

Towards principles of ontology-based annotation of clinical narratives

Stefan Schulz^{1,2,*}, Warren Del-Pinto³, Lifeng Han³, Markus Kreuzthaler¹, Sareh Aghaei¹ and Goran Nenadic³

¹*Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria*

²*Averbis GmbH, Freiburg, Germany*

³*Department of Computer Science, University of Manchester, UK*

Abstract

Despite the increasing availability of ontology-based semantic resources for biomedical content representation, large amounts of clinical data are in narrative form only. Therefore, many clinical information management tasks require to unlock this information using natural language processing (NLP). Clinical corpora annotated by humans are crucial resources. On the one hand, they are needed to train and domain-fine-tune language models with the goal to transform information from unstructured free text into an interoperable form. On the other hand, manually annotated corpora are indispensable for assessing the results of information extraction using NLP. Annotation quality is crucial. Therefore, detailed annotation guidelines are needed to define the form that extracted information should take, to prevent human annotators from making erratic annotation decisions and to guarantee a good inter-annotator agreement. Our hypothesis is that, to this end, human annotations (and subsequently machine annotations learned from human annotations) should (i) be based on ontological principles, and (ii) be consistent with existing clinical documentation standards. With the experience of several annotation projects, we highlight the need for sophisticated guidelines. We formulate a set of abstract principles on which such guidelines should be based, followed by examples of how to keep them, on the one hand, user-friendly and consistent, and on the other hand compatible with the international semantic standards SNOMED CT and FHIR, including their areas of overlap. We sketch the representation of the resulting representations in a knowledge graph as a state-of-the-art semantic representation paradigm, which can be enriched by additional content on A-Box and T-Box levels and on which symbolic and neural reasoning tasks can be applied.

Keywords

Formal Ontologies, Clinical Information Models, Natural Language Processing, Text Annotation Guidelines, Electronic Health Records

Proceedings of the International Conference on Biomedical Ontologies 2023, August 28th-September 1st, 2023, Brasilia, Brazil

*Corresponding author

✉ stefan.schulz@medunigraz.at (S. Schulz); warren.del-pinto@manchester.ac.uk (W. Del-Pinto);

lifeng.han@manchester.ac.uk (L. Han); markus.kreuzthaler@medunigraz.at (M. Kreuzthaler);


sareh.aghaei-dinani@medunigraz.at (S. Aghaei); gnenadic@manchester.ac.uk (G. Nenadic)

ORCID 0000-0001-7222-3287 (S. Schulz); 0000-0003-3307-9432 (W. Del-Pinto); 0000-0002-3221-2185 (L. Han);

0000-0001-9824-9004 (M. Kreuzthaler); 0000-0002-0511-095X (S. Aghaei); 0000-0003-0795-5363 (G. Nenadic)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

1. Introduction

A seamless and effective flow of clinical information is vital for high-quality healthcare and health management. Thus, information must be stored in a way that supports effective communication, search, and analysis. However, most content of electronic health records (EHRs) consists of unstructured text in documents and free-text database fields. As a result, crucial data that require key insights into populations and individuals remain inaccessible without additional processing.

In contrast, the good news regarding EHR interoperability is the increasing support by elaborated semantic standards, *viz.* terminologies, ontologies (e.g., SNOMED CT, LOINC) [1] and information models (e.g., FHIR [2]), which enjoy increasing international adoption. Interoperable clinical data representations would ideally bridge syntactically different but semantically equivalent expressions in different degrees of structure, from data collected in standardised forms to clinical narratives, across disciplines, jurisdictions, and natural languages.

To achieve this goal for narrative data, text passages must be linked to identifiers from a controlled vocabulary, a process referred to as semantic annotation or tagging. Such vocabularies are typically rooted in semantic resources such as those mentioned above. An additional step is the assertion of links between tags, known as relation annotation. To do this manually is resource-intensive and difficult in practice; human annotators need to be trained and monitored, and diverging annotations must be reconciled by subsequent adjudication.

Nevertheless, annotated text corpora are indispensable as a “fuel” for training, domain-fine-tuning, and evaluation of natural language processing (NLP) models. Fig. 1 shows an example of a manual annotation of a clinical text passage.

Annotation guidelines play a vital role in this process. They target consistency and uniformity by providing clear instructions, reducing variability, and ensuring that annotations are standardised across annotators and projects. The inter-annotator agreement serves as a measure of the effectiveness and quality of guided annotation processes. Annotation guidelines provide instructions on, and examples of, handling ambiguous cases, addressing recurring challenges, and resolving annotation discrepancies.

An ideal annotation should produce, with the same input text, the same target representation created by different trained annotators. The same should be achieved based on different paraphrases of the same input or with its translation into a different language. Even where annotations differ, semantic equivalence should be stated based on logical reasoning. Although these desiderata will probably never be completely fulfilled, they justify more effort spent on principled annotation guidelines.

Several guidelines for annotating clinical narratives have been proposed. Examples include the Clinical E-Science Framework (CLEF) [3, 4] which focuses on clinical information extraction tasks, annotation guidelines for the identification of personal health information such as patient names [5] and a range of guidelines that are very specific in terms of language and task, e.g. lung diseases in Japanese [6], family history in Norwegian [7], diseases in Spanish [8], disorders, findings, drugs and body structures in Swedish and Chinese [9, 10]. Guidelines may also specify how to annotate temporal references [11, 12], negation statements [13], and other contextual markers.

Although the recognition of diagnostic statements has been a preferred goal in clinical annotation tasks, the need for a broader coverage of annotation guidelines has been recognised

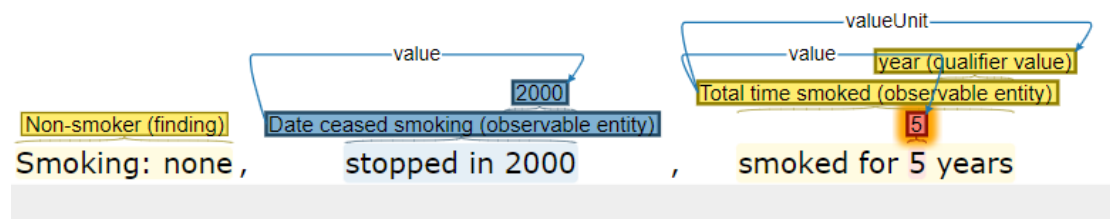


Figure 1: Semantic annotation example of a passage from a discharge summary

by Luo et al. [14], who extended the scope of annotation to “medical problems, treatments and tests”. These examples show that previous works on annotating clinical text have often produced guidelines that were specific in scope, constrained to a particular semantic type or designed for a very use-case-specific data set.

Although the argument that specific use cases necessitate specific annotation strategies is valid, we postulate a need for more general principles that provide a systematic approach to clinical text processing. Such an approach should be based on ontological principles [15, 16] and international semantic standards [17]: a prerequisite to ensuring that data from different sources can be meaningfully compared. Once such principles have been formulated, task-specific guidelines could be instantiated without unnecessary repetition of previous efforts.

In this work, we propose such general annotation principles. Regarding the annotation vocabulary, we focus on SNOMED CT [18], an ontology-based clinical terminology system, and FHIR [2], a set of information templates for recurring clinical documentation tasks. Both are considered leading international semantic standards for healthcare data. We separate ontological aspects – the description of classes of biomedical entities and their properties, which is the domain of SNOMED CT – from contextual information about individuals, which are represented as instances of FHIR resources. This addresses A. Rector’s claim to distinguish “models of meaning” (ontologies) and “models of use” (information models) [19]. Applied to clinical text annotation, this means that FHIR resources provide semantically explicit templates to capture instance-level information on epistemic, temporal, and provenance aspects belonging to a given patient and his/her documentation context. In contrast, SNOMED CT codes represent types of clinical entities¹, whose instances are referred to by an appropriate field in the FHIR template. It is well known that SNOMED CT also supports epistemic and temporal aspects in its *Situation in specific context* hierarchy, whereas FHIR resources include references to HL7 value sets, mostly for roles and qualities, which compete with SNOMED CT *Qualifier value* concepts. This overlap constitutes a complicating factor of clinical information management in general and clinical text annotations in particular. Efforts to principally handle this overlap² have remained unfinished.

¹AKA concepts or T-box entities, modelled as OWL classes.

²TermInfo Project. (This and the following footnotes: click to navigate to the respective project website)

2. Methods

We employed a qualitative and heuristic approach to develop annotation guidelines for standards-aware representation of clinical documents. The goal was to bootstrap annotation rules from sample narratives, inspect the results, discuss disagreements between annotators, identify recurring patterns and structures, and consolidate a final guideline after a series of iterations. The authors have gained experience in numerous annotation activities. These include the use of SNOMED CT for the annotation of clinical summaries in Brazilian Portuguese [20] and for clinical text snippets in five European languages [21], and the application of a set of manually designed annotation rules to TAC2017³ data to supervise the extraction of adverse drug reactions and related entities. SNOMED CT has also been used to provide the semantics for normalising clinical event mentions in diagnostic statements extracted from hospital data in the UK.

The problems of representing context using SNOMED CT and FHIR [22] led the authors to pilot annotation tasks using the German corpus GRASCCO⁴ and a clinical corpus derived from UK hospital data by using FHIR, including HL7 value sets, as a framework in which SNOMED CT is used for the ontological information proper. Soon, it became necessary to formulate rules to manage the overlap between both systems. In addition, it became clear that the handling of the complex structure of FHIR was difficult for the human annotators.

Guideline-based text annotations are also in the centre of the ongoing projects AIDAVA⁵ (Horizon Europe), the German annotation initiative GemTeX⁶, as well as the UK projects JIGSAW⁷ and HIPS⁸. The latter projects required the formulation of annotation guidelines to create diagnosis representations in SNOMED CT, using the *Situation in specific context* hierarchy, which proposes precoordinated content and post-coordination patterns for factuality (e.g. “suspected asthma”), temporality (e.g. “history of major depression”) and family history (“mother died from breast cancer”).

While FHIR is compatible with multiple terminologies, the use of SNOMED CT was motivated in part by its widespread adoption, e.g., by the requirement that all systems of the UK National Health Service (NHS) must use SNOMED CT as a core terminology⁹. Additionally, given the required annotation tasks, the broad coverage of SNOMED CT hierarchies, which include clinical findings (disorders), pharmaceutical products, procedures, and others, was deemed beneficial.

The continuous crafting and revision of annotation guidelines, supported by the outcomes of annotator training sessions, resulted in the creation of a set of annotation principles as formulated in the next section, followed by a case study that serves as an instantiation of these principles, centring on the rooting of annotations into the standards SNOMED CT and FHIR.

³Adverse Drug Reaction Extraction from Drug Labels

⁴Graz Synthetic Clinical Text Corpus

⁵AI-powered Data Curation & Publishing Virtual Assistant

⁶German Medical Text Corpus

⁷Assembling the Data Jigsaw

⁸Healthcare Impact Partnership: Integrating hospital outpatient letters into the healthcare data space

⁹NHS Digital, SCCI0034

3. Results

3.1. Proposed Annotation Principles

Different use cases require different annotation guidelines. For example, secondary uses of clinical data such as epidemiological research have different requirements than data representation for providing direct patient care or data support for billing purposes. Therefore, we first propose a set of general principles from which more specific annotation guidelines for given applications can be derived. This first set of annotation principles is independent of the annotation vocabularies used, i.e. the ontologies and information models (or subsets thereof) used.

- Annotation is limited to the *semantic* aspects of narratives. It comprises the assignment of codes for unary predicates (types, but also individuals) from a domain ontology, together with literals such as decimal numbers to spans of text. These annotations are additionally connected by binary predicates (relations). Both unary and binary predicates are rooted in a semantic reference such as a domain ontology or a clinical information model.
- The endpoint is a canonical form of representing clinical narrative information as a primary knowledge graph (KG), with subject-predicate-object triples describing individual patients and related clinical entities. This primary KG should then be transferable, by applying supporting rules and resources, into a knowledge graph that follows the structures of the underlying standards and is committed to Applied Ontology principles. Such a KG makes a clear T-Box/A-Box distinction, i.e. between nodes that point to entity types as given by an ontology and those that point to individual entities that instantiate the types provided by the annotation. The difference is relevant, because, e.g., “asthma” should point to an individual entity (particular) in an affirmative context (“The patient has asthma”), but to an entity type (concept) in a negative context (“no signs of asthma”). Another ontological distinction is the one between information entities [23] (instances of FHIR resources) and the clinical entities they are about.
- This complexity needs to be hidden from annotators. Predicate types should be restricted to the necessary. E.g., the relation between “tumour” and “right” is ontologically very different from the one between “tumour” and “suspected”. Nevertheless, a literal annotation (by using predicates such as “laterality” and “verificationStatus” in the same way) would suffice for the annotators, provided that annotation post-processing is sufficiently specified by transformation rules.
- Scope and granularity are determined by the underlying annotation vocabulary. For restricted annotation tasks, subsets of the maximum vocabulary are provided to the annotators. For example, a targeted cancer annotation task might not require fine-grained reference to unrelated diseases like heart disorders or injuries. This could motivate the pruning of sub-hierarchies, e.g., underneath *Coronary disease* or *Fracture of bone*.
- Annotations are descriptive and not interpretative: annotators annotate only what they read, without complicating their task by seeking individual interpretations. For example, the annotations “fever” after “hip replacement” are only linked with a predicate for causality if there is a causality statement in the text. Two exceptions to this rule are highlighted: (i) word sense disambiguation, as long as meaning can be derived from the context; and

- (ii) co-reference, as long as the antecedent to which an anaphoric reference points back is identifiable.
- The granularity of annotation spans is not given by a named entity recognition step prior to annotation, which would yield entity types, such as *Disorder* and *Body part* as annotations of the spans “fracture of skull” or “left ventricle”. Instead, the spans are determined by the underlying annotation vocabulary. The principle of longest match is followed and pre-coordinated expressions are used as preferred if they correspond to a contiguous span. Otherwise, e.g. in “the skull exhibited the sign of a fracture”, shorter text spans (“skull”, “fracture”) are annotated and linked afterwards. If the representation of the meaning of a span requires more than one code, a conjunction (logical AND) is preferred over a construction that additionally requires binary predicates.
 - The annotation vocabulary is distinguished between *core* content and *supporting* content. The former one characterises the central focus of the given annotation task, as defined by the intended use case, while the latter one provides additional, mostly refining information to the core. For example, for the identification of diagnosis statements, core content would be given by the concepts of the Clinical Finding hierarchy of SNOMED CT, such as diseases. Supporting concepts would be those under *Body Structure*, which specify a particular *Clinical finding* concept, as well as those that capture factuality, such as *Probably present*.

The following annotation principles explicitly refer to the use of SNOMED CT and HL7 FHIR as annotation vocabularies.

- “Core” hierarchies as introduced above are *Clinical finding*, *Event*, *Observable entity*, *Pharmaceutical / biologic product*, *Procedure*, *Specimen* in SNOMED CT. They have a high proportion of fully defined concepts, expressed by OWL ‘equivalentTo’ axioms, in which concepts from “non-core” hierarchies such as *Substance*, *Organism*, *Body structure* and *Physical object* are referred to by existentially quantified links. Ambiguous annotations are addressed by preference rules, e.g. to prefer C_1 over C_2 , if C_1 belongs to a core hierarchy. For example, SNOMED CT offers different codes for “Sarcoma”, viz. *Sarcoma (disorder)* and for *Sarcoma (morphological abnormality)*. The former is preferred for annotation of clinical texts because it is fully defined and axiomatically implies the latter.
- The hierarchy *Situation in specific context* – although it would formally correspond to a core hierarchy – is not used for annotation because FHIR has been shown to be more granular, actively maintained, and frequently used to represent the context of clinical statements.
- A set of binary predicates with their own namespace “anno:” was introduced for close-to-user relation annotation. These binary predicates were grounded in (i) SNOMED CT object properties or chains thereof, (ii) relational chains of FHIR elements, or (iii) both. For example, the predicate **site** between a SNOMED CT clinical finding and a body structure, is mapped to the linkage concept (relation) ‘**finding site**’ as well as the concatenation of the inverse of the FHIR element **Condition.code** with **Condition.body**, cf. Table 2.
- SNOMED CT mappings to HL7 value sets are proposed, e.g. *hl7:Recurrence* to *sct:Recurrent* or *hl7:Confirmed* to ‘*sct:Confirmed present*’ (see Table 2).

Guidelines designed for a specific annotation task can be viewed as an instantiation of the general principles outlined above. It is then necessary to define sets of permitted codes for both the core and supporting concepts to be drawn from, for example, SNOMED CT reference sets. To summarise, the steps in utilising such a guideline to annotate clinical text are as follows:

- Identify a core concept mentioned explicitly in the document.
- Assign to the identified phrase a suitable concept from the set of possible core concepts.
- If present, identify phrases corresponding to supporting concepts that refine the core concept.
- Assign to each identified supporting concept a suitable SNOMED CT concept, drawn from the set of possible supporting concepts.
- For each of the supporting concepts, link them directly to the core concept and identify the type of relation.

Given a clinical document, the above steps can be repeated until all of the relevant clinical mentions have been captured. The final annotation shall be a semantic representation of the explicit meaning of the original clinical text.

3.2. Examples

Here, we provide examples of applying these general principles to produce task-specific templates for annotating clinical text.

3.2.1. Example from JIGSAW/HIPS

The JIGSAW and HIPS projects include the task to extract diagnosis information from both semi-structured lists and free-text narratives in outpatient letters for secondary use purposes such as specifying patient cohorts for epidemiological studies. The annotation task was decided to be uniform across both projects to ensure consistent capture of relevant information. The combination of SNOMED CT with FHIR naturally resulted from the NHS use of SNOMED CT as its preferred terminology, and the need to represent and communicate instance-level information about patients. Following the principles outlined above, core and supporting concepts were identified. As a source for core concepts all concepts in the SNOMED CT *Clinical Finding* hierarchy were taken. The supporting concepts were primarily taken by the elements of the FHIR Condition resource¹⁰, with additional relations from the SNOMED CT concept model including *Associated morphology* and *Causative agent*. In order to keep the annotation task consistent and simple, the initial assignment of codes to diagnostic statements in the text has been entirely done using concepts from the SNOMED CT *Clinical finding* hierarchy. The linking of supporting concepts to their corresponding core concepts follows the strategy outlined in the previous section: a set of binary predicates were specified for relation annotation, to avoid the need for annotators to be familiar with all SNOMED CT “linkage concepts”¹¹ and FHIR elements. In this context, given the narrative phrase “osteoarthritis of the spine”, the FHIR Condition resource specifies that

¹⁰FHIR Condition

¹¹Corresponding to OWL object and datatype properties.

a *Clinical finding* code should be provided for the condition being diagnosed and, if necessary, a body part can also be provided in the form of a SNOMED CT *Body structure* code. Therefore, *Osteoarthritis* is a core concept that is refined by the SNOMED CT *Clinical finding* concept *Osteoarthritis*. Meanwhile, “spine” acts as supporting information. It is annotated using the *Body Structure* concept *Joint structure of spine*. As a result, it is necessary to perform post-processing both to map the provided SNOMED CT codes to appropriate FHIR values and to map the relations to either SNOMED CT object properties or FHIR elements where applicable.

For FHIR elements with value sets that are specified as SNOMED CT codes by default, such as **Condition.bodySite**, no mapping was required. For those that specify an alternative FHIR value set, such as **Condition.verificationStatus** and **Condition.clinicalStatus**, mappings were specified between appropriate SNOMED CT concepts and the FHIR values. These mappings were based upon discussion with clinicians regarding which information is needed for their use case, such as the need to specify uncertainty of diagnoses via SNOMED CT concepts such as *Probably present*, and how these should be interpreted in FHIR, for example *Provisional*.

3.2.2. Example from AIDAVA

Regarding the annotation of composed expressions, AIDAVA puts more emphasis on the use of pre-coordinated SNOMED CT content. For the contiguous span “osteoarthritis of the spine” preference would be given to the single concept *Spondylosis*, whereas a passage such as “osteoarthritis of knees, hips, spine” would require a post-coordinated expression such as exemplified in 3.2.1. The important is here that due to the formal axioms in SNOMED CT, equivalence can be stated between pre-coordinated content and post-coordinated expressions.

AIDAVA also required mappings between SNOMED CT and FHIR as shown in Table 1. In order to support, alternatively, queries on SNOMED CT and FHIR, more attention was paid to the interoperability of predications, i.e. the use of binary predicates and their anchoring in both standards. In general, predications are straightforward at a text level but complex at an ontology-based representation level. The phenomenon that representations may compete, e.g. SNOMED CT only vs. SNOMED CT in FHIR contexts had to be addressed, cf. Table 2. Figure 2 demonstrates the generation of an ontology-based knowledge graph, as intended by AIDAVA, consisting of SNOMED CT concepts, instances thereof, and literals like dates and identifiers, emerging from guideline-driven annotations of clinical narratives. It shows how, according to the annotation predicates chosen, different FHIR resources, viz. *Condition* and *FamilyMemberHistory* are instantiated. It also shows how the choice of the predicate **anno:verificationStatus** related to *sct: Suspected* leads to a reference to the concept '*sct: Neoplasm of breast*', whereas the annotation without modifier (which assumes that the diagnosis is confirmed) the same FHIR relation points to an *individual referent* that instantiates '*sct: Neoplasm of breast*'. Other details are only hinted at, such as the class *Information content entity* from the Information Artifact Ontology¹², as well as the inferred relation '**occurs in**' from the Basic Formal Ontology (BFO)[24].

¹²Information Artifact Ontology (IAO)

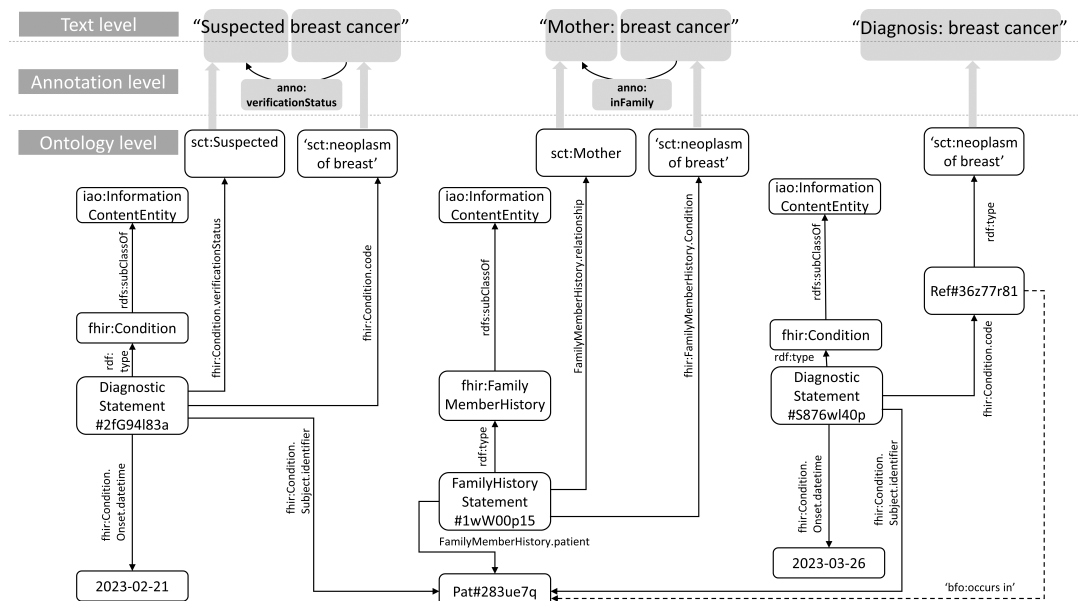


Figure 2: From text to ontology-based knowledge graphs. The text level shows three text snippets (quoted) belonging to clinical texts about one patient. The human annotations (grey background) use SNOMED CT and “anno:” as annotation vocabularies. The lower, most complex level shows the instantiations of FHIR resources with references to SNOMED concepts, instances thereof, and literals. Nodes with random IDs represent individuals, linked to the concept whose code appears in the annotation. An example of an inferred predication is shown by a dashed arrow.

Table 1

Examples for mappings between HL7 FHIR values and SNOMED CT concepts

HL7 FHIR	SNOMED CT (from the <i>Qualifier value</i> hierarchy)
<i>Unconfirmed</i>	<i>Suspected OR Probably not present</i>
<i>Provisional</i>	<i>Probably present OR Suspected</i>
<i>Confirmed</i>	<i>Confirmed present</i>
<i>Refuted</i>	<i>Known absent</i>

4. Discussion

Past clinical annotation projects were often based on UMLS CUIs, as freely accessible but semantically often shallow concept identifiers, or on in-house annotation languages which restrict their openness such as in [10]. In other cases, annotations were limited to high-level entity types, such as *Disorder* and *Body part*, with a focus on relations. Many different ontologies were used, among which the Human Phenotype Ontology (HPO) or the Disease ontology (DO) should be emphasised.

The reason why we focus on SNOMED CT is its growing international acceptance as a standard for all health record content, its scope and granularity, and particularly its logical under-fitting, which facilitates the bridging between pre-coordinated and post-coordinated expressions.

However, our approach also have some reservations. It has been argued that SNOMED CT is little used in routine, particularly in continental Europe, that current licenses exclude important jurisdictions, and that translations are still missing. We reply that the status quo in clinical terminologies, with national ICD versions, national procedure classifications and drug catalogues, does not offer a convincing interoperability perspective without SNOMED CT. Furthermore, SNOMED CT is fully available to the research community, so the fact that its productive use requires a licence, should not be an obstacle.

One specific limitation is the still unresolved management of the overlap between SNOMED CT, FHIR, and related value sets. A further continuation of the Terminfo work in the light of FHIR would be desirable. Another limitation is that numerous SNOMED CT concepts lack formal and textual definitions, and pose challenges to annotators, particularly with texts in languages for which no official translation exists.

We are convinced that in times when large language models are skyrocketing, and under the hypothesis that machine understanding of clinical language is a realistic goal, semantic standards do not become obsolete. On the contrary, large language model technology has to be leveraged to generate canonical, standardised representations. Such representations as a gold standard for clinical content representation need to be elaborated and refined. We understand the proposed annotation principles as a step in this direction.

5. Conclusion and Future Work

Clinical annotation guidelines are crucial for structured data preparation, content representation for human reading and searching, enhancing supervised learning and for benchmarking language models via gold-standard data sets. Clinical annotation tasks become even more sophisticated due to the diversity of the text and broad coverage of knowledge across multiple dimensions, such as diseases, signs, symptoms, findings, procedures, as well as temporality, factuality, and other

Table 2

Example binary predicates for relation annotations with their translation into SNOMED CT and FHIR syntax. “INV” = inverse, “||” = concatenation. The relation paths marked with [a] refer to the concatenation of SNOMED CT relations, those marked with [b] to roughly equivalent FHIR elements

anno:	Domain	Target path	Range
site	‘sct:Clinical finding’	[a]‘sct:Finding site’ [b] INV(fhir:Condition.code) fhir:Condition.body	‘sct:Body structure’
site	‘sct:Procedure’	[a]‘sct:Procedure - Direct’ [b] INV(fhir:Procedure.code) fhir:Procedure.body	‘sct:Body structure’
inFamily	‘sct:Clinical finding’	[b] INV(fhir:FamilyMemberHistory.condition) fhir:FamilyMemberHistory.relationship [a] INV(‘sct:Associated finding’) ‘sct:Subject relationship context’	‘sct:Person’
verification status	‘sct:Clinical finding’	[b] INV(fhir:Condition.code) fhir:Condition.verificationStatus [a] INV(‘sct:Associated finding’) ‘sct:Finding context’	Qualifier value’ (cf. Tab. 1)

contexts. Existing annotation guidelines have been mostly motivated by shared task organisers to solve NLP challenges such as entity recognition and relation extraction, which tended to lead to shallow policies. With the purpose to address clinical annotations from an ontology-driven methodology and to take advantage of the rich content of SNOMED CT and FHIR, we proposed a set of annotation principles and designed the mapping of annotations into SNOMED CT and FHIR via entity and relation linking, as well as expression normalisation. The full design and implementation of our annotation principles cover a broad and in-depth semantic representation of the original clinical text, for which we recommend a representation as knowledge graphs, which can then be enriched by additional content on A-Box and T-Box level and on which both symbolic and neural reasoning tasks can be applied. For practical applications, we do suggest the users carry out the core annotation first and choose the level of depths of annotations according to their needs.

We understand that clinical free text annotation is a huge and challenging task, and the goal of achieving our full guideline principles might take a long journey. But there is indeed such a need to unify the clinical annotation tasks so that it can facilitate clinical research across different sectors, and languages, as well as for current large NLP model training.

In future work, we will prepare more annotation examples with a full set of instructions. We also plan to carry out the evaluation of our principles from the NLP perspective, in addition to case studies for measuring the inter-rater agreement [25] levels via trained workers.

6. Acknowledgments

The work was partially supported by Grant 101057062 “AIDAVA” (funder: the European Commission, HORIZON-HLTH-2021), Grant “Assembling the Data Jigsaw: Powering Robust Research on the Causes, Determinants and Outcomes of MSK Disease” (The project has been funded by the Nuffield Foundation, but the views expressed are those of the authors and not necessarily the Foundation. Visit www.nuffieldfoundation.org) and Grant EP/V047949/1 “Integrating hospital outpatient letters into the healthcare data space” (funder: UKRI/EPSRC).

References

- [1] O. Bodenreider, R. Cornet, D. J. Vreeman, Recent developments in clinical terminologies—SNOMED CT, LOINC, and RxNorm, *Yearb Med Inform* 27 (2018) 129–139.
- [2] D. Bender, K. Sartipi, HL7 FHIR: An agile and RESTful approach to healthcare information exchange, in: *Proc 26th IEEE Symp on Computer-based Med Syst*, 2013, pp. 326–331.
- [3] A. Roberts, R. Gaizauskas, M. Hepple, et al., The CLEF corpus: semantic annotation of clinical text, in: *AMIA Annu Symp Proc*, 2007, pp. 625–9.
- [4] A. Roberts, R. Gaizauskas, M. Hepple, et al., Building a semantically annotated corpus of clinical texts, *J Biomed Inform* 42 (2009) 950–966.
- [5] B. R. South, D. Mowery, Y. Suo, et al., Evaluating the effects of machine pre-annotation and an interactive annotation interface on manual de-identification of clinical text, *J Biomed Inform* 50 (2014) 162–172.

- [6] S. Yada, A. Joh, R. Tanaka, et al., Towards a versatile medical-annotation guideline feasible without heavy medical knowledge: starting from critical lung diseases, in: Proc of the 12th LREC, 2020, pp. 4565–4572.
- [7] T. Rama, P. Brekke, Ø. Nytrø, L. Øvreliid, Iterative development of family history annotation guidelines using a synthetic corpus of clinical text, in: Proc of the 9th Intl Workshop on Health Text Mining and Information Analysis, ACL, 2018, pp. 111–121.
- [8] A. Miranda-Escalada, L. Gascó, S. Lima-López, et al., Overview of DisTEMIST at BioASQ: Automatic detection and normalization of diseases from clinical texts: results, methods, evaluation and multilingual resources 13390 (2022).
- [9] M. Skeppstedt, M. Kvist, G. H. Nilsson, et al., Automatic recognition of disorders, findings, pharmaceuticals and body structures from clinical text, *J Biomed Inform* 49 (2014) 148–158.
- [10] E. Zhu, Q. Sheng, H. Yang, et al., A unified framework of medical information annotation and extraction for chinese clinical text, *Artif Intell Med* (2023) 102573.
- [11] W. F. Styler IV, S. Bethard, S. Finan, et al., Temporal annotation in the clinical domain, *Transactions of the ACL* 2 (2014) 143–154.
- [12] D. L. Mowery, H. Harkema, W. Chapman, Temporal annotation of clinical text, in: Proc of the Workshop on Current Trends in Biomedical NLP, 2008, pp. 106–107.
- [13] M. Marimon, J. Vivaldi, N. Bel Rafecas, Annotation of negation in the IULA Spanish corpus, *SemBEaR –Computational Semantics Beyond Events and Roles* (2017) 43–52.
- [14] Y.-F. Luo, W. Sun, A. Rumshisky, MCN: a comprehensive corpus for medical concept normalization, *J Biomed Inform* 92 (2019) 103132.
- [15] N. Guarino, Formal ontology, conceptual analysis and knowledge representation, *Int Journal of Human-computer Studies* 43 (1995) 625–640.
- [16] B. Smith, M. Ashburner, C. Rosse, et al., The OBO foundry: coordinated evolution of ontologies to support biomedical data integration, *Nature biotech* 25 (2007) 1251–1255.
- [17] S. Schulz, R. Stegwee, C. Chronaki, Standards in healthcare data, in: K. et al. (Ed.), *Fundamentals of Clinical Data Science*, Springer, Cham(CH), 2019.
- [18] J. Millar, The need for a global language–SNOMED CT introduction, *Stud Health Technol Inform* 225 (2016) 683–685.
- [19] A. L. Rector, R. Qamar, T. Marley, Binding ontologies and coding systems to electronic health records and messages, *Applied Ontology* 4 (2009) 51–69.
- [20] E. J. Pacheco, *MorphoMap: Mapeamento automático de narrativas clínicas para uma terminologia médica*, PhD dissertation. UTFPR, Brazil, 2009.
- [21] J. A. Miñarro-Giménez, R. Cornet, M.-C. Jaulent, et al., Quantitative analysis of manual annotation of clinical text samples, *Int J Med Inform* 123 (2019) 37–48.
- [22] M. Ayaz, M. F. Pasha, M. Y. Alzahrani, R. Budiarto, D. Stiawan, The Fast Health Interoperability Resources (FHIR) standard, *JMIR Med Inform* 9 (2021) e21929.
- [23] E. M. Sanfilippo, Ontologies for information entities: State of the art and open challenges, *Appl Ontology* 16 (2021) 111–135.
- [24] J. N. Otte, J. Beverley, A. Ruttenberg, BFO: Basic Formal Ontology, *Appl Ontology* 17 (2022) 17–43.
- [25] S. Gladkoff, L. Han, G. Nenadic, Student’s t-distribution: On measuring the inter-rater reliability when the observations are scarce, in: Proc of the RANLP, 2023.