|SNOMED CT concept|), followed by its direct subtypes, called Top Level Concepts. Each concept can also have several parent concepts, which results in a polyhierarchical structure. Attribute relationships represent different aspects of the meaning of a concept and connect concepts in different hierarchies. For example, the concept |appendicitis| is linked to |Inflammation| via the relation type |Associated morphology| [30].

It should also be noted that SNOMED CT is an English-language terminology standard translated into 16 languages or language varieties (as of October 2022). High-quality translations are essential for the reliable use of this terminology system. Translations of SNOMED CT are larger projects, coordinated by the National Release Center (NRC) of a SNOMED International Member Country and regulated by guidelines for the management of both the translation process and the translation approach. Due to the considerable size of the terminology, only SNOMED CT content determined by the respective NRCs as a priority is translated, rather than a full translation. This is usually carried out by interdisciplinary teams (experts in medical informatics, (bio)medicine, linguistics, terminology).

The translation of SNOMED CT terms should be concept-based, whereby the precise description of the so-called "clinical idea" is of pivotal importance. For this purpose, translations should start from the FSN, a description that unambiguously reflects the meaning of a concept, regardless of the context in which it is used. The FSN is composed of a term and a "semantic tag" between parenthesis (e.g. |Myocardial biopsy (procedure)|). The tag indicates the hierarchy to which the concept belongs (e.g. procedure, disorder, etc.). Each concept has a FSN, which is unique in SNOMED CT in a given language, even in cases where identical terms refer to concepts belonging to different categories. For example, |Hematoma (morphologic abnormality)| is the FSN that represents what the pathologist sees under the microscope, whereas |Hematoma (disorder)| is the FSN that indicates the clinical diagnosis of a hematoma. After looking at the FSN, translators then focus on how the concept has been modelled in the ontology, to identify its hierarchical position and its relationships to other concepts. In case of ambiguity, the meaning of a concept can be checked via a textual definition, when it exists. If necessary, the use of a given target term should be analysed in context and/or domain experts should be consulted.

---

[4] The resource can be accessed at https://www.snomed.org/.

SNOMED CT concepts are defined in three different ways: (1) The FSN, which expresses the meaning of a SNOMED CT concept in a human-readable form. (2) A formal concept definition, which provides a computer-processable representation of a concept's meaning. Conceptual relations are generated in two ways: either by human users proposing a new concept, or automatically, by the SNOMED CT OWL description logic classifier. The manually generated representation is called a stated view, whereas the automatically generated representation is an inferred view. Concepts can also be either fully defined (i.e. their definition is sufficient to computably distinguish it from other concepts) or primitive (i.e. their definition is not sufficient to computably distinguish it from other concepts); (3) Textual definitions, which describe the meaning of a concept in natural language [31].

For translation purposes, the three types of definitions can be useful. However, professional translators without medical expertise will have greater difficulty understanding all SNOMED CT concepts starting from the FSN and/or their formal relationships. Moreover, the formal representation of concepts may contain errors which can only be detected by translators with sufficient domain knowledge. Therefore, natural language definitions and textual resources that provide information about the contextual use of a term are key additional tools. In fact, a 2010 survey to SNOMED CT direct users showed that 63% of the respondents claimed textual definitions would be useful [32]. However, for the time being, textual definitions in SNOMED CT are optional and only provided for a limited number of concepts[5], when there is a requirement for additional detail. As there is a clear need for more natural language definitions for translation purposes, efforts are being undertaken by SNOMED International to enrich the resource with this type of information. A relevant research question is, therefore, how such natural language definitions should be created in SNOMED CT and in which way the formal definitions can be used as templates for their generation.

The interconnection of formal and textual definitions in healthcare has been explored over the last decade. On the one hand, work has been done on the automatic generation of formal concept definitions from texts [33], [34], [35], with SNOMED CT being regarded as a 'model' for how formal definitions should be represented. On the other hand, the reverse has also been tackled, i.e. automatically generating textual definitions from formal ones: [36] proposed a basis for ICD-11's textual definitions; [37] created OntoVerbal, a Protégé plugin based on SNOMED CT which is no longer available; and, more recently, [38] developed Sno2Eng, a promising project which creates natural language definitions from SNOMED CT formal definitions, as shown in **Annex 3**, albeit only for diseases. However, as mentioned earlier, the formal definitions in SNOMED CT may not be stable or may contain errors, which will directly impact the quality of the automatically generated textual definitions. This is where Terminology can help improve the consistency of such results, both at the conceptual and linguistic levels.

4. Definition(s) in SNOMED CT and the role of Terminology: a case study

Due to space limitations, only one SNOMED CT example will be analysed in this paper: it refers to |Single-port laparoscopic cholecystectomy (procedure)|, a primitive concept without a natural language definition proposal. As we can see from the inferred view diagram depicted in **Annex 4**, this concept is defined as a subtype of |Laparoscopic cholecystectomy (procedure)|, thereby inheriting its attribute relationships and respective values (e.g. |Method (attribute)| |Excision-action (qualifier value)|. |Single-port laparoscopic cholecystectomy (procedure)| has, therefore, the same formal definition of |Laparoscopic cholecystectomy (procedure)| - a fully defined concept -, which might lead one to believe that there is no difference between these two concepts. However, the delimiting characteristic of the concept under analysis is **not** represented in the formal definition, despite its presence in the FSN ("single-port"). In fact, there is currently no conceptual distinction in SNOMED CT between a conventional laparoscopic approach (with multiple skin incisions) and an approach with a single incision (also called laparoendoscopic single-site approach[6]). Although the concept of |Incision of skin (procedure)| exists in SNOMED CT, its subtype concepts are related to the location of the incision

[5] The current US version of SNOMED CT contains 4,327 textual definitions, covering 2,608 diseases, which corresponds to only 1.3% of all SNOMED CT concepts and 3.3% of disease concepts [38].

[6] Cf. [39] for a more comprehensive review of this type of surgical approach and its terminological implications.

rather than to the number of incisions. It is believed that such a distinction could be relevant and should be clearly stated in an EHR, since these differences in surgical approach may have implications in medical coding, insurance, and billing. One proposal to incorporate this essential characteristic into the formal definition, thereby allowing the concept to become fully defined, is based on EndoTerm[7].

The EndoTerm project combines an Aristotelian-based, specific difference approach in the main hierarchies with a customised set of non-hierarchical concept relations, mostly based on SNOMED CT and UMLS[8], as well as with a categorial structure to prevent logical errors [40]. The concept systems within EndoTerm were then encoded as a formal ontology, which enables interoperability. One of EndoTerm's hierarchies is the <Surgical approach> one (a semi-formal visualisation of the concept system can be seen in **Annex 5**), in which the number of skin incisions is introduced as an axis of analysis, thereby allowing single-skin-incision and multiple-skin-incision approaches to be explicitly distinguished. This hierarchy was integrated into EndoTerm's ontology, with an example being provided in **Annexes 6** and **7** for the micro-concept system and the formal definition, respectively, of |Laparoendoscopic single-site total hysterectomy (procedure)|[9], a surgical procedure with the same surgical approach as |Single-port laparoscopic cholecystectomy (procedure)|. We therefore propose that the content of this <Surgical approach> hierarchy can be adapted into SNOMED CT as values for the |Access (attribute)| relationship (|Surgical access values (qualifier value)|, as shown in examples for |Laparoendoscopic single-site total hysterectomy (procedure)| and |Single-port laparoscopic cholecystectomy (procedure)|, represented using SNOMED CT's compositional grammar (**Annex 8**).

With the formal definition stabilised, one can then resort to the definitional template-like structure referred to in Section 2 – and explored within the EndoTerm project – to optimise the drafting of the natural language definition. Thus, a template is put forward with a customised version of a categorial structure (available in **Annex 9**), which will allow a more effective and consistent drafting process. If the order of the different elements of the template is followed, a natural language definition proposal for |Laparoendoscopic single-site total hysterectomy (procedure)| could be *Procedure consisting of the excision of the uterus and cervix using a laparoscope to access the site via a minimally invasive surgical approach with a single skin incision*. Following this approach, only a minor, indispensable change would be required for |Single-port laparoscopic cholecystectomy (procedure)|: *Procedure consisting of the excision of the* ***gallbladder structure*** *using a laparoscope to access the site via a minimally invasive surgical approach with a single skin incision*.

5. Concluding remarks

As seen in Section 3, concepts are defined in three different ways in SNOMED CT – the FSN, the formal concept definition, and the textual definition, when available – and ideally, the three should align. However, the preceding example has demonstrated that the formal concept definition may sometimes lack one (or more) delimiting characteristics, therefore hindering that concept's ability to differentiate itself from other concepts and determining the quality of a potential textual definition. Other analysed examples (not included here due to space limitation) show ambiguity in the FSN, as well as textual definitions containing information that is nowhere to be found in the formal definition.

Given the growing trend in incorporating textual definitions into biomedical terminological resources (as is the case with MeSH and ICD-11), on the one hand, and the importance that such natural language definitions have for translators and, potentially, for other user groups (namely patients[10]), on the other hand, there are a number of questions that need to be addressed by SNOMED International if this matter becomes a priority, namely: 1) the scope of such definitions, i.e. whether these textual definitions would match the formal definitions entirely or they would go beyond them; 2) what inclusion

---

7 The project is described in detail in [14].

8 https://www.nlm.nih.gov/research/umls/index.html

9 This concept is currently inexistent in SNOMED CT, but this type of minimally invasive surgical procedure is becoming increasingly prevalent in several gynaecological conditions [41].

10 From November 2022, patients started having access to their General Practitioner health record through the NHS App and other online services. In England, all GP systems capture clinical terms using SNOMED CT. (https://digital.nhs.uk/services/nhs-app/nhs-app-guidance-for-gp-practices/guidance-on-nhs-app-features/accelerating-patient-access-to-their-record/update-from-nhs-england)

criteria would be followed if extra information was to be added; 3) if textual definitions would be limited to fully defined concepts, or if they would also comprise primitive concepts (concerning the latter, how would a consistent textual definition be drafted if the formal definition is not yet stable?); 4) who would draft such definitions and based on which workflow – the guidelines put forward by [42] within the scope of the OBO Foundry might provide a relevant starting point.

Ultimately, this paper aimed to show the added value of the synergies involving Terminology – with its double-dimensional nature – and ontologies in what concerns knowledge representation, organisation, and sharing. Based on the example of EndoTerm, we aimed to illustrate how a solid conceptual framework allows for logical consistency in drafting natural language definitions, while also contributing towards more effective communication within a multilingual setting, as is the case with SNOMED CT translation work. In addition, it paves the way to interoperability, also at the linguistic level, if Linked Data is leveraged, which is currently being explored. Still part of our ongoing work and an essential aspect to be further developed is how to automate (part of) the definition drafting process.

6. Acknowledgements

Part of this research is supported by Portuguese national funding through the FCT – Portuguese Foundation for Science and Technology, I.P. within the scope of the project UIDB/LIN/03213/2020 and UIDP/LIN/03213/2020 – Linguistics Research Centre of NOVA University Lisbon (CLUNL).

We would like to thank Dr Marie-Alexandra Lambot, Assistant Medical Director at the Centre Hospitalier Universitaire Saint-Pierre, Brussels (CHU Saint-Pierre) and coordinator of the Community Support for Clinical Terminologies (CSCT), for providing the SNOMED CT example analysed in this paper.

7. References

[1] Moss, R. J. et al. “eHealth and mHealth.” European journal of hospital pharmacy: science and practice vol. 26,1, 2019: 57-58. doi:10.1136/ejhpharm-2018-001819

[2] World Health Organization. mHealth: New Horizons for Health through Mobile Technologies, Global Observatory for eHealth series - Volume 3, WHO, Geneva, 2011.

[3] Kashani, M. H. et al. “A systematic review of IoT in healthcare: Applications, techniques, and trends.” Journal of Network and Computer Applications, vol. 192, 2021: 103-164. https://doi.org/10.1016/j.jnca.2021.103164

[4] Selvaraj, S., Sundaravaradhan, S. Challenges and opportunities in IoT healthcare systems: a systematic review. SN Appl. Sci. 2, 139, 2020. https://doi.org/10.1007/s42452-019-1925-y

[5] Supriya, M., Deepa, AJ. Machine learning approach on healthcare big data: a review. Big Data and Information Analytics, 5(1), 2020: 58-75. doi: 10.3934/bdia.2020005

[6] Ngiam, K. Y., Khor, W. I. “Big data and machine learning algorithms for health-care delivery.” The Lancet. Oncology vol. 20,5, 2019: e262-e273. doi:10.1016/S1470-2045(19)30149-4

[7] Healthcare Information and Management Systems Society (HIMSS), Interoperability in healthcare, 2022. URL: https://www.himss.org/resources/interoperability-healthcare

[8] DigitalHealth Europe Consortium. DigitalHealthEurope recommendations on the European Health Data Space, DHE, 2021.

[9] Berners-Lee, T., Hendler, J., & Lassila, O. “The Semantic Web. A New Form of Web Content That Is Meaningful to Computers Will Unleash a Revolution of New Possibilities”. Scientific American, 284, 2001:1-5.

[10] Heath, T. and Bizer, C. *Linked Data: Evolving the Web into a Global Data Space* (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool, 2011.

[11] Gruber, T. “A translation approach to portable ontology specifications”, Knowledge Acquisition, Volume 5, Issue 2, 1993: 199-220, https://doi.org/10.1006/knac.1993.1008.

[12] Costa, R. "Plurality of theoretical approaches to terminology". *Modern approaches to terminological theories and applications*, Peter Lang Verlag, 2006: 79-89.

[13] Santos, C. and Costa, R. "Domain Specificity: Semasiological and Onomasiological Knowledge Representation". Handbook of Terminology. Ed. Hendrik J. Kockaert and Frieda Steurs. Amsterdam / Philadelphia: John Benjamins, 2015: 153-179.

[14] Carvalho, S. A terminological approach to knowledge organization within the scope of endometriosis: the EndoTerm project. PhD, Univ. NOVA de Lisboa - FCSH / Communauté Université Grenoble Alpes, 2018.

[15] Smith, R. "Aristotle's Logic". Stanford Encyclopedia of Philosophy, 2022. URL: https://plato.stanford.edu/entries/aristotle-logic/

[16] ISO. ISO 1087:2019. *Terminology work and terminology science — Vocabulary*, Geneva, 2019.

[17] ISO. ISO 704:2022. *Terminology work — Principles and methods*, Geneva, 2022.

[18] Felber, H. and UNESCO. *Terminology manual*. United Nations Educational, Scientific and Cultural Organization: International Information Centre for Terminology, Paris, 1984.

[19] Sager, J. *A Practical Course in Terminology Processing*. Amsterdam, John Benjamins, 1990.

[20] Suonuuti, H. *Guide to Terminology*. 2nd ed., Sanastokeskus, 2001.

[21] Löckinger, G., Kockaert, H. and Budin, G. "Intensional definitions". Handbook of Terminology. Ed. Hendrik J. Kockaert and Frieda Steurs. Amsterdam / Philadelphia: John Benjamins, 2015: 60-81.

[22] Seppälä, S. "An Ontological Framework for Modeling the Contents of Definitions", *Terminology*, vol. 21, no. 1, 2015: 23-50.

[23] Kageura, K. "Multifaceted/Multidimensional Concept Systems", Handbook of Terminology Management: Volume 1: Basic Aspects of Terminology Management. Ed. Sue Ellen Wright and Gerhard Budin, 1997: 119-132.

[24] Martin, W. "Frames as definition models for terms". In *Proceedings of the International Conference on Professional Communication and Knowledge Transfer. Vol .2.*, 1998: 189-220. Vienna: TermNet.

[25] Faber, P. (ed.). *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin, Boston: De Gruyter Mouton, 2012.

[26] Faber, P. "Frames as a Framework for Terminology". Handbook of Terminology. Ed. Hendrik J. Kockaert and Frieda Steurs. Amsterdam/Philadelphia: John Benjamins, 2015: 14-33.

[27] García de Quesada, M. et al. "Propuesta de estructura definicional terminográfica en OntoTerm®", Terminology, vol. 8, no. 1, 2002: 57-90.

[28] Faber, P. et al. "EcoLexicon: New Features and Challenges", GLOBALEX 2016: Lexicographic Resources for Human Language Technology in conjunction with the 10th edition of the Language Resources and Evaluation Conference, 2016: 73-80.

[29] Durán-Muñoz, I. "Producing frame-based definitions". *Terminology,* 22:2, 2016: 223–249.

[30] SNOMED International. *SNOMED CT Starter Guide*, 2022. URL: https://confluence.ihtsdotools.org/display/DOCSTART

[31] SNOMED International. *SNOMED CT Editorial Guide*, 2022. URL: https://confluence.ihtsdotools.org/display/DOCEG

[32] Elhanan, G. et al. "A survey of SNOMED CT direct users, 2010: impressions and preferences regarding content and quality." *Journal of the American Medical Informatics Association: JAMIA* vol. 18 Suppl 1, 2011: i36-44. doi:10.1136/amiajnl-2011-000341

[33] Ma, Y. and Distel, F. "Learning Formal Definitions for Snomed CT from Text". In: Peek, N., Marín Morales, R., Peleg, M. (eds) Artificial Intelligence in Medicine. AIME 2013. Lecture Notes in Computer Science, vol 7885. Springer, Berlin, Heidelberg, 2013. https://doi.org/10.1007/978-3-642-38326-7_11

[34] Petrova, A. et al. "Formalizing biomedical concepts from textual definitions." *Journal of biomedical semantics* vol. 6, 22, 2 Apr, 2015, doi:10.1186/s13326-015-0015-3

[35] Prokhorov, V. et al. "Generating Knowledge Graph Paths from Textual Definitions using Sequence-to-Sequence Models". In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,*

Volume 1 (Long and Short Papers), 2019: 1968–1976, Minneapolis, Minnesota. Association for Computational Linguistics.

[36] Jiang, G., Solbrig, H. and Chute, C. "Using Semantic Web technology to support icd-11 textual definitions authoring" Journal of Biomedical Semantics, vol. 4, no. 1, 2013. doi:10.1186/2041-1480-4-11

[37] *Liang, S. F. et al. "OntoVerbal: a Generic Tool and Practical Application to SNOMED CT." ArXiv abs/1312.2798, 2013.*

[38] Lyudovyk, O. and Weng, C. "SNOMEDtxt: Natural Language Generation from SNOMED Ontology." *Studies in health technology and informatics* vol. 264, 2019: 1263-1267. doi:10.3233/SHTI190429

[39] Carvalho, S., Costa, R. and Roche, C. "LESS Can Indeed Be More: Linguistic and Conceptual Challenges in the Age of Interoperability". *12th International conference on Terminology and Knowledge Engineering*, Copenhagen Business School, Jun 2016, Copenhagen, Denmark, 2016. URL: https://hal.archives-ouvertes.fr/hal-01354331

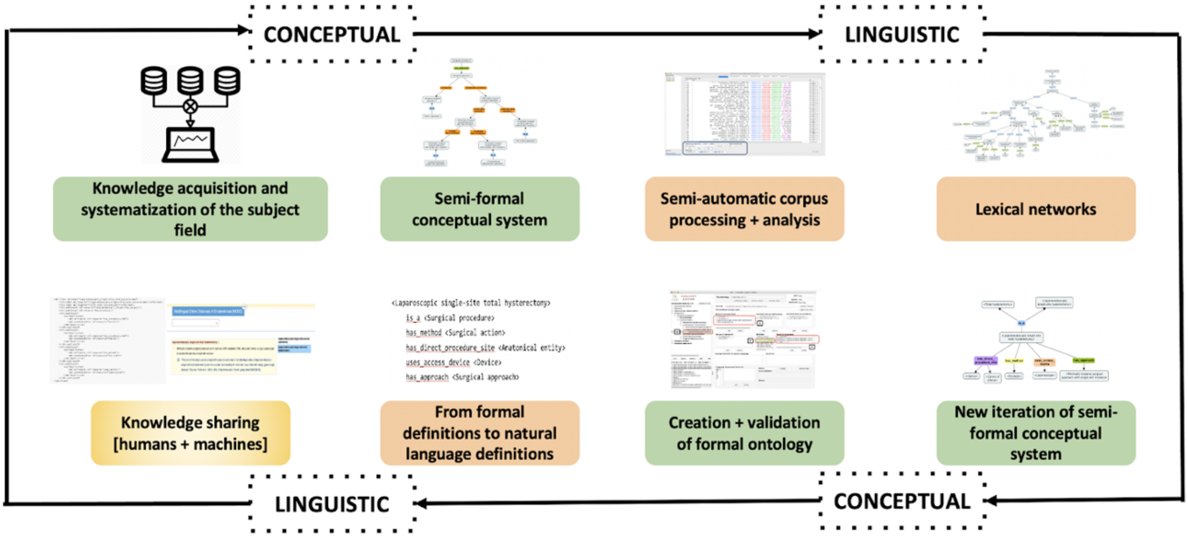
[40] ISO. ISO 1828:2012. *Health informatics — Categorial structure for terminological systems of surgical procedures*, Geneva, 2012.

[41] Lin, Y. et al. "Laparoendoscopic single-site surgery compared with conventional laparoscopic surgery for benign ovarian masses: a systematic review and meta-analysis." *BMJ open* vol. 10,2 e032331. 16 Feb. 2020, doi:10.1136/bmjopen-2019-032331

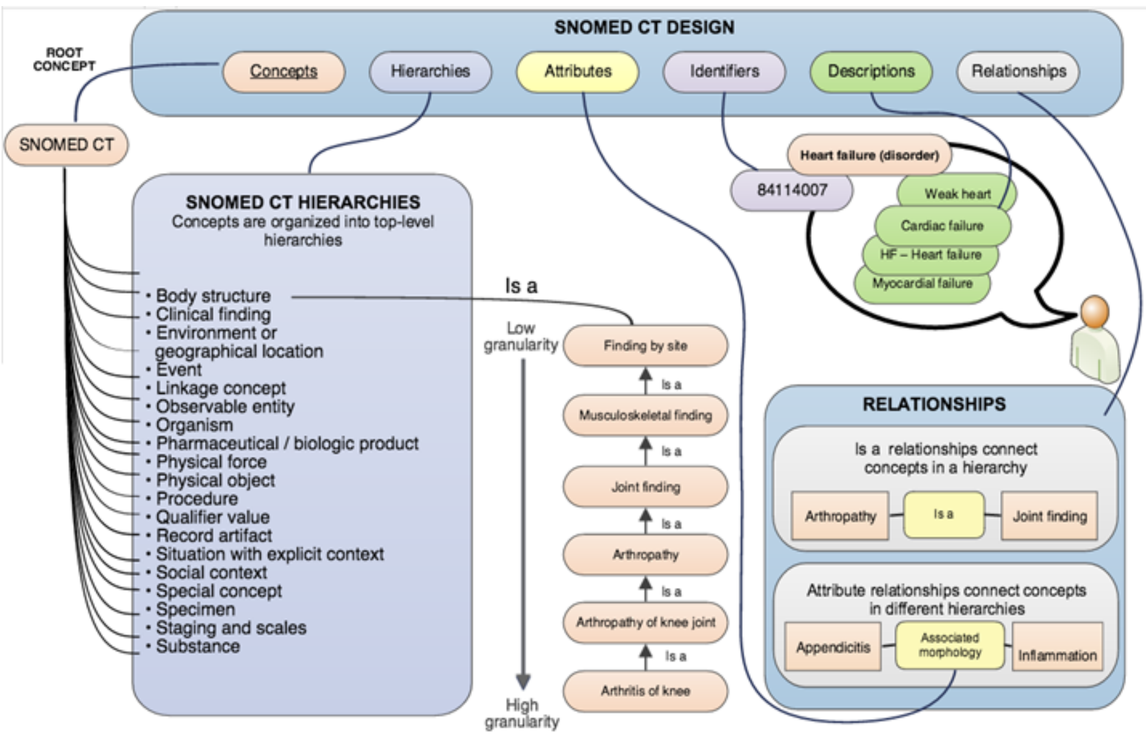
[42] Seppälä, S., Ruttenberg, A. and Smith, B. "Guidelines for Writing Definitions in Ontologies". *Ciência Da Informação*, vol. 46, nº 1, dezembro de 2017, doi:10.18225/ci.inf. v46i1.4015.

8. Annexes

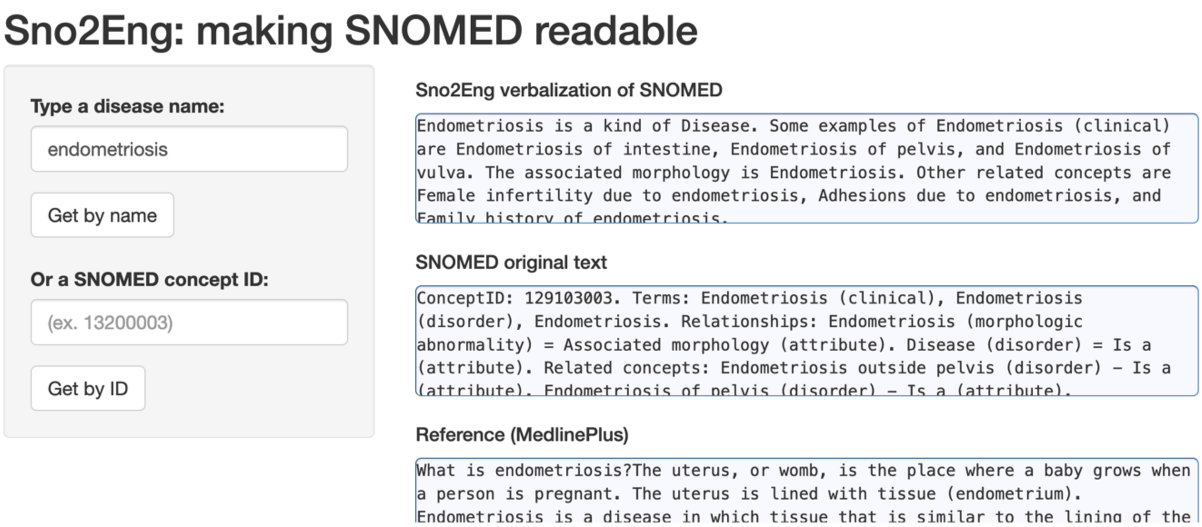
Annex 1: Methodological framework of the EndoTerm project [14]



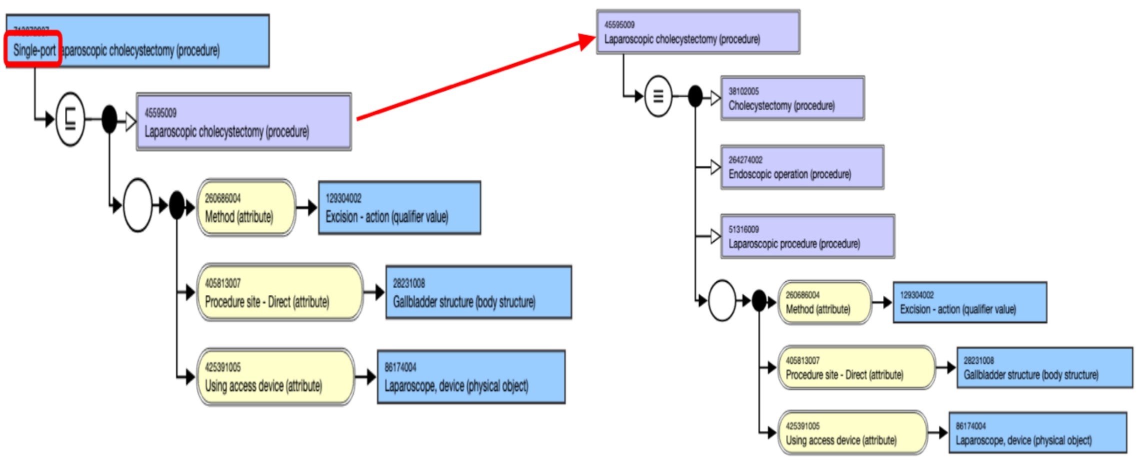
Annex 2: SNOMED CT's design [31]



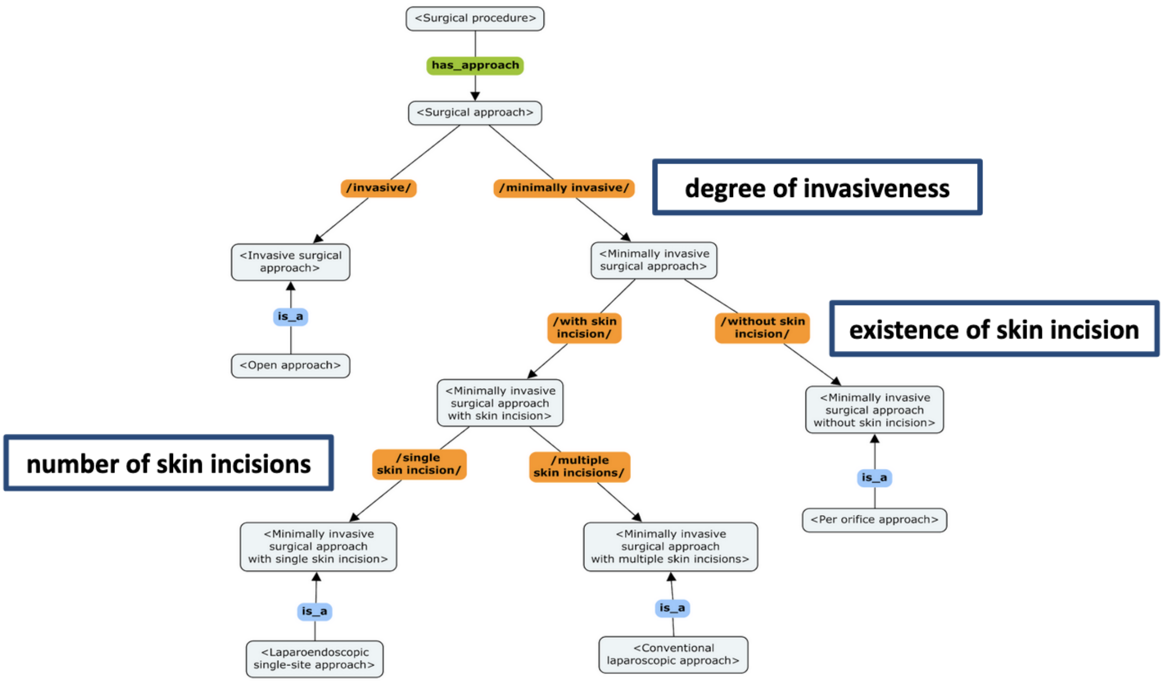
Annex 3: SNO2Eng – example with ‘endometriosis’ [https://github.com/WengLab-InformaticsResearch/SNOMEDtxt]



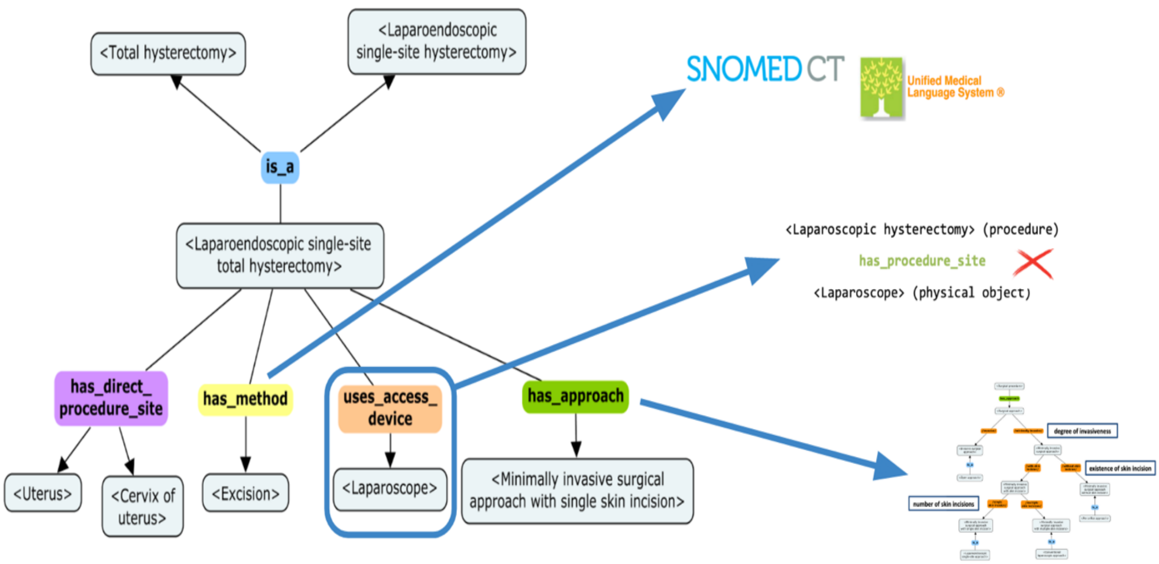
Annex 4: Example of the formal definitions of |Single-port laparoscopic cholecystectomy (procedure)| and of |Laparoscopic cholecystectomy (procedure)| (https://browser.ihtsdotools.org/)



Annex 5: EndoTerm's <Surgical approach> hierarchy [14]



Annex 6: EndoTerm's micro-concept system for <Laparoendoscopic single-site total hysterectomy> [14]



Annex 7: EndoTerm's formal definition of <Laparoendoscopic single-site total hysterectomy> in RDF/OWL [14]

```xml
<owl:Class rdf:about="Laparoendoscopic_single-site_total_hysterectomy">
    <rdfs:label xml:lang="en">"laparoendoscopic_single-site_total_hysterectomy"</rdfs:label>
    <rdfs:label xml:lang="en">"LESS_total_hysterectomy"</rdfs:label>
    <rdfs:subClassOf rdf:resource="Laparoendoscopic_single-site_surgery"/>
    <rdfs:subClassOf rdf:resource="Hysterectomy"/>
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="has_procedure_site"/>
            <owl:someValuesFrom rdf:resource="Cervix"/>
        </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="has_procedure_site"/>
            <owl:someValuesFrom rdf:resource="Uterus"/>
        </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="has_method"/>
            <owl:someValuesFrom rdf:resource="Excision"/>
        </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="uses_device"/>
            <owl:someValuesFrom rdf:resource="Laparoscope"/>
        </owl:Restriction>
    </rdfs:subClassOf>
</owl:Class>
```

Annex 8: Example of SNOMED CT's formal concept definitions for |Laparoendoscopic single-site total hysterectomy (procedure)| and |Single-port laparoscopic cholecystectomy (procedure)|using compositional grammar

```
000000000 |Laparoendoscopic single-site total hysterectomy (procedure)|

=== 236887006 |Laparoscopic hysterectomy (procedure)| +
  116140006 |Total hysterectomy (procedure)| :
    { 260686004 |Method (attribute)| = 129304002 |Excision - action (qualifier value)|,
     405813007 |Procedure site - Direct (attribute)| = 181452004 |Entire uterus (body structure)|,
     425391005 |Using access device (attribute)| = 86174004 |Laparoscope, device (physical object)|,
     260507000 |Access (attribute| =000000000 |Minimally invasive surgical approach with single skin incision - access (qualifier value)| }

713872007 |Single-port laparoscopic cholecystectomy (procedure)|

=== 45595009 |Laparoscopic cholecystectomy (procedure)| :
    { 260686004 |Method (attribute)| = 129304002 |Excision - action (qualifier value)|,
     405813007 |Procedure site - Direct (attribute)| = 28231008 |Gallbladder structure (body structure)|,
     425391005 |Using access device (attribute)| = 86174004 |Laparoscope, device (physical object)|,
     260507000 |Access (attribute| = 000000000 |Minimally invasive surgical approach with single skin incision - access (qualifier value)| }
```

Annex 9: Definitional template for the concepts under analysis

CONCEPT	
is_a	|Procedure (procedure)|
|Method (attribute)|	|Surgical action (qualifier value)|
|Procedure site - Direct (attribute)|	|Body structure (body structure)|
|Using access device (attribute)|	|Physical object (physical object)|
|Access (attribute)|	|Surgical access values (qualifier value)|