The AIDAVA Annotation Guide (AAG)

Instructions for semantic annotations of clinical narratives based on SNOMED CT and FHIR

Authors / Contributors:

Stefan Schulz[1,2], Sareh Aghaei[1], Akhila Naz Kuppassery[1], Alexander Beger[1], Daniel Dür[1], Larissa Hammer[1], Kristian Kankainen[3], Heba Alloch[1], Jan Lauritsch[1], Jennifer Rode[1], Natthanaphop Isaradech[4], Catalina Martínez-Costa[5], Andrea Riedel[6], Luise Modersohn[7], Christina Lohr[8], Goran Nenadic[9], Warren Del-Pinto[9], Lifeng Han[9], Markus Kreuzthaler[1]

[1]Medical University of Graz, Austria, [2]Averbis GmbH, Freiburg, Germany, [3]North Estonia Medical Centre and Tallinn University of Technology, Estonia, [4]Chiang Mai University, Chiang Mai, Thailand, [5]University of Murcia, Spain, [6]University of Erlangen, Germany, [7]Technical University of Munich, Germany, [8]University of Leipzig, Germany, [9]University of Manchester, UK

Live document - Last modifications 2026-09-07

Guests are welcome! Feel free to drop comments. E-mail contact: stefan.schulz@medunigraz.at

This site is currently subject to frequent modifications, which may affect the numbering of subsections.

New: Annotation Cheat Sheet [pptx][pdf]

1. Introduction

Large amounts of clinical information are only available as narratives in electronic health records (EHRs), characterised by unstandardized language with jargon expressions, short forms, spelling variants, errors, and typos. Notwithstanding, international standards such as SNOMED CT, LOINC, and FHIR promise interoperable and computable representations of clinical content. This includes the linkage to units of clinical language (words, word parts, multiword expressions) and, in consequence, the domain entities they denote. Important goals for standards-based interoperable systems have been addressed for a while, but they have only partially been met, even in advanced clinical computing environments:

1. Univocal standardised representations of a given portion of clinical reality
2. Identification of syntactically different but semantically identical or similar representations
3. Transformation of narrative content into such representations

Bridging between human language and semantic standards requires technology and resources in natural language processing (NLP), from rules and dictionaries to deep learning and large language models. They need narrative data to be trained, fine-tuned, and benchmarked. Text collections that are semantically annotated by human annotators are therefore an essential resource. They constitute the "fuel" for models that reliably convert the content of narratives into interoperable expressions rooted in standavards. Both HL7 and SNOMED International explicitly recommend the use of the SNOMED CT and FHIR and work on interoperability issues in regular meetings. A third standard, LOINC is currently being harmonised with SNOMED CT and will be accessible using SNOMED CT codes.

High-quality human annotations should approximate the following goals:

1. With the same input text, different human annotators produce the same target representation.
2. With different paraphrases of (1.), different human annotators produce target representations for which semantic equivalence between (1.) and (2) can be inferred.
3. With the translation of (1.) and (2.) to different human languages, different human annotators produce target representations for which semantic equivalence could be stated.

This is the rationale of our effort to propose a set of annotation rules within this set of annotation instructions, known as this Annotation Guide (AAG), developed within the AIDAVA project. AAG is guided by overarching principles, supported by examples. They enforce that the annotation result conforms with the building principles set of by standardisation organisations like SNOMED International and HL7[1]. In particular this means:

1. The result of annotation is given by an annotation graph.
2. Its nodes consist of SNOMED CT codes (some of them mapped to HL7 value set elements)
3. Its edges are provided by a set of user-friendly binary predicates, which are introduced in this guideline as the namespace "anno:".
4. All "anno:" predicates are rooted in relations or relational patterns that refer to SNOMED CT and FHIR

[1] A draft paper on general annotation principles for ontology-aware annotation was published at ICBO 2023 in Brasilia together with the Manchester NLP group gnTEAM.

2. Background

2.1. Annotation strategies

Strategies for annotation of narratives have been diverse, multifaceted, and only partly comparable, regarding:

- **Which types of annotations are performed**: nded
  - Syntactic: tokens, part of speech, chunks, syntactic relations
  - Semantic: text spans assigned to codes for concepts or concept types, semantic relations/predicates
  - Discourse: coreference, document type, document section type
- **Whether semantic annotations use high-level semantic types (i) or the whole depth (ii) of an ontology**.
  - In a "shallow" annotation strategy (i), "Open fracture of left femur" could be annotated:
    i. Open fracture as Condition
    ii. Left as Quality
    iii. Femur as Body part
  - This is contrasted a "deep" annotation, which could result in:
    i. Open fracture as 397181002 |Open fracture (disorder)|
    ii. Left as 7771000 |Left (qualifier value)|
    iii. Femur as 421235005 |Structure of femur (body structure)|

- **How text spans that are candidates for annotation are delineated**: either are they delineated by a named entity recognition process prior to annotation, or by the annotators themselves. In the latter case, annotators are guided by an existing annotation vocabulary, where the longest matching span is given preference. E.g., "Open fracture of left thumb" might have already been delineated by an entity recognition task as one of
  - "Open fracture" [Condition] "of" "left" [Quality] "thumb" [Body part]
  - "Open fracture of left thumb" [Condition]
  - "Open [Quality] "fracture" [Condition] "of" "left thumb" [Body part]

  Whatever the output of an automated entity recognizer, it is a process that is difficult to control, but it relieves annotators of the burden of finding appropriate boundaries, allowing them to focus on finding the most appropriate concepts for the already recognized words or passages.
  If, instead, delineation is left to the annotators - **the strategy proposed here** - it must follow clear rules, such as defining the span according to the longest matching term in the terminology. A span like "the left elbow shows exposed bones" or "left elbow with open fracture" would be considered a single span because it semantically matches the SNOMED CT concept "10820261000119101 |Open fracture of left elbow (disorder)".
- **Whether only words, word sequences or word parts are annotated, or also relations (binary predicates) between them**. In the latter case, their source needs to be specified. A case where there is no single (pre-coordinated) code for the whole expression is the following: "Open fracture of left thumb". One solution is annotating "open fracture" with 397181002 |Open fracture (disorder)| and "left thumb" with "734143007 |Structure of left thumb (body structure)|". However, the exact meaning requires linking both annotations by a predicate annotation. This could be 363698007 |Finding site (attribute)|, taken from SNOMED CT (<<106237007 |Linkage concept (linkage concept)|), but also the FHIR slot *Condition.bodySite*, or a binary predicate from another source such as 'has-location' from the OBO Relation ontology. **AAG proposes its own set of predicates**, which is user-friendly on the one hand and fully compatible with SNOMED CT and FHIR on the other hand
- **To which extent the values for span and predicate annotations are constrained**. They could be taken from a complete ontology or from use-case-specific subsets. E.g., in an annotation task focusing on cancer, such a vocabulary might constrain the granularity of the annotation of only marginally related comorbidities. "Open fracture of left elbow" could then only be annotated as "fracture of bone" [sct:125605004], if the subclasses of the latter were excluded from the annotation vocabulary. Or a subset could only contain atomic concepts such as 397181002 |Open fracture (disorder)|, 7771000 |Left (qualifier value)|, and 421235005 |Structure of femur (body structure)|, but not the pre-coordinated

concept “10820261000119101 |Open fracture of left elbow (disorder)”. In this case, shorter text elements must be annotated.

- **Whether annotation spans are tolerated to span over unrelated text**. Assuming that there is no code for “open fracture of left femur”, but a code for “open fracture of femur” and for “left”. When annotating the whole phrase with the most precise code, it spans over “left”, which must be annotated separately. This entails that annotations overlap.
- **Whether the choice of the concept or predicate takes local context into account**, or whether the annotation is done literally, even if the annotator knows from the previous text that the referent of a more specific type
    - In “After the procedure, the patient was instructed to avoid…”, with “procedure” referring to a more specific concept introduced before, e.g. “tonsillectomy”.
    - In “Since the patient started taking anticonvulsants, no seizure has occurred”, the onset of the medication and the absence of seizures are temporally related. Annotating it with a predicate that expresses a causal relationship could be seen as an overinterpretation.
- **How polysemy should be represented and distinguished from composed meanings**. Regarding ambiguous annotations, it is normally expected that the annotator performs the disambiguation as long as it is clear from the context which readings can be ruled out.
- **How overlapping areas of the semantic resources used are dealt with**. This occurs with SNOMED CT and FHIR, which recommends HL7 values sets, e.g. for units of measurements or factuality modifiers. One solution is that everything that can be expressed by FHIR data structures and HL7 value sets should be used. A contrary position is to use SNOMED CT only and make use of the SNOMED context approach. Overlaps could also be addressed by mappings so that the downstream representation can go either way.

2.2. Related work

In contrast to previous works such as the Annotation guideline for ASSESS-CT, which are often confined to specific data or limited scopes, this **AIDAVA Annotation Guide (AAG)** offers a broader applicability and flexibility. It can be instantiated and applied to various clinical use cases, providing a versatile framework. For instance, it has been employed to annotate data pertaining to breast cancer and cardiovascular disease within the AIDAVA project. Its principles can be applied to the whole of SNOMED CT or confined to subsets thereof.

The guideline developed in [Medical/Clinical Text Annotation Guidelines] provides a specification for tagging different types of medical entities and the relations between them. For instance, it defines the XML tag <d></d> to indicate disease mentions within the documents, and incorporates the attribute 'certainly' with predefined values ('positive', 'negative', 'suspicious', and 'general') to tag the degree of the disease. However, compared to AAG, this work is limited to recognizing types of medical entities at a high level without considering further details. It lacks interoperability because clinical documentation standards such as SNOMED CT, LOINC or FHIR are not considered.

The work described in [https://ars.els-cdn.com/content/image/1-s2.0-S1532046419300504-mmc1.pdf] focuses primarily on annotating concepts, with no consideration of relations between them. This approach utilises RXNORM for annotating medication spans and selects the first-ranked result from the UTS SNOMED CT Browser to identify the optimal Concept Unique Identifier (CUI) for other recognized mentions. To ensure accuracy, it addresses abbreviations, misspellings, unclear or ambiguous concepts by employing Google for resolution. Furthermore, modifiers falling outside of the concept mention are disregarded, and a mention can be annotated with one or multiple annotations. In terms of preference, plural forms of concepts are prioritised over similar concepts in their singular form and other ambiguities between concepts are not discussed.

To normalise medical mentions, CUI, SNOMED CT, and RxNorm are utilised in [Research article MCN: A comprehensive corpus for medical concept normalisation]. In cases where a CUI is not available, annotators can first normalise the mention to any suitable concept in SNOMED CT. However, considering the incomplete coverage of medications in SNOMED CT, annotators are also allowed to employ RxNorm for normalisation. In case of compositional concepts, annotators are requested to use multiple concepts to represent that mention by splitting

the mention span into the largest mention span that can be normalised to a concept and the other smaller mention span. Moreover, singular forms of concepts are given priority over plural forms.

A method presents in [A method for encoding clinical datasets with SNOMED CT | BMC Medical Informatics and Decision Making | Full Text] for encoding clinical relational databases with SNOMED CT, including three main parts: (1) identifying potential data items in a given database, (2) cleaning data items, and (3) encoding the cleaned data items. Compared to the AIDAVA Annotation Guide (AAG), there are inevitably several ambiguities between two concepts that remain unsolved in this method, such as observable entities and diseases. Additionally, numbers, dates, and measurements are often ignored, despite their potential as terms for annotation. Also, the relations between concepts are not addressed.

Similar to AIDAVA Annotation Guide (AAG), the situation with explicit context hierarchy is not used in the proposed annotation process in [https://www.sciencedirect.com/science/article/pii/S1319157815000919] and the default context is assumed for all the concepts. However, the coding rules are limited to only pre-coordinated concepts, and new custom concepts are recommended for terms that do not match existing ones. Moreover, the prioritised hierarchies are only discussed for clinical findings, observable entities, and procedures (i.e., if there are matches in more than one hierarchy, the highest priority is given to clinical findings, followed by observable entities and procedures in that order). However, we propose intensive rules regarding different hierarchies to avoid ambiguities as much as possible (as discussed in Table X). Additionally, the scope of use cases where the proposed coding rules can be applied is quite limited.

Table [Literature Review] presents a comparison between the related works described and the AAG, focusing on the underlying reference terminology, annotation tool, and input data source.

Reference Vocabulary	**Annotation Tool**	**Data**	**Source**
SNOMED CT RxNORM	Multi-document Annotation Environment	Clinical discharge summaries from the MIMIC database	https://ars.els-cdn.com/content/image/1-s2.0-S1532046419300504-mmc1.pdf
SCT ONLY UMLS EXT LOCAL	Excel	Clinical text snippets provided by ASSESS CT project	Annotation guideline for ASSESS-CT
-	Tools support XML tag format[2]	Medical concepts including disease, illness, and body regions as well as the relations between the concepts	Medical/Clinical Text Annotation Guidelines
-	Tools support JSON format[3]	Clinical categories including anatomical structure, body function, body measurement, laboratory, medical condition, medical device, medical procedure, medication, substance abuse, and patient status	https://github.com/google/healthcare-text-annotation
SNOMED CT	-	palliative care dataset	https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-10-53
SNOMED CT	-	diabetes diagnosis CBR systems	https://www.sciencedirect.com/science/article/pii/S1319157815000919
CUI SNOMED CT RxNorm	MAE MAE2	discharge summaries	https://www.sciencedirect.com/science/article/pii/S1532046419300504

[2] The name of the applied tool has not been reported in the source.

[3] The name of the applied tool has not been reported in the source.

- MIMIC-IV on FHIR: converting a decade of in-patient data into an exchangeable, interoperable format
- Work specific on SNOMED CT:
    - Lee DH, Lau FY, Quan H. A method for encoding clinical datasets with SNOMED CT. BMC Med Inform Decis Mak. 2010;10(1):53. doi:10.1186/1472-6947-10-53
    - Lau FY, Simkus R, Lee D. A Methodology for Encoding Problem Lists with SNOMED CT in General Practice. In: Proceedings of the Third International Conference on Knowledge Representation in Medicine, Phoenix, Arizona, USA, May 31st - June 2nd, 2008. ; 2008. http://ceur-ws.org/Vol-410/Paper17.pdf
    - Randorff Højen A, Rosenbeck Gøeg K. SNOMED CT Implementation: Mapping Guidelines Facilitating Reuse of Data. Methods Inf Med. 2012;51(06):529-538. doi:10.3414/ME11-02-0023
    - El-Sappagh S, Elmogy M. An encoding methodology for medical knowledge using SNOMED CT ontology. J King Saud Univ - Comput Inf Sci. 2016;28(3):311-329.
- Work specific on FHIR
- Only type and relation annotations: Lohr C et al. Evolutionary Approach to the Annotation of Discharge Summaries. Stud Health Technol Inform. 2020, 16;270:28-32.

3. Objectives

- To propose general principles for ontology-based clinical document annotation, based on SNOMED CT and FHIR (possibly enhanced by some upper-level ontology), precise enough to achieve that, as much as possible, that two independent annotators agree in their annotation results and that the annotation corresponds to a knowledge graph that represents the content of the text in an ontological manner, based on standards.
- To create and iteratively enhance concrete annotation rules based on these principles. This implies that this annotation guide is incrementally refined, using clinical documents from several sources in several languages.
- To downstream it to concrete use cases, particularly texts from the ASSESS-CT corpus, the GRASSCO corpus, AIDAVA use cases, from the GeMTeX Methods Platform
- To build webinars and educational materials base on the AAG
- To achieve convergence with other guidelines for similar purposes
- To submit it to a final assessment with inter-annotator agreement as an endpoint.

4. Tools and resources

4.1. Annotation Tool

The AIDAVA Annotation Guide (AAG) is not committed to a specific tool, but recommends one that supports online cooperation and, pre-defined vocabularies and relation (directed links, binary predicates) annotations. This functionality is, e.g. supported by INCEpTION.

4.2. Annotation vocabularies

4.2.1. SNOMED CT

The AIDAVA Annotation Guide (AAG) uses SNOMED CT as annotation vocabulary for concepts and recommends the SNOMED CT browser to find active codes of the most recent release. The decision which code to select should be made according to

- The wording of the Fully Specified Name (FSN) of a concept, in one of the official languages of the international version (English or Spanish)
- The concept's text definition (if available)

- Its formal axioms, as well as its taxonomic ancestors and descendants

Annotators that are not native English speaker must pay attention that the wording of a fully specified name may be misleading due to different meaning of cognates. E.g., the English adjective “pathologic” such as in “23875004 |No pathologic diagnosis (finding)|” has a much narrower meaning than in some other European languages in which, e.g. “pathologisch” or “patológico” means “abnormal” in a general sense. In such cases, only the inspection of the parent concept - here “250537006 |Histopathology finding (finding)|” reveals that its scope is restricted to histopathology.

According to the recent SNOMED - LOINC agreement, interoperation between these two terminologies is aimed at. The plan is that every concept in LOINC (observables) will have an equivalent SNOMED CT concept (observable) by the end of 2023. This means that LOINC as an additional terminology does not need to be considered.
See also the new LOINC Ontology Browser. Until further notice, the AAG recommends the use of evaluation procedure concepts for lab values as suggested by SNOMED International.

4.2.2. HL-7 FHIR

For everything beyond terminology aspects proper, FHIR is used as a guiding framework. Clear rules are established according to which an annotation result can be expressed as the instantiation of a FHIR resource. FHIR Condition, Observation, and Procedure are the most relevant ones. Which additional resources are needed will depend on the documents. For HL7 value sets, selected SNOMED CT content is used, mapped to HL7 values in the background. Standard situations such as a confirmed diagnosis or surgery done are expressed by the SNOMED code without any FHIR “envelope”.

4.2.3. Value set for predicate annotations

The linkage between annotations requires binary predicates (relations). Both SNOMED CT and FHIR provide a rich inventory of predicates or relational expressions By introducing the “anno:” namespace, the AAG proposes a set of simplified “alias” predicates with user-friendly naming, together with their translation into relational expressions of the source standards.

4.2.4. Post-annotation processing and export

The adherence to the two standards is enforced by postprocessing of the annotations, which has to be specified in a separate document. This step is characterised by the addition of namespaces, the translation of the annotation predicates into relational expressions rooted in SNOMED CT and/or FHIR, as well as the inference of additional nodes. A detailed specification is pending

5. Basic assumptions and decisions

5.1. Annotation philosophy

SNOMED CT is a health care ontology. This means that its representational units (aka concepts, with codes, associated labels, definitions, and axioms) denote types of clinically relevant things from diseases over drugs and body parts to lab parameters etc. In contrast, FHIR specifies templates to represent individual patient-level information. The interpretation of a SNOMED CT code used in the annotation of some text span in a clinical document in the context of HL7 FHIR is the following:
The related SNOMED concept is referred to implicitly or explicitly by some FHIR instance. This FHIR instance describes the portion of reality referred to by the document during the episode of treatment and is related to the subject of record, i.e. the patient the document is about. The FHIR instance specifies whether an instance of the concept can be related to the subject of record, e.g. in the case of a confirmed diagnosis of a completed procedure. If not, it may refer to a situation of uncertainty or negation or to another individual such as a family member in the case of family history information.
According to FHIR, all contextual information at the instance level should be consistent with the value sets proposed by the FHIR specifications whereas all ontological information (referred to by “code” in a FHIR resource) is provided by an ontology such as SNOMED CT.
E.g., laterality, aetiology, or chronicity of a condition is ontological, as well as dose form and strength of a drug. The same is true for anatomical location of a surgical procedure. All this information should be expressed by SNOMED

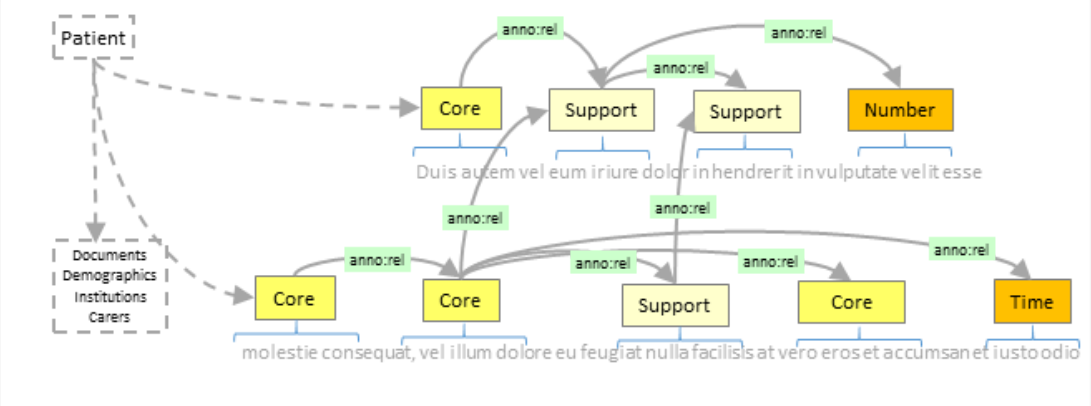
CT. In contrast, diagnostic certainty, subject relationship (patient or family), temporal contexts of conditions and procedures are contextual, as well as determinants of the provenance of such information. FHIR proposes existing HL7 value sets (e.g. with the value “differential” for diagnostic certainty). These values, however, largely overlap between FHIR and SNOMED. This is why both communities are currently working on mappings between FHIR (HL7) value set elements and SNOMED CT codes.
In order to avoid that annotators have to deal with different ontologies we propose the restriction to SNOMED CT concepts even in these cases where FHIR suggests HL7 values sets. In these cases SNOMED CT - HL7 mappings are maintained in the background.

5.2. General Annotation Principles

5.2.1. Methodology and scope of annotation

Regarding the options (Section 2.3), the following preferences are suggested for annotating clinical documents. Clinical documents are very diverse. A focus is laid on documents and document parts that describe clinical processes, instructions, findings and diagnoses. The AAG, just as SNOMED CT as a whole is, however not ideally prepared for representing in-depth descriptions of pathological structures (microscopic and macroscopic) as well as images and detailed surgical actions.
It must be emphasised that both the annotation vocabulary and the annotation instructions are complex. Good quality annotation requires time and resources. An intensive training phase is mandatory. Even if annotators have acquired a certain routine, the annotation guideline and the resulting annotation cheat sheet must be at hand. The scope of document-centred annotations described by this document are depicted in Fig. [schema]

Figure [Schema]: document-centred annotation schema

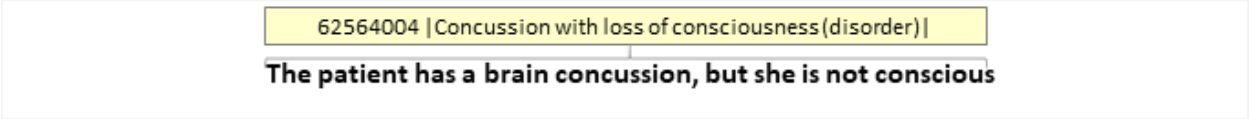


Out of scope regarding this annotation guide is the representation of patient metadata as visualised in grey. The coloured elements symbolise the annotations proper, split into span annotations and relation annotations. They form what is named annotation graphs. One of several annotation graphs represent a clinical document. The top node of each annotation graph must be an annotation by a concept under selected semantic types (core). Each core annotation should represent an entity directly related to the patient (who, by him/herself, is not annotated). Such core annotations are expected to represent self-standing pieces of information (e.g., a disease, a procedure), modified by others (body parts, qualities, numbers etc.), which are not meaningful, in isolation and are therefore considered as supportive.

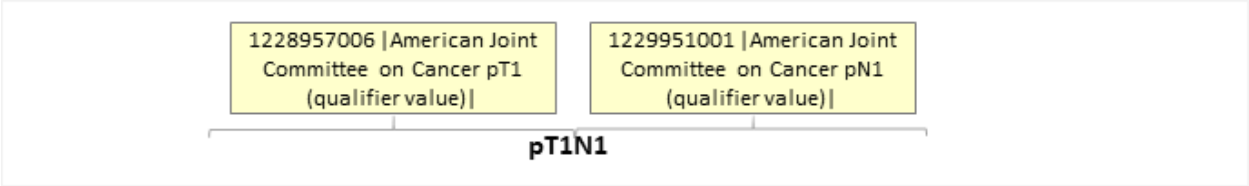
5.2.2. Annotation spans

All annotations are rooted in annotation spans, i.e. contiguous sequences of character, ranging from word parts over single words to multi-word expressions. The AIDAVA Annotation Guide (AAG) only describes semantic annotations and coreference annotations and not, e.g., annotations at a document or document section level or syntactic annotations. The breadth of the annotation span is given by the concept. Due to pre-coordination in

SNOMED CT, even complex expressions such as whole sentences can often be accurately annotated by a single code. The most specific SNOMED CT concept that captures the exact meaning of a span is always preferred. However, annotation spans must not cross sentence or paragraph boundaries.

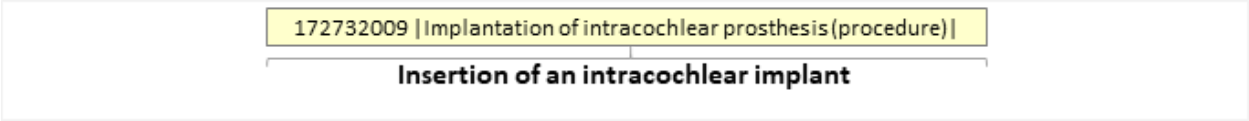


Annotations on a subword level are allowed wherever a word is clearly composed by parts that could alternatively be separated by spaces or hyphens (here between “pT1” and “N1”):

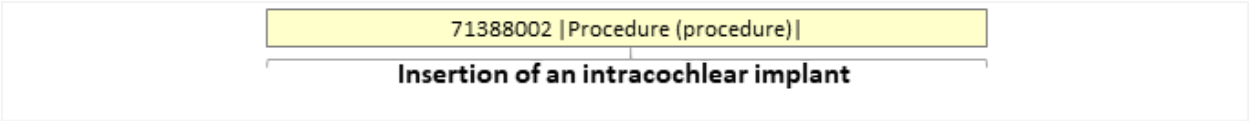


5.2.3. Annotation vocabulary

Annotations of text spans use the whole depth of the annotation vocabulary. This means that always the annotators choose the concept that comes closest to the passage to be annotated. The delineation of text passages follows the ontology. For instance,



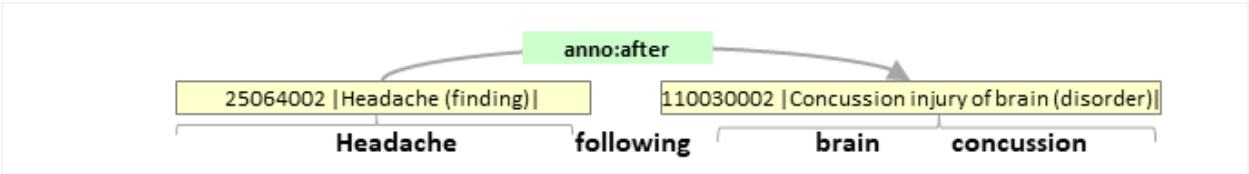
instead of



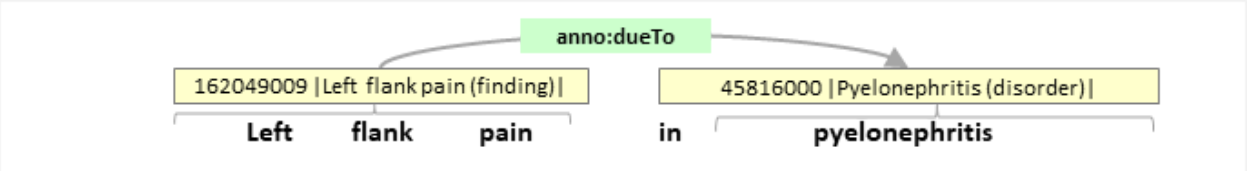
(which is easily inferable from the above, in case that only annotations at a high level are required for certain use cases).

5.2.4. Facts

Facts (or annotation triples) are constituted by the connection of annotated spans with a predicate (relation). They are annotated as long they can be unambiguously derived from the text without any additional interpretation.
This example suggests causality, but it is nevertheless annotated with a temporal predicate, because only a temporal relation is stated in the text.



This is a counterexample, in which the association between disease and symptom is sufficiently expressed by the preposition “in”



5.3. Annotation symbols

We distinguish between sources for concepts and sources for predicates. For concepts we use most of SNOMED CT with some restrictions as explained below. For predicates we use a closed set of relations derived from SNOMED CT and FHIR (namespace "anno"). The SNOMED annotations are characterised by the use of the typical SNOMED syntax that coordinates identifier and label using the pipe character.

5.3.1. Concepts

By "concepts" we understand units of non-relational meaning from a controlled vocabulary, an ontology, or a value set. The guideline document focuses on the use of SNOMED CT and FHIR, but could also be used for other vocabularies. Whether a concept in an annotation denotes a particular meaning or a universal meaning is not distinguished at the annotation level ("John Doe has asthma" vs. "John Doe is examined for asthma"), but is subject to downstream interpretations[4]. Top-level concepts in SNOMED CT (the heads of the SNOMED CT hierarchies) are also referred to by "Semantic Types".

SNOMED CT often presents ambiguities, i.e. two or more concepts with same or similar names. This requires guidance in terms of preference rules. A distinction between core concepts and non-core (supportive) concepts is therefore fundamental. The former ones are always to be preferred in case of doubt.

Core concepts are ideally fully defined. Even standing alone they express a clear meaning, e.g. that a patient has a disease or underwent some diagnostic or therapeutic intervention. They do infrequently modify other concepts, but are often modified, e.g. by qualifiers. Many core concepts are partly or fully defined using axioms with relations and restrictions as prescribed by the SNOMED concept model. Core concepts come typically from the hierarchies Clinical Conditions (SNOMED findings / disorders / events), Procedures, Observables, Staging and scales, Pharma products. Table [Core] provides an overview of the SNOMED CT hierarchies and their use for annotation.

Table [Priority] highlights typical cases for priority decisions between competing SNOMED CT concepts. Table [Core] gives an overview of SNOMED CT hierarchies, split into "Core", "Non-core" and "Not-to-use". Note the difference between SNOMED CT hierarchies and SNOMED CT semantic tags - the expressions that follow a SNOMED CT Fully Specified Name (FSN). Whereas the top concepts of the hierarchies often have the same name as a general semantic tag, in many hierarchies more specific semantic tags are used the more specific concepts are chosen[5].

However, these preferences apply once the span to be annotated by a single concept, has always been determined. Capturing the meaning of one expression by a single annotation has always priority:

166830008 |Serum cholesterol above reference range (finding)| would therefore always be preferred over the combination of
412808005 |Serum total cholesterol measurement (procedure)|
with
281302008 |Above reference range (qualifier value)|

[4] In the first example, an instance of asthma exists (in John). In the second example there is only a hypothesis referring to the concept asthma

[5] Semantic Tag - SNOMED CT Editorial Guide - SNOMED Confluence (ihtsdotools.org)

Table [Priority]: SNOMED Hierarchies, their use in annotation, their hierarchy tags (expression in brackets appended to the fully specified name) and their use for annotation

fully defined findings/disorders > observable > procedure/regime/therapy > finding/disorder/event > morphological abnormality > (medicinal) product (form)> clinical drug > specimen > physical object > body structure ("structure") > body structure ("entire") > cell structure > substance > organism > environment > occupation > unit of presentation > *all others > qualifier value Not to be used: record artifact, situation, metadata, linkage concept, link assertion, attribute, namespace concept, foundation metadata concept, OWL metadata concept, navigational concept

Table [Core]: SNOMED Hierarchies, their use in annotation, their hierarchy tags (expression in brackets appended to the fully specified name) and their use for annotation

Hierarchies and semantic tags	**Characterisation /** *Use for AAG annotation*
Core concepts - can stand alone - with an implicit link to the patient	
Clinical finding (finding, disorder)	**Any phenomenon a human body may exhibit and which is of clinical interest: diseases, injuries, risks, signs, symptoms, but also normal phenotypic characteristics** *Concepts that correspond to negative fact statements should be avoided, as long as they are primitive and can easily be expressed with their positive correlate and a verificationStatus assertion*
Event (event)	**Mostly short-term events that "just happen" and are reportable from a clinical perspective.** *Clinical findings, disorders and events are treated the same in the AAG, as they cannot easily be distinguished*
Observable entity (observable entity)	**Measurable or countable qualities of clinical relevance, which are only informative together with a qualitative value, a quantitative one, or a quantitative value with a unit.** *Observables without value are only acceptable to represent open questions*
Procedure (procedure, regime/therapy)	**Events that are planned and/or executed by a health care professional for therapeutic or diagnostic reasons.** *As an important exception, procedures concepts under 386053000 |Evaluation procedure (procedure)| are treated like observables and are completed a qualitative or quantitative values (+ units)*
Social context (social concept, ethnic group, life style, occupation, person, racial group, religion / philosophy, social status)	**Social, familial, professional and other roles of people and groups** *Ontologically, all social context concepts are roles, i.e. characteristics of people and groups that are often not essential (e.g. mother, student), as they (may) change. They are, however, treated like persons or groups themselves. The concept patient 116154003 |Patient (person)| will not be used for annotation, unless a clear relation to other individuals is required*
Specimen (specimen)	**Some matter (organ part, tissue, substance) taken out of the body and examined in the lab**
Staging and scales (staging scale, assessment scale, tumour staging)	**Similar to observables. Need to be completed by nominally or ordinally scaled values**
Supportive concepts - cannot stand alone, modify core concepts	
Body structure (body structure, morphological abnormality, cell, cell structure)	**From cells to organs and body, everything that constitutes a human body. Including abnormal morphological structures and features.** *Morphology concepts only in those cases where no corresponding finding concept is available*
Environment or geographical location (environment, geographic location)	**Spaces, from rooms to buildings to environments in which a patient can be located.**
Organism (organism)	**Entire biological organisms (viruses, bacteria, fungi, plants, animals) of relevance for health**
Pharmaceutical / biologic product (product, medicinal product, clinical drug, medicinal product form)	**All kinds of substances or combination of substances in a defined dose form used for therapy and prevention.** *Needs to be linked to a procedure or finding concept. Should always be given preference over the annotation with the substance alone*
Physical force (physical force)	**Non-biological external factors that have an influence on health.**
Physical object (physical object)	**Manufactured, macroscopic objects that are not drugs and have relevance for health**
Qualifier value (qualifier value,	**Broad range of concepts used to refine the meaning of other concepts, mostly corresponding to adjectives in**

dose form, basic dose form, disposition, role, unit of presentation, intended site, administration method, transformation, release characteristic, state of matter, supplier, product name)	**human language, but also including units of measurements** *The qualifier value hierarchy also includes clinical processes, dispositions etc. used for defining core concepts. Whenever possible, the more "meaningful" core concepts should be used.*
Substance (substance)	**Amounts of matter, chemically defined or not** *For lab parameters give preference to measurement procedures, for drugs use pharmaceutical drugs*
Not-To-Use Concepts	
Record artifact (record artifact) - *documentation related to administration, including types of documents and document parts*	
Situation with explicit context (situation) - *disorders and procedures in a context such as past history, negation, family history*	
SNOMED CT Model Component (metadata, linkage concept, link assertion, attribute,namespace concept, foundation metadata concept, OWL metadata concept) - *codes that support descriptions of the terminology itself and its use, but not the clinical domain*	
Special concept (navigational concept) - *to ignore, because they will be retired soon*	

All annotations with a non-core concept should be linked to some node of a core concept. In case the core concept is only explicit, the core annotation should be added as a zero-with annotation
Table [Priority]: Priorities between SNOMED Hierarchies in case of ambiguity
Avoid "isolated" concepts, i.e. those not referring to others and not being referred to by others

First priority	Second priority	Third priority	Example
body structure	qualifier value		**449826002 |Phimosis (disorder)|** > 1286931009 |Phimotic (qualifier value)|
Body structure ("structure")	body structure ("entire")		**64033007 |Kidney structure (body structure)|** > 181414000 |Entire kidney (body structure)|
finding / disorder / event	morphological abnormality	qualifier value	**118600007 |Malignant lymphoma (disorder)|** > 1163043007 |Malignant lymphoma (morphologic abnormality)|
observable	procedure (with value)	finding / disorder / event, substance	**75367002 |Blood pressure (observable entity)|** > 392570002 |Blood pressure finding (finding)| **79301008 |Electrolytes measurement (procedure)|** > 365758001 |Finding of electrolyte levels (finding)| > **59573005 |Potassium measurement (procedure)|** > 88480006 |Potassium (substance)|
substance	organism		**735971005 |Fish (substance)|** > 90580008 |Fish (organism)|
product	substance		**363598004 |Product containing diclofenac (medicinal product)|** > 7034005 |Diclofenac (substance)|
*	*	qualifier value	**86273004 |Biopsy (procedure)|** > 129314006 |Biopsy - action (qualifier value)| **363346000 |Malignant neoplastic disease (disorder)|** 21594007 |Malignant (qualifier value)|
specimen	body structure, substance		**119297000 |Blood specimen (specimen)|** > 256906008 |Blood material (substance)|
procedure/regime/therapy	substance		**182764009 |Anticoagulant therapy (procedure)|** > 372862008 |Anticoagulant (substance)|
finding / disorder / event	procedure/regime/therapy (without value)	Physical object	**441509002 |Cardiac pacemaker in situ (finding)|** > 307280005 |Implantation of cardiac pacemaker (procedure)| > 14106009 |Cardiac pacemaker, device (physical object)|

Table [Not-To-Use]: Semantic types not to be used

Type	Explanation
attribute, linkage concept, link assertion	(represents SNOMED relations. We will use **anno:relations** instead
situation	(e.g. family history, past history, negation, plan) Instead, postcoordination is preferred

metadata concept	Describes the terminology and not the domain
record artifact	Describes documents and their parts

Core concepts are typically related to supportive concepts via outgoing relations (following the SNOMED concept model), such as <Clinical condition, *causative agent,* Organism> or <Administration of drug or medicament, *Direct substance, S*ubstance>. Ideally, supportive concepts should only be used in case they are clinically important and not expressible as core concepts, and when the interpretation of other parts of the text depends on them. Qualifier values and units of measurement are used only when related to other concepts.

5.3.2. Predicates and their definitions

Table [predicates] lists the "close to user" alias predicates to be used, together with their origin in SNOMED CT and FHIR. The main reason for specific predicates is to shield users away from the complexity of the internal wiring of SNOMED CT and FHIR, including the possibility of redundant representations. The suggested predicates are expected to cover 95% of the relational assertions needed. Where a new predicate is required that is not in the list a new one can be suggested by the annotators, for which a rooting in FHIR, SNOMED or both can subsequently be sought by the maintainers of the AAG.

Regarding SNOMED CT predicates (linkage concepts in SNOMED CT, object properties in OWL) , ambiguous mappings (e.g. "site") - which are deliberately introduced to keep the set of predicates small - can be disambiguated in terms of finding or procedure site regarding their domain type. But they also support redundant representations, e.g. by mapping "site" also to FHIR representations in which it maps to bodySite.

"INV" indicates an inverse relation, '||' the concatenation operator. "<<" specifies the allowed values for domain and range, according to the SNOMED Expression Constraint Syntax (ECL).

Range restrictions with specific value sets (including the mapping from SNOMED CT codes to HL7-FHIR value sets are given in Table 2 (following SNOMED International mapping recommendations).

Table [predicates]: Predicate values, their domain and range restrictions and their rooting in SNOMED CT and FHIR-
Note: "medicinal/product/form" means the semantic tags "product", "medicinal product", or "medicinal product from"

Predicates (namespace: anno:)	**Domain** (as semantic tags or SNOMED ECL expressions)	**Relational expression** (rooted in **SNOMED** or **FHIR** linkage concepts / slots / relations)	**Range** (as as semantic tags or SNOMED ECL expressions)
actionStatus	medicinal/product/form, procedure, regime/therapy	INV(MedicationAdministration.medication) || MedicationAdministration.status	Cf. table values [E]
after	finding, disorder, event, procedure, regime/therapy	255234002 |After (attribute)|	finding, disorder, event, procedure, regime/therapy
beginAge	finding, disorder, event, procedure, regime/therapy	INV(Condition.code) || Condition.onsetAge.quantity.value	decimal
beginAgeUnit	finding,	INV(Condition.code) || Condition.onsetAge.q	< 767524001 |Unit of

	disorder, event, procedure, regime/therapy	quantity.code	measure (qualifier value)|
beginTime	finding, disorder, event, procedure, regime/therapy	INV(Condition.code) || Condition.onSet.dateTime	dateTime Qualifier value
		INV(ServiceRequest.code) || ServiceRequest.occurrence.dateTime	
		INV(Procedure.code) || Procedure.occurrence.dateTime	
clinicalStatus	finding, disorder, event	INV(Condition.code) || Condition.clinicalStatus	Cf. table values [C]
device	procedure, regime/therapy	363699004 |Direct device (attribute)|	physical object
doseForm	medicinal/product/form, clinical drug,	763032000 |Has unit of presentation (attribute)| 411116001 |Has manufactured dose form (attribute)|	unit of presentation dose form
doseMethod	Clinical drug	411116001 |Has manufactured dose form (attribute)| || 736472000 |Has dose form administration method (attribute)|	administration method
doseQuantity	procedure	Dos age.doseAndRate.dose	decimal
doseRate		Dosage.doseAndRate.rate (per unit of time)	decimal
doseRoute	procedure	410675002 |Route of administration (attribute)|	< 284009009 |Route of administration value (qualifier value)|
doseTiming	procedure	Dosage.timing	qualifier value
duration	*		decimal
durationUnit	*		<<767524001 |Unit of measure (qualifier value)|
dueTo	finding, disorder, event, procedure, observable regime/therapy	246075003 |Causative agent (attribute)|	organism, physical force, substance, medicinal/product/form
		42752001 |Due to (attribute)|	finding, disorder, event, procedure, regime/therapy,
endAge	finding, disorder, event, procedure, regime/therapy	INV(Condition.code) || Condition.abatementAge.quantity.value	decimal
endAgeUnit	finding, disorder, event, procedure, regime/therapy	INV(Condition.code) || Condition.abatementAge.quantity.code	<<767524001 |Unit of measure (qualifier value)|
endTime	finding, disorder, event procedure, regime/therapy	INV(Condition.code) || Condition.abatement.dateTime INV(Procedure.code) || Procedure.occurrence.period.end.dateTime	dateTime
evidence	finding, disorder, event procedure,	INV(Condition.code) || Condition.evidence	
hasFocus	procedure, regime/therapy	363702006 |Has focus (attribute)|	finding, disorder, event, procedure, regime/therapy

			specimen
hasIntent	procedure, regime/therapy	363703001 |Has intent (attribute)|	qualifier value
interprets	finding	363714003 |Interprets (attribute)|	procedure, regime/therapy, observable entity
inFamily	finding, disorder, event	INV(FamilyMemberHistory.condition) || FamilyMemberHistory.relationship	person
		INV(246090004 |Associated finding (attribute)|) || 408732007 |Subject relationship context (attribute)|	
informant	finding, disorder, event	INV(Condition.code) || INV(Provenance.target) || Provenance.agent.type	
	procedure, regime/therapy	INV(Procedure.code) || INV(Provenance.target) || Provenance.agent.type	
	observable entity, procedure, staging scale	INV(Observation.code) || INV(Provenance.target) || Provenance.agent.type	
ingredient	medicinal/product/form	127489000 |Has active ingredient (attribute)|	< 105590001 |Substance (substance)|
familyDeath	finding, disorder, event	INV(FamilyMemberHistory.relationship) || FamilyMemberHistory.condition.contributedToDeath	boolean
inheresIn	observable entity, procedure,	INV(Observation.code) || Observation.interpretation 704319004 |Inheres in (attribute)| || 718497002 |Inherent location (attribute)| **For procedures use only where they have the meaning of observable**	body structure, morphologic abnormality, specimen, person, physical object, medicinal/product/form
laterality	body structure	272741003 |Laterality (attribute)|	<<182353008 |Side (qualifier value)|
	finding disorder	363698007 |Finding site (attribute) || 272741003 |Laterality (attribute)|	
	procedure, regime/therapy	405813007 |Procedure site - Direct (attribute)| || 272741003 |Laterality (attribute)|	
	specimen	118169006 |Specimen source topography (attribute)| || 272741003 |Laterality (attribute)|	
medicationStatus			
morphology	procedure	363700003 |Direct Morphology (attribute)|	morphologic abnormality
	finding, disorder, event	116676008 |Associated morphology (attribute)|	
	specimen	118168003 |Specimen source morphology (attribute)|	
notPerformedReason	procedure, regime/therapy	408730004 |Procedure context (attribute)| Procedure.StatusReason.ProcedureNotPerformedReason	Cf. table values [F]
or	any	Logical OR	any
otherThan	any	**Specifies the domain entity by excluding any entities of the range type**	any
partOf	body structure	774081006 |Proper part of (attribute)|	body structure, specimen
	finding, disorder,	**only entities of the same semantic type can be related by this predicate**	finding, disorder,

	event		event
	procedure		procedure
	specimen		Specimen, body structure
	observable entity		observable entity
requestIntent	procedure, regime/therapy	INV(ServiceRequest.code) || ServiceRequest.intent 363703001 |Has intent (attribute)|	Cf. table values [G]
sameAs	any	**Coreference between sub- and superconcepts (is-a)**	any
severity	finding, disorder, event	INV(Condition.code) || Condition.severity	Cf. table values [D]
		246112005 |Severity (attribute)|	
siteDirect	finding, disorder, event	363698007 |Finding site (attribute)|	body structure
		INV(Condition.code) || Condition.bodySite	
	observable	704327008 |Direct site (attribute)|	
	procedure, regime/therapy	405813007 |Procedure site - Direct (attribute)|	
		INV(Procedure.code) || Procedure.bodySite	
	specimen	118169006 |Specimen source topography (attribute)|	
	morphologic abnormality	inv(116676008 |Associated morphology (attribute)|) || 363698007 |Finding site (attribute)|	
siteIndirect	procedure, regime/therapy	405814001 |Procedure site - Indirect (attribute)|	
specimen	procedure, regime/therapy	116686009 |Has specimen (attribute)|	specimen
value	finding, disorder, event	363713009 |Has interpretation (attribute)|	qualifier value
	Body structure	To be defined	qualifier value
	observable entity, staging scale, <<386053000 |Evaluation procedure (procedure)|, << 763158003 |Medicinal product (product)|, <<105590001 |Substance (substance)|	INV(Observation.code) || Observation.value.quantity.value	decimal
		1142138002 |Has concentration strength numerator value (attribute)| 1142135004 |Has presentation strength numerator value (attribute)| INV(Observation.code) || Observation.value.CodeableConcept	qualifier value
	*	also to express cardinality (count of entities of the same type)	decimal
valueDenominator		1142137007 |Has concentration strength denominator value (attribute)| 732947008 |Has presentation strength denominator unit (attribute)|	
valueLow		INV(Observation.code) || Observation.value.Range.low	decimal
valueHigh		INV(Observation.code) || Observation.value.Range.high	decimal
valueComparator	medicinal/product/form observable entity, staging scale, <<386053000 |Evaluation procedure (procedure)|	INV(Observation.code) || Observation.value.quantity.comparator	Cf. table values [A]

unit	medicinal/product/form observable entity, <<386053000 |Evaluation procedure (procedure)|	INV(Observation.code) || Observation.value.quantity.code	<<767524001 |Unit of measure (qualifier value)|
	373873005 |Pharmaceutical / biologic product (product)| <<105590001 |Substance (substance)|	733725009 |Has concentration strength numerator unit (attribute)| 732945000 |Has presentation strength numerator unit (attribute)|	
unitDenominator		733722007 |Has concentration strength denominator unit (attribute)| 732947008 |Has presentation strength denominator unit (attribute)|	
usingSubstance	procedure, regime/therapy	363701004 |Direct substance (attribute)| 246093002 |Component (attribute)|	substance medicinal/product/form [6]
verificationStatus	finding, disorder, event	INV(Condition.code) || Condition.verificationStatus	Cf. table values [B]
		INV(246090004 |Associated finding (attribute)|) || 408729009 |Finding context (attribute)|	

[6] That 'sct:direct substance' can here be linked to product instead of substance only is explicitly allowed on the annotation level, but has to be dealt with in further KG processing

5.3.3. Predicate values and their definitions

Table [values]: Predicate values and their rooting in SNOMED CT and FHIR

Relation alias prefix anno:	FHIR value	Default	Corresponding SNOMED CT concepts
[A] anno:valueComparator	<		276139006 |Less-than symbol < (qualifier value)|
	< =		276137008 |Less-than-or-equal symbol <= (qualifier value)|
	=	x	276136004 |Equal symbol = (qualifier value)|
	>		276140008 |Greater-than symbol > (qualifier value)|
	> =		276138003 |Greater-than-or-equal symbol >= (qualifier value)|
	<>		431878004 |Inequality symbol <> (qualifier value)|
[B] anno:verificationStatus	Unconfirmed		415684004 |Suspected (qualifier value)|
	Provisional		410592001|Probably present (qualifier value)|
	Confirmed	x	410605003 |Confirmed present (qualifier value)|
	Refuted		410516002 |Known absent (qualifier value)|
	Entered-in-error		723510000 |Entered in error (qualifier value)|
	Unknown		261665006 |Unknown (qualifier value)|
[C] anno:clinicalStatus	Active	x	394774009 |Active problem (qualifier value)|
	Inactive		394775005 |Inactive problem (qualifier value)|
	Resolved		410513005 |In the past (qualifier value)|
	Recurrence		255227004 |Recurrent (qualifier value)|
	Remission		277022003 |Remission phase (qualifier value)|
	Relapse		263855007 |Relapse phase (qualifier value)|
[D] anno:severity	Mild		255604002 |Mild (qualifier value)|
	Mild to moderate		371923003 |Mild to moderate (qualifier value)|
	Moderate		6736007 |Moderate (severity modifier) (qualifier value)|
	Moderate to severe		371924009 |Moderate to severe (qualifier value)|
	Severe		24484000 |Severe (severity modifier) (qualifier value)|
	Life threatening severity		442452003 |Life threatening severity (qualifier value)|
	Fatal		399166001 |Fatal (qualifier value)|
[E] anno:actionStatus	in-progress		385651009 |In progress (qualifier value)|
	not-done		385660001 |Not done (qualifier value)|
	on-hold		385655000 |Suspended (qualifier value)|
	completed	x	410513005 |In the past (qualifier value)|
	entered-in-error		723510000 |Entered in error (qualifier value)|
	stopped		410545000 |Stopped before completion (qualifier value)|
	unknown		410537005 |Action status unknown (qualifier value)|
[F] anno:notPerformedReason	contraindicated		410536001 |Contraindicated (qualifier value)|
	discontinued		410546004 |Discontinued (qualifier value)|
	not done		385660001 |Not done (qualifier value)|
	not indicated		410534003 |Not indicated (qualifier value)|
	not offered		410530007 |Not offered (qualifier value)|
	not wanted		410528005 |Not wanted (qualifier value)|
	refused		443390004 |Declined (qualifier value)|
	stopped		385654001 |To be stopped (qualifier value)|
			385647007 |Rejected by performer (qualifier value)|
[G] anno:requestIntent	Considered and not done		385661002 |Considered and not done (qualifier value)|
	Planned		397943006 |Planned (qualifier value)|
	Not to be stopped		385653007 |Not to be stopped (qualifier value)|

The main interesting point of the relations tabulated in Table 1 is that they can be transformed into a set of connected concepts, FHIR elements, or SNOMED CT relations ("linkage concepts" corresponding to OWL object and datatype properties), or both in further steps, which provides a significant level of interoperability. In order to ensure clarity and understanding, we give a detailed presentation of the construction of the 'inFamily' and 'clinicalStatus' relations using both SNOMED CT and FHIR.

Several semantic assumptions have been formulated in this annotation guide that are considered essential to be followed throughout the entire annotation process regarding clinical documents:

- The default subject throughout the entire process is the patient (e.g., the subject of care) and there is no need to annotate it.
- Conditional, hypothetic and imperative expressions, as well as questions, should not be annotated.
- The default value for the presence of a clinical condition is 'known' does not need to be annotated.

Note: The default values for the ranges of the introduced predicates are shown in the 'default' column of the Table[values].

5.3.4. Metadata

(except the descendants of 900000000000441003 |SNOMED CT Model Component (metadata)|).

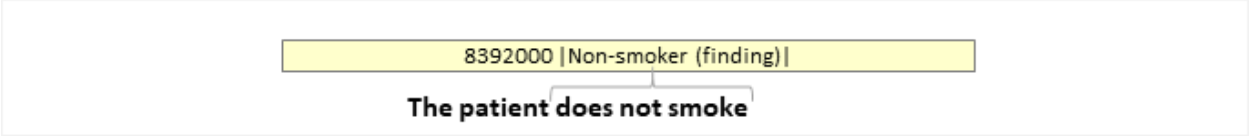
Metadata is an essential part of the annotation to understand, reproduce, and improve the annotation process. Following preliminary work[7], for the AAG the scores "full coverage", "partial coverage", "inferred coverage", and "none" are recommended, with "full coverage" being the default value for entity and relation annotations.
The assignment of coverage scores should be carried out on a pragmatic level in order not to unnecessarily slow down the annotation process. In general, the assignment of concept coverage scores should be carried out on a pragmatic level in order not to unnecessarily slow down the annotation process.

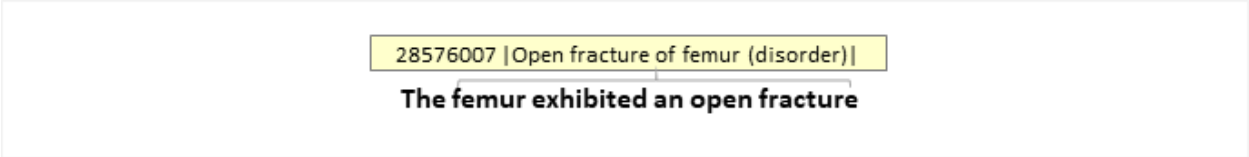
[7] Miñarro-Giménez JA, Cornet R, Jaulent MC, Dewenter H, Thun S, Gøeg KR, Karlsson D, Schulz S. Quantitative analysis of manual annotation of clinical text samples. Int J Med Inform. 2019 Mar;123:37-48. doi: 10.1016/j.ijmedinf.2018.12.011. Epub 2018 Dec 31. PMID: 30654902.

5.4. Specific annotation rules

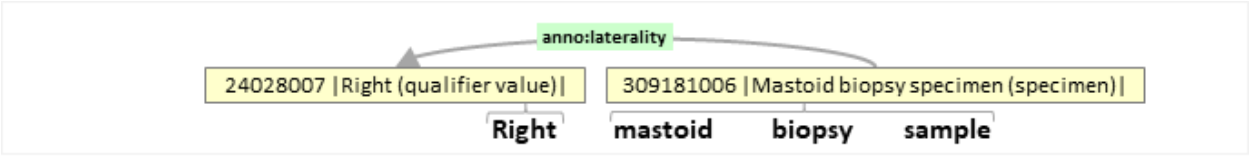
5.4.1. Delineation of annotation spans

Preference is given to the longest matching span. Unspecific words such as articles “the”, “a”, pronouns such as “we”, auxiliary or modal verbs, e.g., “is”, “will”, but also “patient” or proper names are not covered.



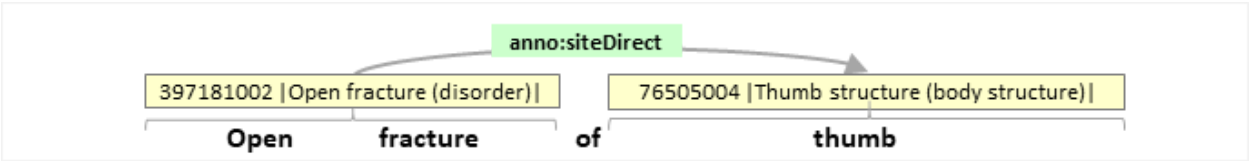


In case of different possibilities of partitioning a span, preference is given to the partition that produces a maximally fine grained core concept. Although an optimal match is often possible by using a concept with the “situation” tag (e.g. 394967008 |Suspected asthma (situation)|), this hierarchy will not be used and post-coordination will be done instead..



Although semantically the same, this would be preferred over the combination of 258415003 |Biopsy specimen (specimen)| and 763872005 |Structure of right mastoid bone (body structure)|

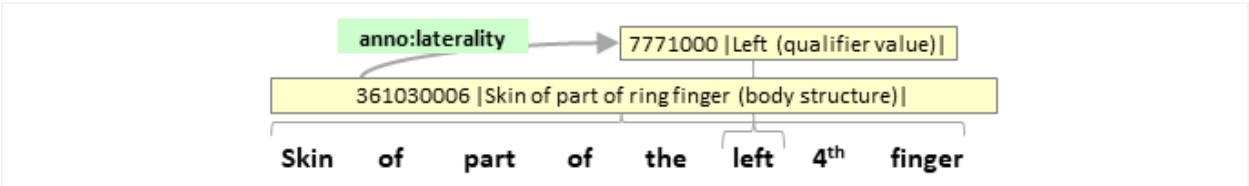
Pre-coordinated content in SNOMED CT is not evenly distributed as it follows the requests of the SNOMED users. Frequent and severe conditions and procedures are more likely pre-coordinated. E.g., there is a single code for “open fracture of femur” (see above) but not for “open fracture of thumb”.



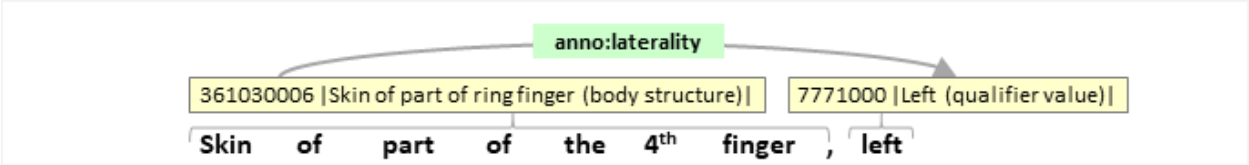
Annotating “open fracture of femur” the same way would not be semantically wrong, but the AAG gives priority to the longest match principle in order to facilitate agreement between annotators.

5.4.2. Overlaps and subword annotations

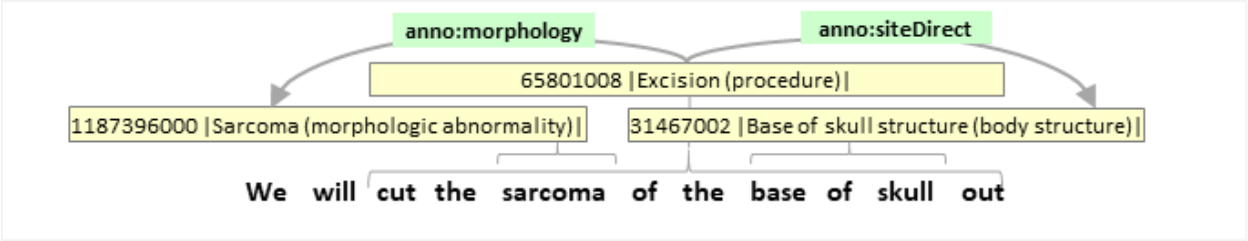
An annotation can refer to a non-contiguous passage. i.e. may bridge tokens that are not or differently annotated. However, the must not include sentence boundaries or paragraphs.



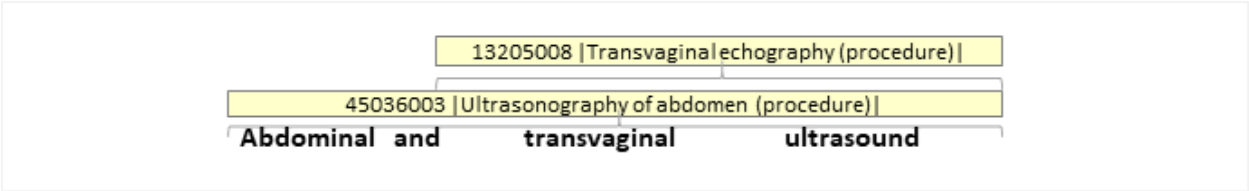
Annotations can therefore overlap. The previous one is semantically equivalent to the following, non-overlapping one:



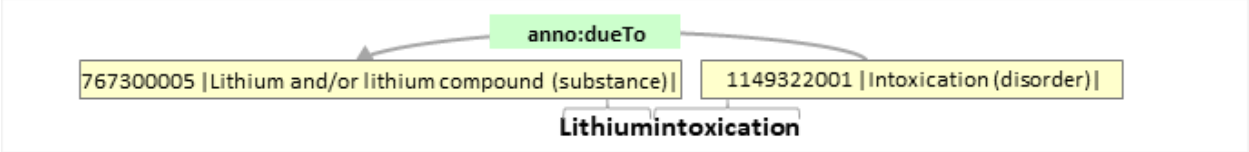
Split verb prefixes of English phrasal verbs sometime require a long annotation span. “Cut … out” is here the correct grounding for the concept 65801008 |Excision (procedure)|



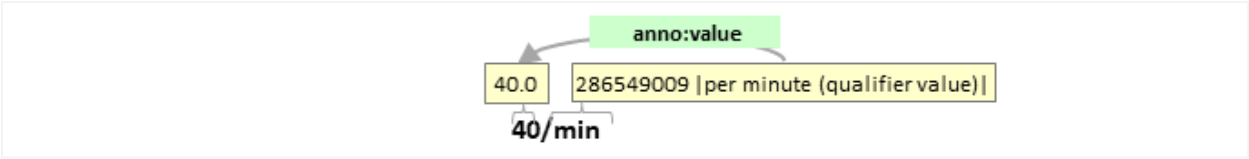
Typical non-contiguous passages occur in coordinations, i.e the juxtaposition of noun phrases connected with “and” and commas.



Single-word compounds are common in some languages, e.g. German. This requires that word parts are annotated separately:



The same may be necessary in case of fusion between numeric values and units, which often occur without being separated by spaces.



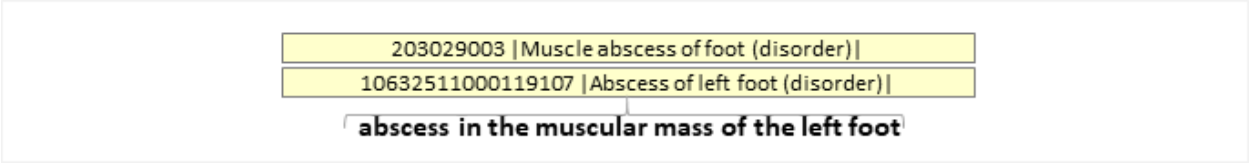
Annotation spans should not go beyond the word(s) to which the annotation applies. This means that it does not include articles, prepositions or conjunctions. Therefore the word “The” should not be in the annotation span. That “a” is in the annotation does not contradict the rule, because “fracture” is still needed and annotations cannot be discontinuous.

Annotation spans must not cross boundaries of sentences or paragraphs. A sentence is bounded by a period, exclamation sign, interrogation sign or a new line character. Commas, colons and semicolons are not sentence

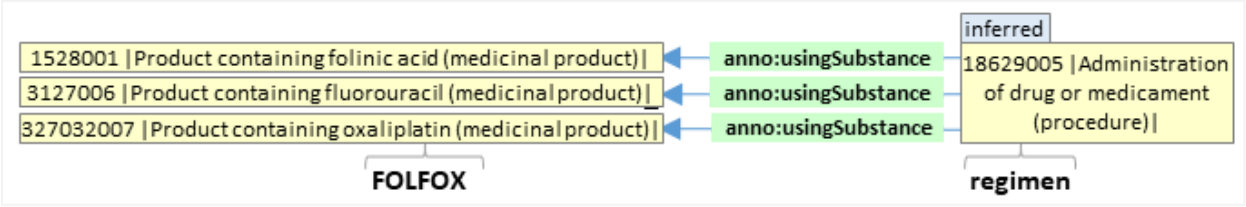
boundaries. For instance, the annotation of the upper example would not changed with “nose: fracture”, or “nose, fractured”. However, if split into sentences such as “The nose was examined. It had a fracture”, “fracture” and “nose” would be annotated separately and linked by the predicate “directSite”.
The longest match rule should not be applied in cases where the longest matching concepts does not have any internal semantics. For instance, “255249005 |48 hours (qualifier value)|” does not have any defining attributes, and there is no concept for, e.g. 47 or 49 hours.

5.4.3. Superposition

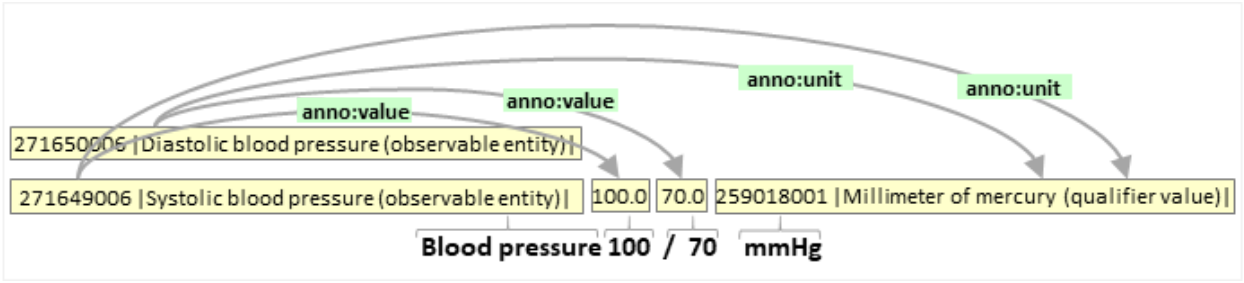
One and the same span may have two or more coinciding annotations. A typical case is where the meaning of the span can only be sufficiently expressed by more than one concept. This corresponds to a logical “AND”, but requires that the concepts are of the same semantic type. If such a combination is possible, we prefer concept superposition of concepts that are complementary in meaning, over complicated structures with predicate annotations. However, the annotation span needs to be exactly the same.



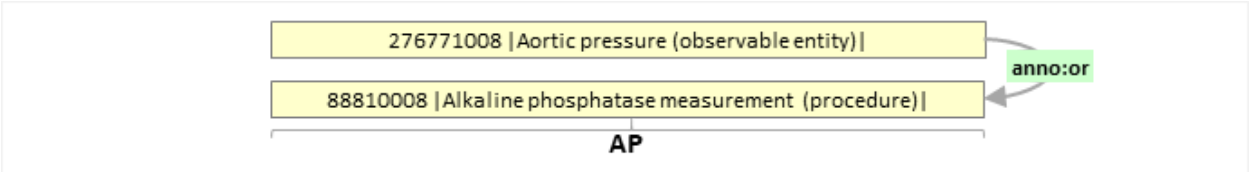
In other cases there is no typical logical “AND”, but the interpretation derives from the relational structure:



A typical case for this is also the mention of blood pressure, a combination of diastolic and systolic blood pressure, each one with their own attributes:

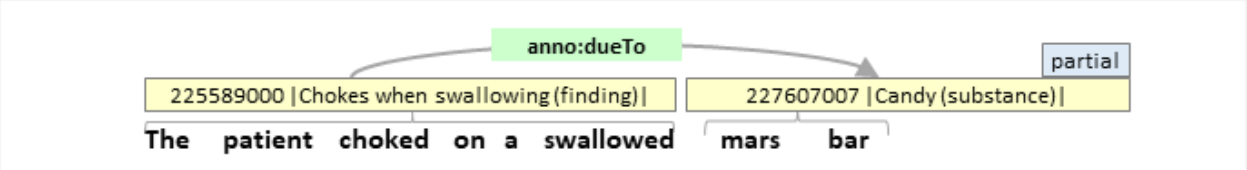


Wherever a logical combination corresponds to an “AND”, this is stated in the annotation. Note that the implicit meaning of findings, disorders, procedures and products is always “finding with…”, “product containing…”, which makes a logical “AND” plausible.
Lexical ambiguities should be resolved by the annotator if the context is clear. In cases where annotators cannot decide, two or more annotations can be added, which then need to be pairwise annotated by anno:or.

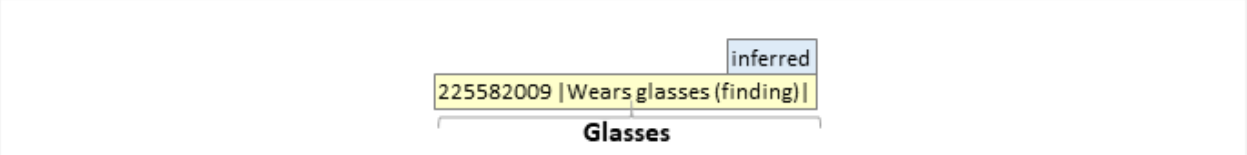


5.4.4. Partial and inferred annotations

Unspecific annotations due to limited granularity of the annotation vocabulary receive a meta-annotation "Incomplete".



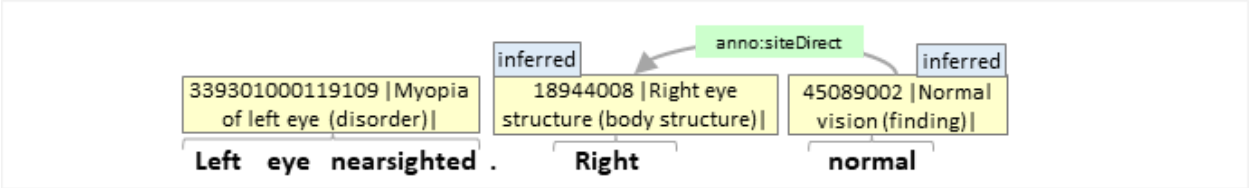
In case the meaning is inferred from the context, we add a meta annotation "inferred"



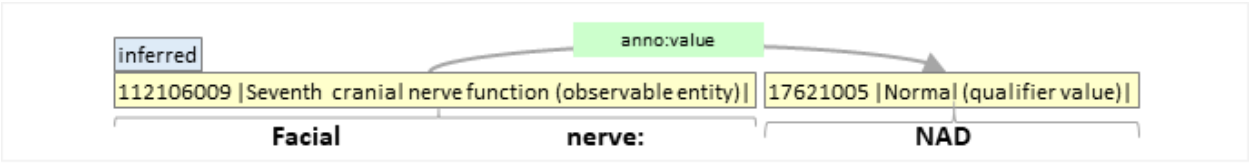
5.4.5. Ellipsis and zero-width annotation

Ellipsis means the omission of text content that can be inferred from domain knowledge and context. Expressing "patient wears glasses" just by "glasses" is a common phenomenon in clinical narratives. However, the AAG would not allow to literally assign the concept 50121007 |Eye glasses, device (physical object)|, because we need a core concept for annotation. In this case it is 225582009 |Wears glasses (finding)|, but we

In obvious cases, annotators should choose the code that is really meant for annotation. This is corroborated by the rule to give preference to concepts from the core hierarchies. However, for elliptic annotations the property "elliptic" should be set.

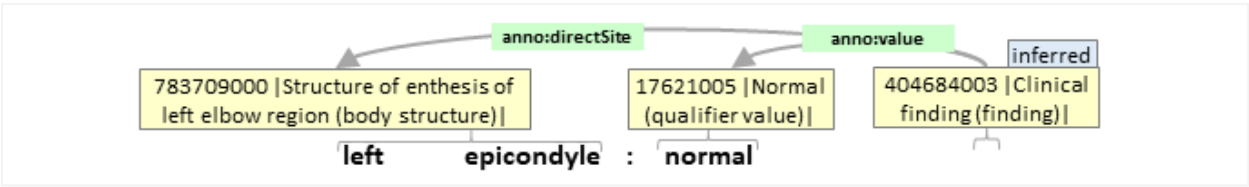


Very common are ellipses in expressions that assign a property to some body structure, which however means that the examination of that organ yielded a normal result.

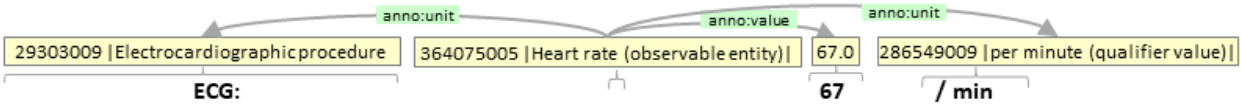


Wherever the whole statement can be expressed by a single core concept, follow the longest match principle. In bothe of the following two examples, core concepts could be inferred:

Wherever a core concept is meant but not explicitly stated, we add a zero-width annotation. Instead of anchoring this concept in a text expression, it is anchored in the space where the missing word would be expected (In this example the missing word would be “finding”.



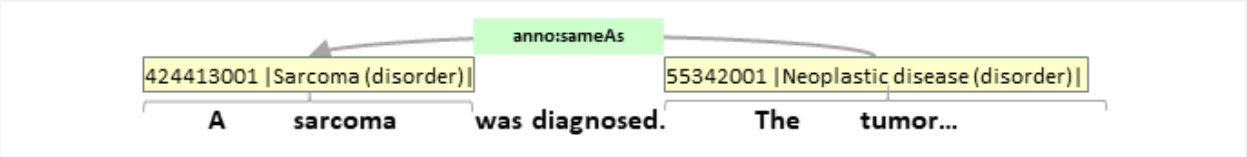
In many cases, however, there is already a coordination of a finding and an anatomical site, such as 301230006 |Lung finding (finding)|, where even the combination with “normal” exists: 827032003 |Normal lung (finding)|. These precoordinated concepts must be given preference, according to our longest match rule.



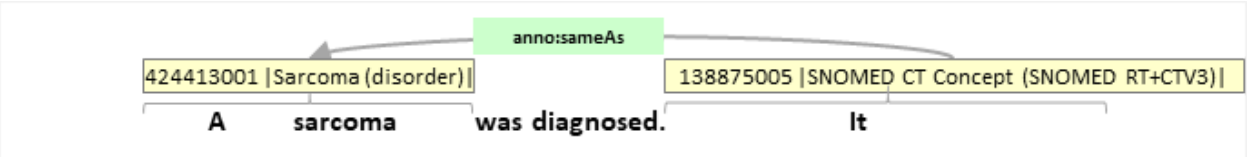
Here, the central concept “heart rate” is missing (because it is assumed that heart rate is the only parameter in an ECG with the dimension 1/min). Zero-width annotations always carry the meta-annotation “inferred”

5.4.6. Coreference

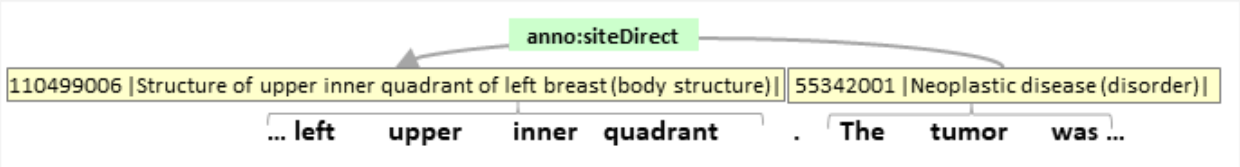
The same individual thing, e.g. a particular disease or procedure of a patient, is often mentioned more than once in a clinical document. After a first mention that typically uses a precise term, subsequent mentions use a more general wording. This is known as nominal anaphor and constitutes a rather frequent phenomenon. We annotate the anaphoric expression at the level of granularity it appears and relate it to the precise annotation using the coreference predicate “anno:sameAs”.

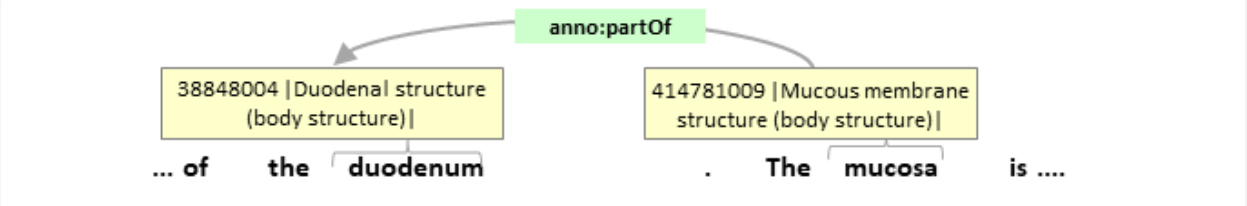


The same with pronominal anaphora, which is the same phenomenon using a personal or a possessive pronoun The pronoun is annotated with the top concept of SNOMED.

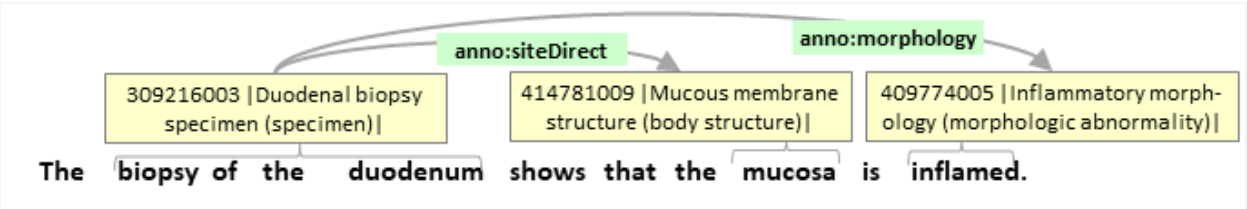


Coreference also includes the phenomenon of bridging anaphora. Here, the anaphoric expression does not refer to a prior one that is more specific, but which modifies and refines it. Bridging anaphora are just annotated with the appropriate relation, often across sentence boundaries.





Bridging anaphora may use the relation “anno:partOf” between anatomy terms. However, this is probably less common, because the anatomy terms are often part of a larger span that is annotated by a core concept in which the anatomy is implicit, see below.



5.4.7. Products, drugs and prescriptions

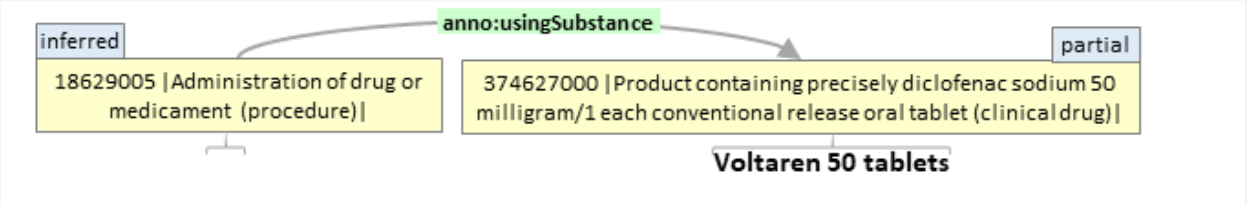
Drug-related statements are complex, as is the current drug model of SNOMED CT. The following distinctions are important for a better understanding:

- Branded drugs or clinical drugs are products. They contain one or more substances.
- Branded drugs are not in SNOMED. Clinical drugs are the most fine-grained concepts (including strengths and dose forms). They cover many common drugs, otherwise they can be post-coordinated
- Typical drug prescription statements have two parts. The “ontological” part (generally left part) specifies the type of drug products, the “epistemic” part (generally right part) specifies what was done or what

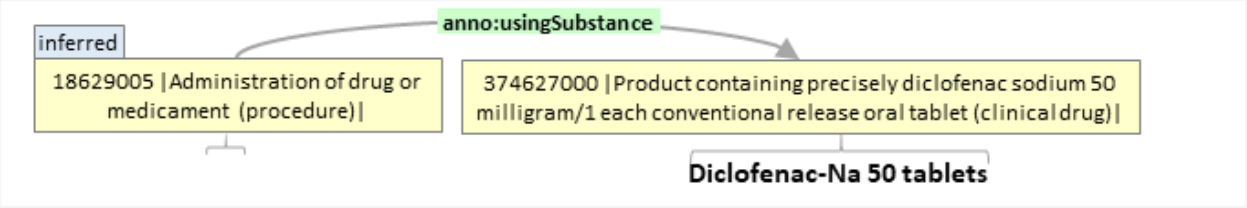
should be done with this drug in a patient. Therefore the first part relates a drug product with its attributes, i.e. ingredients, strengths and dose forms (i.e. the information a pharmacist needs to know for selling the product), whereas the second part relates the administration procedure with the product to be given, the route and the timing of the administration

- A complete drug statement depends on an instance of 18629005 |Administration of drug or medicament (procedure)| or any of its descendants

The mention of a pharmacological product in a prescription or medication administration statement is annotated with a code from the Pharmaceutical / biologic product hierarchy. Where this administration procedure is not explicitly stated, we add this concept as a zero-width annotation.

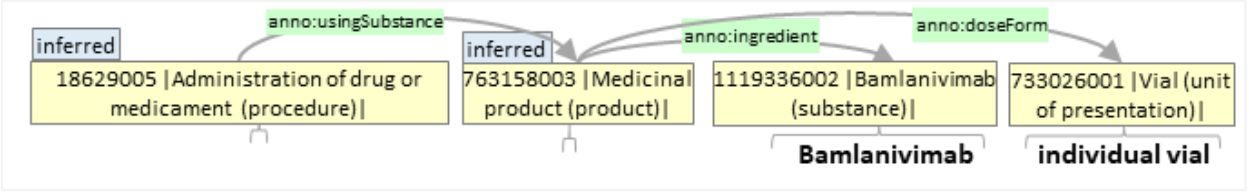


The annotation would be the same as with, e.g. “Dedolor 50 tablets”. The differences between brands in terms of salts and excipients cannot be accounted for, hence the meta-annotation partial.
However, if the text uses the substance term and not a brand name, the annotation is no longer partial.



In other contexts like overdose, poisoning or allergy, the annotation is done with the substance concept (in case there are not always pre-coordinated concepts such as 293613006 |Allergy to diclofenac (finding)|)

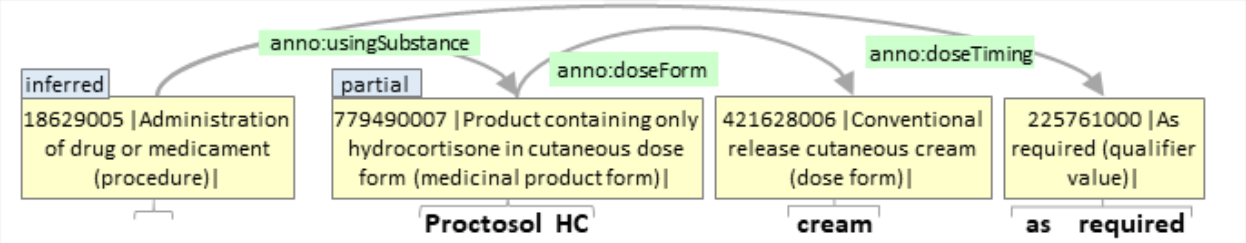
Where there is no concept from this hierarchy, the ingredients and other characteristics need to be linked, introducing 763158003 |Medicinal product (product)| as a zero-width annotation.



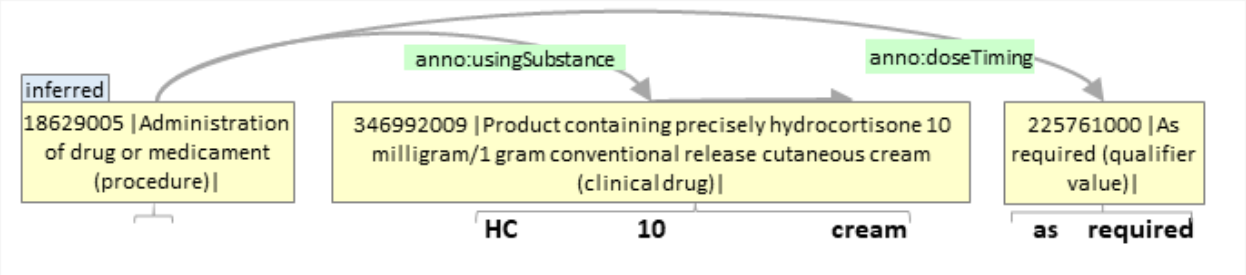
We make the following simplifying assumptions:

- If the unit (mostly milligrams) is missing in the text we assume the unit that is included in the SNOMED concept that matches in all other aspects (there is never any choice between micrograms and milligrams, or milligrams and grams)
- The most common dose form is tablet and capsule. Unless stated that they are prolonged releases we assume conventional release.

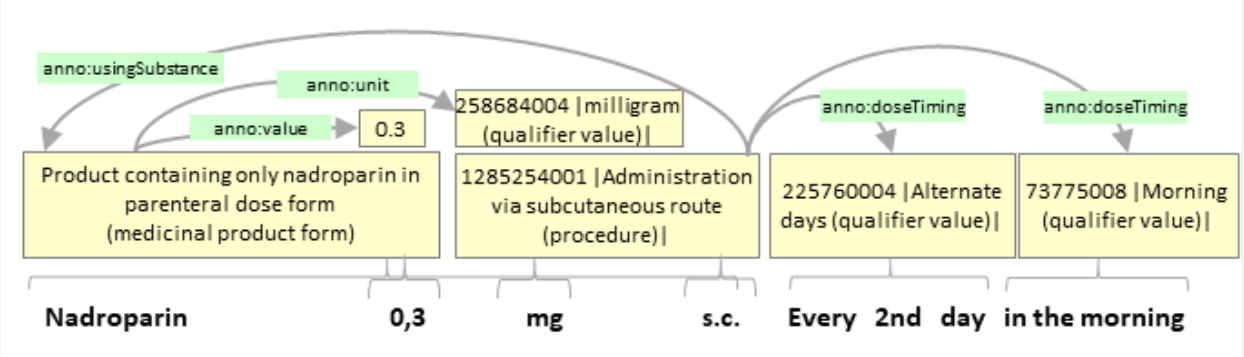
Nevertheless we only annotate what is actually there. In the following example the strength is not specified.



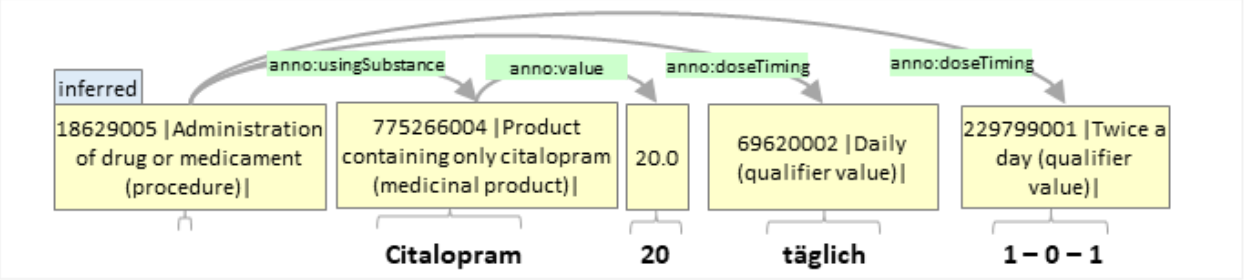
Compare with the following:

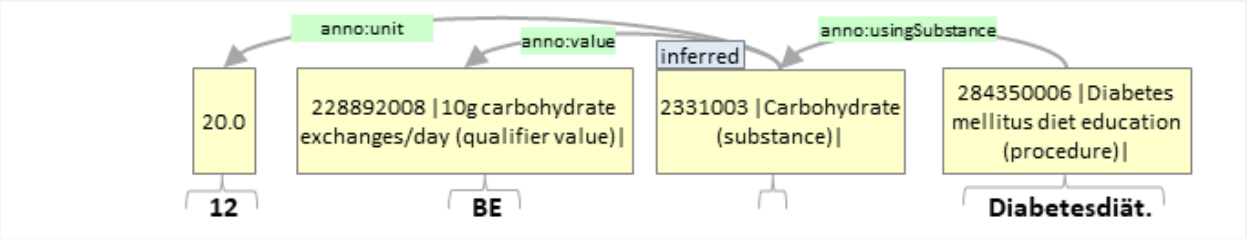


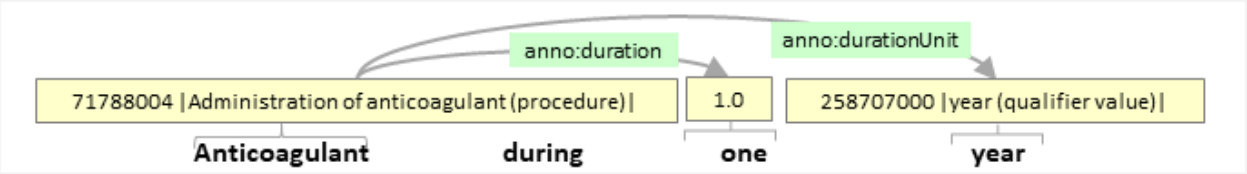
Here, the meaning of the SNOMED concept corresponds exactly to the annotated span. Although the denominator (/ gram) is not specified in the span, the fact that the strength of all HC creams in SNOMED are related to one gram allows to infer that this is also the case here.



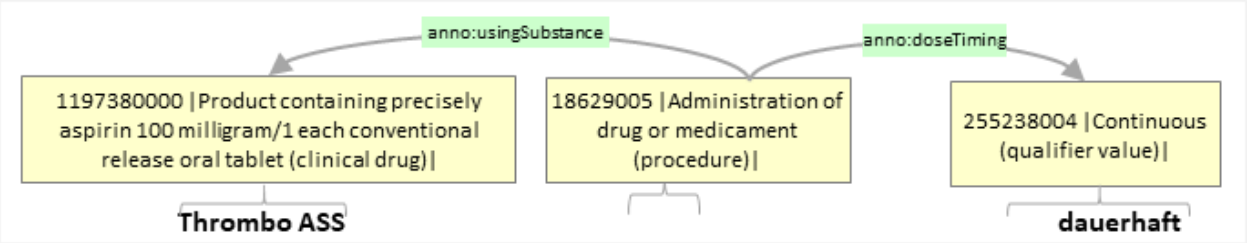
,



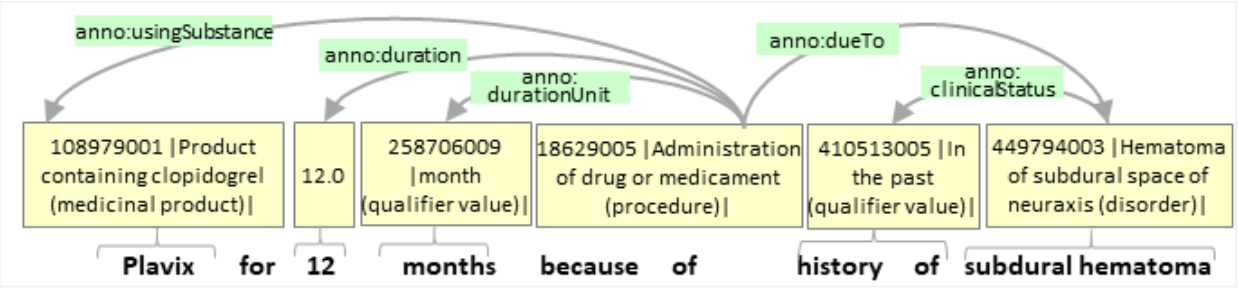




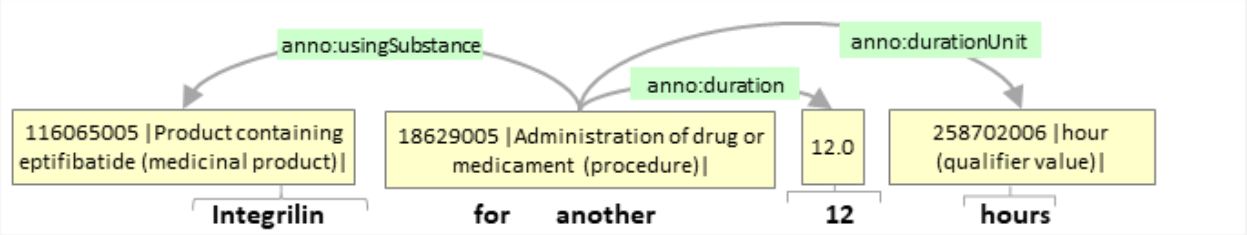
These examples also show that subconcepts of 18629005 |Administration of drug or medicament (procedure)| can be used, following the longest match principle.



This example shows the precise annotation of a clinical drug (which requires lookup in the drug documentation), but incomplete timing information. Probably it was assumed that the standard dose (1 * 100 mg) as recommended for the use of aspirin as an anticoagulant was given. Annotators, however, must not complete this information. The example also shows that the zero-width annotation does not need to use the concept "71788004 |Administration of anticoagulant (procedure)|". It would not be wrong but redundant, because the information of the type of drug follows from the very precise clinical drug concept (see axioms in the SNOMED browser)



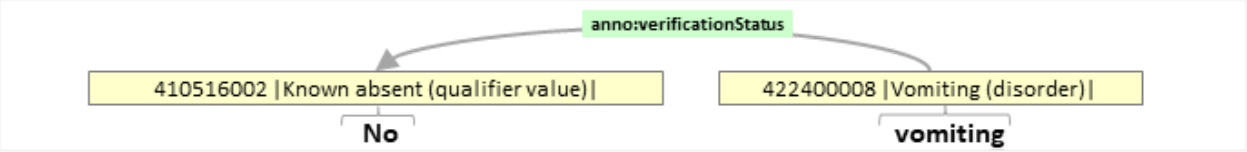
This last example might suggest that it describes a prescription rather than an administration. This can only be decided from the document context. Anyway, 33633005 |Prescription of drug (procedure)| could be an alternative. It shows also that the combination of a decimal number together with 258702006 |hour (qualifier value)| is given preference although 123035007 |12 hours (qualifier value)| exists. This is done because the latter concept does not have any internal semantics, from which the value "12" and the unit "hour" could be inferred.

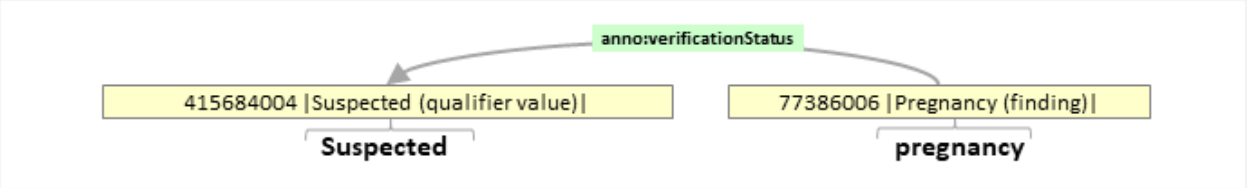


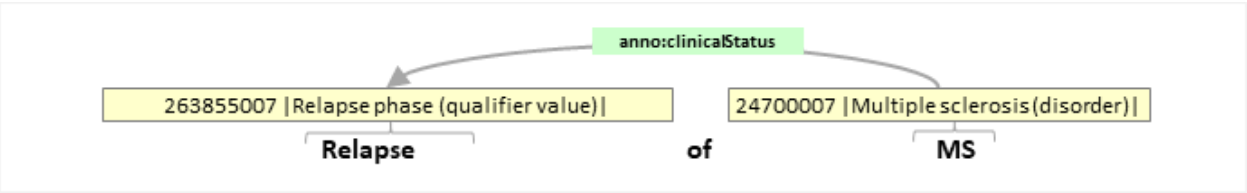
5.4.8. Negation, Uncertainty, Clinical status and Severity

Concepts with a negative meaning such as 162062008 |No vomiting (situation)| are to be avoided whenever they are expressible by combining the positive meaning (as given by SNOMED CT) with the value "refuted" in Condition.verificationStatus as given by FHIR. The use of SNOMED CT concept s with negative meaning is limited to those cases where there is no alternative, e.g. 249695006 |Absence of rib (finding)|. On the one hand, the very absence of a cody part is the definitional feature of a condition; on the other hand there is no such condition as "presence of rib" in SNOMED.

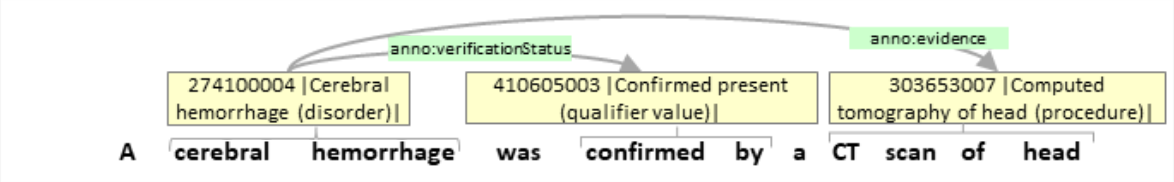
Then there is negative meaning embodied in overly popular terms such as "non-smoker" or "tumor-free", "asystolic", where the corresponding concepts should be used.





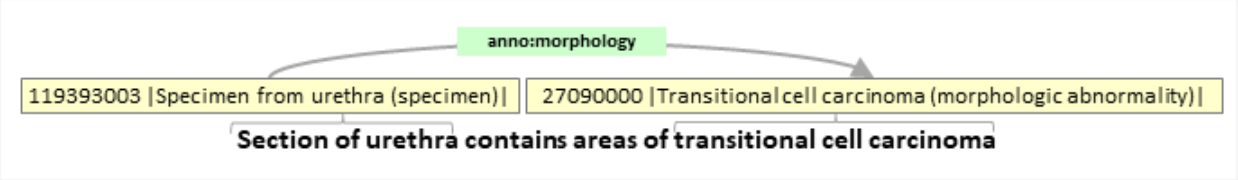


The fact that a condition is confirmed or refuted can be further specified using anno:evidence

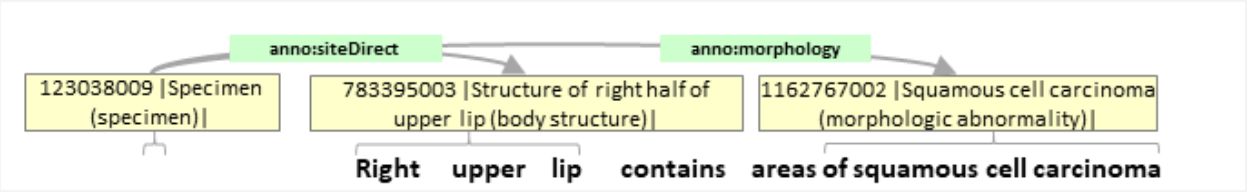


5.4.9. Specimens

The description of specimens should be distinguished from the description of patients and diseases. Specimens are parts of organs, tissues or body fluids taken out of the body for examination. Typically, specimens are examined by pathologists and bacteriologists. Blood specimens are examined for a large range of parameters (chemicals, enzymes, cells).
When specimens are referred to, always check whether there is a fitting concept from the specimen hierarchy. If specimens refer to abnormal morphology, use the morphology concept, even if there is a disorder concept of the same name.

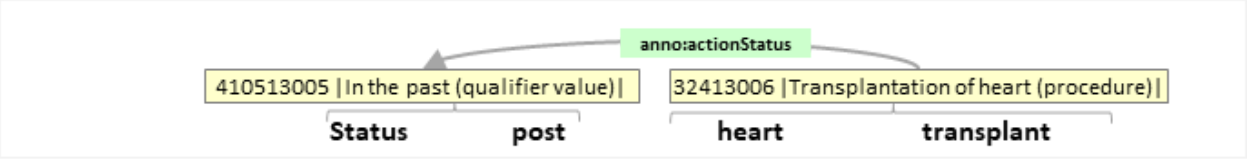


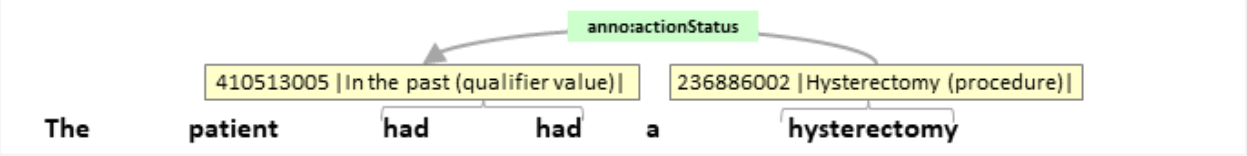
In case there is no specimen concept, refer to the appropriate body part concept from a zero-width specimen annotation.

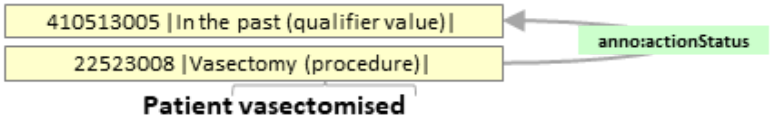


5.4.10. Procedure status and intent

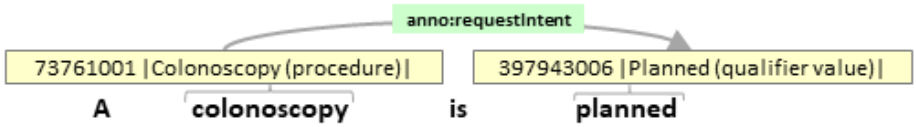
Procedure status is only annotated if the procedure was done prior to the episode of care described in the document.







Planned procedures are expressed with anno:requestIntent



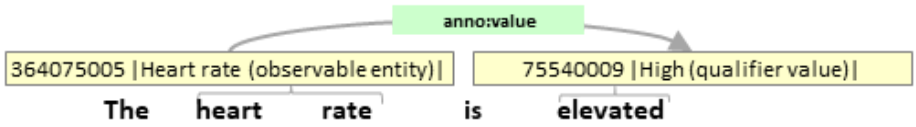
5.4.11. Qualitative Values

Concepts of the hierarchies “Observable entity” and “Staging and scales” are only used with a value. Wherever this hierarchy does not provide a concept to express the measurement of something use subconcepts of 785673007 |Measurement of level of substance in blood (procedure)| instead.

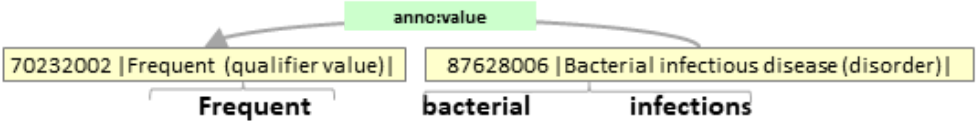
A current drawback of Observables is that they are often not related to their defining concepts, e.g. 446089006 |Volume of lower limb (observable entity)| is not related to the lower limb concept. It is therefore undefined how to refine observables via post-coordination, such as Volume of left lower limb. We here suggest that for laterality, the relation laterality is used in the same way as for body parts.

Values can be quantitative or qualitative. Qualitative values, like the attribution of “elevated” to an observable such as “systolic blood pressure” just requires the linkage of two concepts.

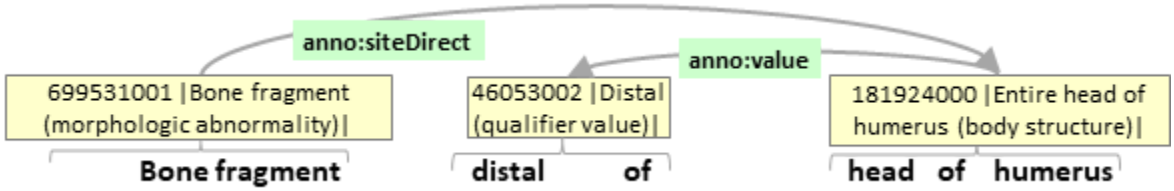
For relating observable entities (as well as SNOMED procedures substituting observables) with values we use the predicate anno:value It can be linked to codes when qualitative values are used:



Also other types of concepts may be refined by the association of values, e.g.

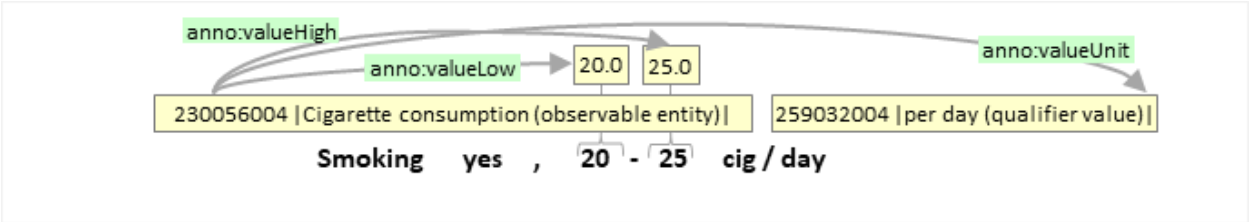


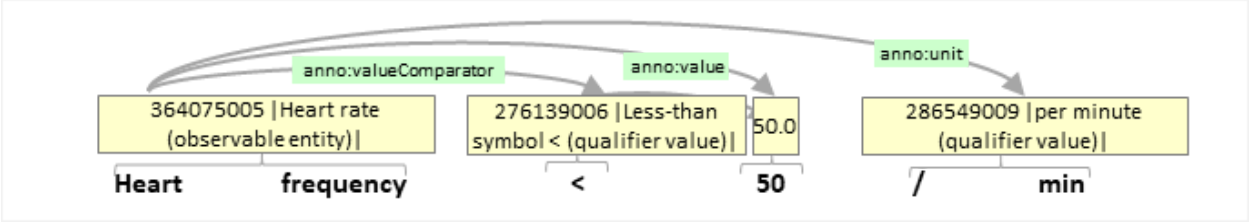
Anatomy is often further refined by adjectives such as “cranial”, “caudal”, i.e. descendants of 309825002 |Spatial and relational concepts (qualifier value)|. They are linked to the anatomy annotations using the anno:value predicate:



5.4.12. Quantitative values

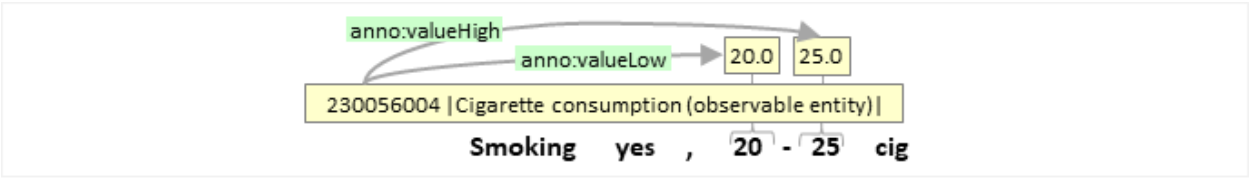
Quantitative values are more complex. Quantitative values are more diverse, e.g. “Heart rate < 40 /min” vs. “Heart rate 40 /min” vs. “Heart rate 40”. For relating observable entities with quantitative values we often have to consider units and in rare cases comparators. If there is no unit in the text, it is left out. As general, missing or unclear information is not guessed in the annotation process. Only concepts of the hierarchies “Observable entity” and “Procedure” use (non-temporal) quantitative values.

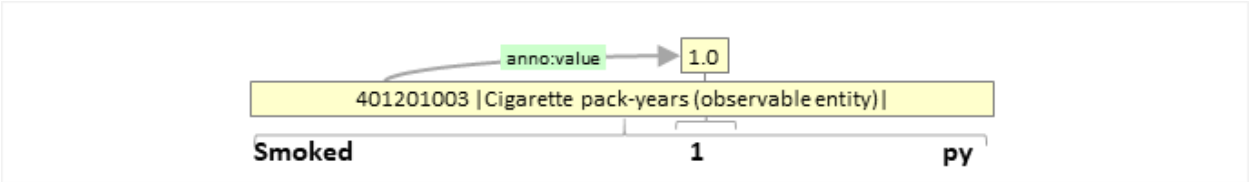


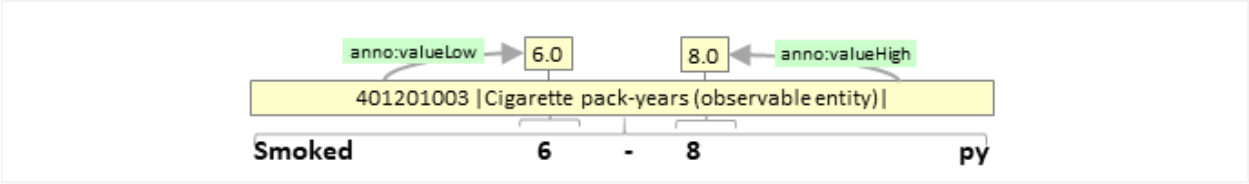


Often the entities to be counted are part of the meaning of the observable concept.

Often units of measurement are omitted, as they are easily inferred by the informed reader. This, however, is not a reason for a zero-width annotation.

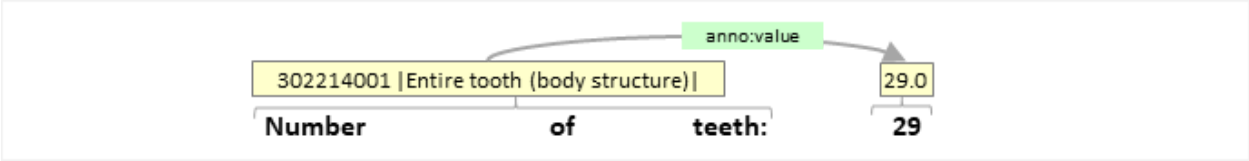






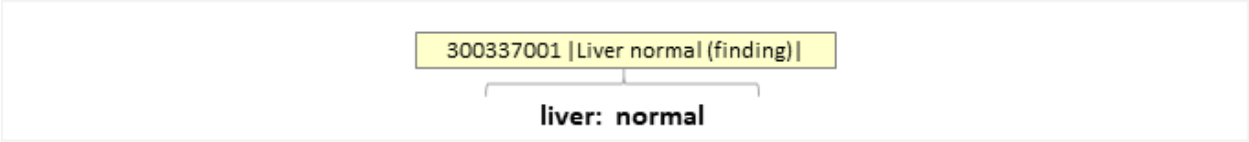
Numeric values are annotated as decimals. Note that the decimal separator is always the period. We use ".0" also for integer values. Values between 0 and 1 are always written with a leading 0: "0.1" (not ".1")

In cases where things (body parts, events, tablets) are to be counted without any appropriate observable concept available, we may directly assert a value link to an integer number.

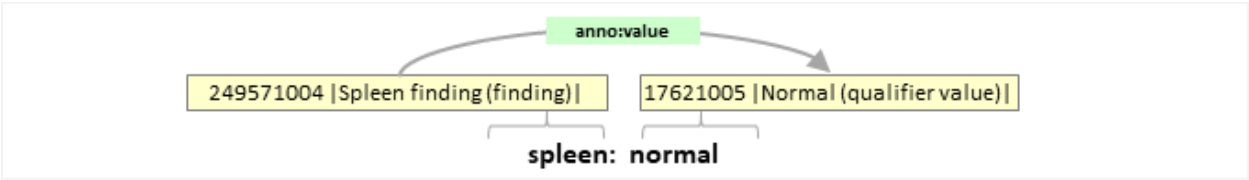


5.4.13. Normal and abnormal

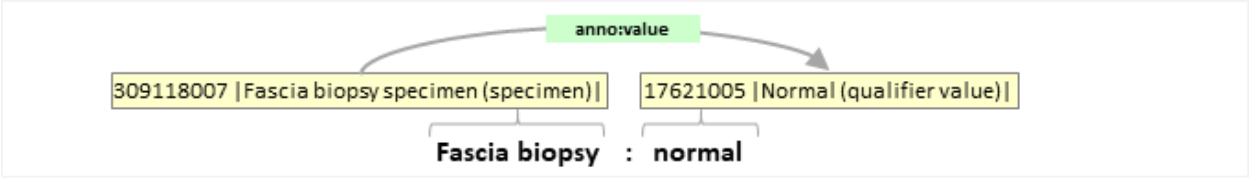
The qualifier values 263654008 |Abnormal (qualifier value)| and 17621005 |Normal (qualifier value)| can be related to observables and findings/disorders via the predicate anno:value. However, there are numerous pre-coordinated concepts that should be given preference according to the longest match rule.



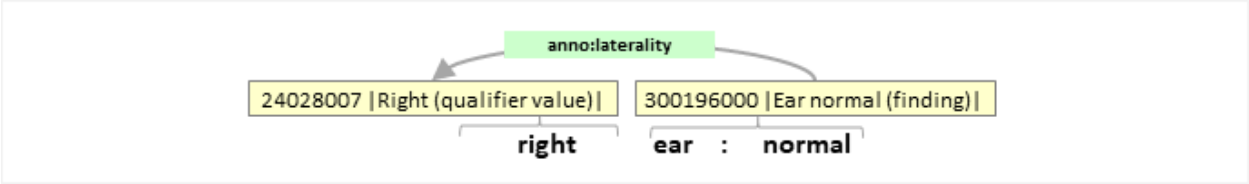
In other cases use always finding concepts (or, in second priority observables) to express analogous expressions for which no pre-coordinated normality finding exist:



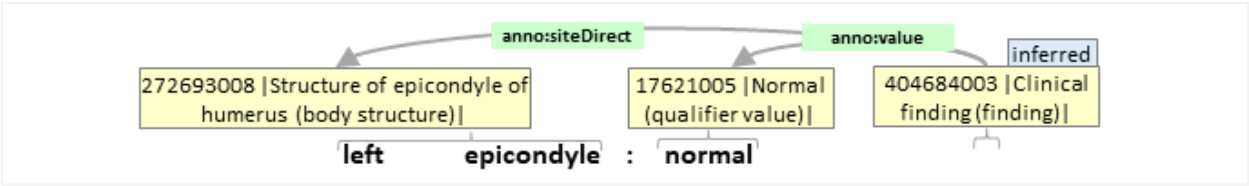
In no way body structure concepts such as 78961009 |Splenic structure (body structure)| can be linked to qualifiers, because body structures are not core. However specimens are core concepts; so the following is allowed:



Laterality qualifiers can be attached to findings, disorders and procedures:



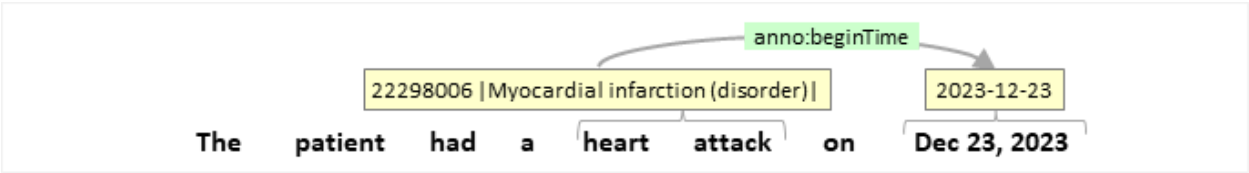
Whenever there is no appropriate core concept, we need to introduce the a general on as a zero-width annotation:



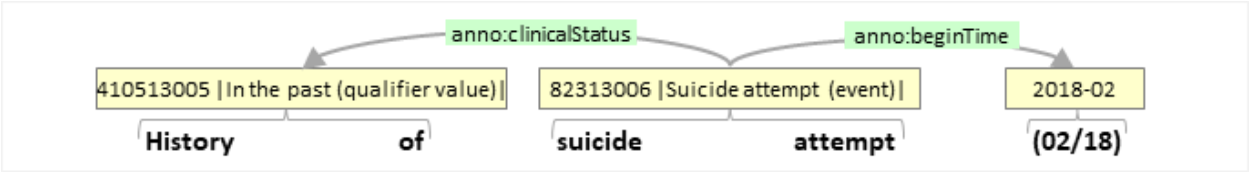
Choosing 116309007 |Finding of elbow region (finding)| instead of 404684003 |Clinical finding (finding)| would not be wrong, but unnecessary, because the exact anatomy is already covered by the body structure concept.

5.4.14. Temporal values

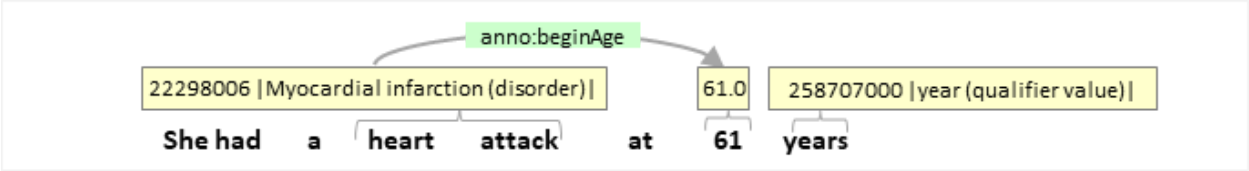
Everything can be related to a temporal value, such as a date, but also a time period or an age. For dates and time use the specification valid for FHIR: Datatypes - FHIR v6.0.0-cibuild: YYYY-MM-DD or YYYY-MM-DDThh:mm:ss. Incomplete dates and times are possible such as YYYY or YYYY-MM. If there is only one time reference the predicate anno:beginTime (although there is no end time) is used.



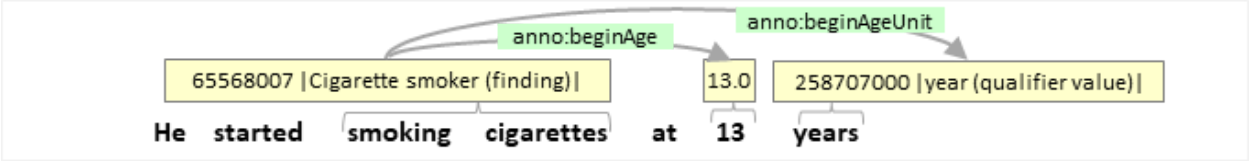
Reference to time can also be related with clinical status values. Incomplete dates use only yyyy or yyyy-mm. The missing year is completed from the context, e.g. Dec 23: “2023-12-23” . If the year is not clear, don’t annotate.



Reference to age :

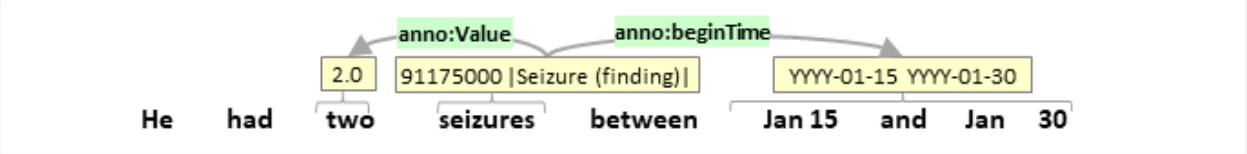


For very frequent conditions there are observable concepts that take an age or a time as a value, such as 228488005 |Age at starting smoking (observable entity)| . For the sake of annotation homogeneity we do not use them, instead

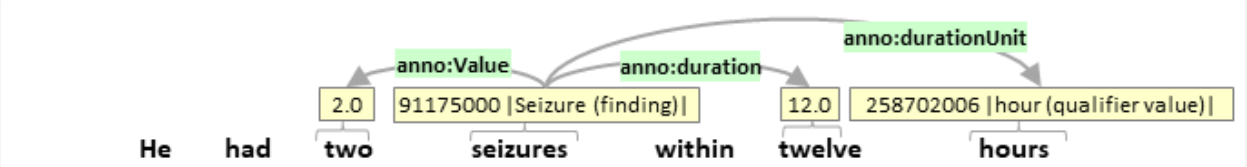


Time intervals are be expressed by incomplete time (e.g. “2022-01” means at some time between first and 31st of January 2022) if possible, otherwise or by the coordination of the time boundaries (“2023-11-12 2023-12-01” means at some time between twelfth of November and first of december of 2023). Note that time intervals do not

refer to anything that begins and ends at the interval bounds. For this “beginTime” and “endTime” is used separately.



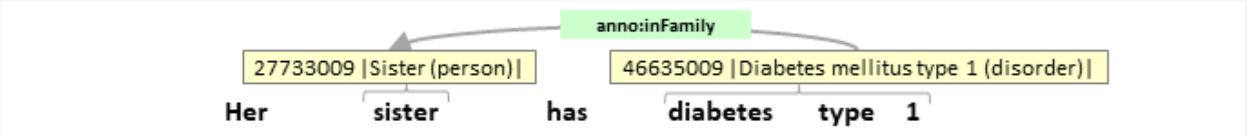
Date information that have incomplete years or months are filled with “YYYY” or “MM”, respectively.

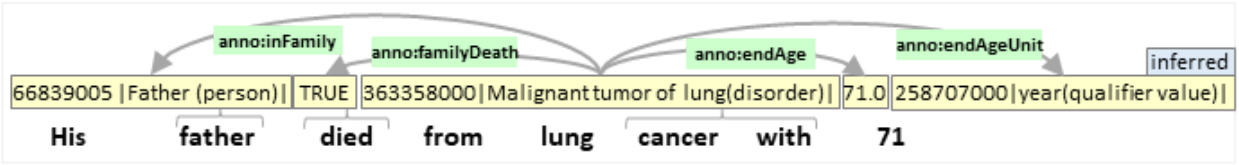


Finally, a duration that is not given by the delineating time points uses the predicates anno:duration and anno:durationUnit. SNOMED concepts that already include a time, but without referring to numbers and unit concepts, such as “123035007 |12 hours (qualifier value)|” are not used.

5.4.15. Family History

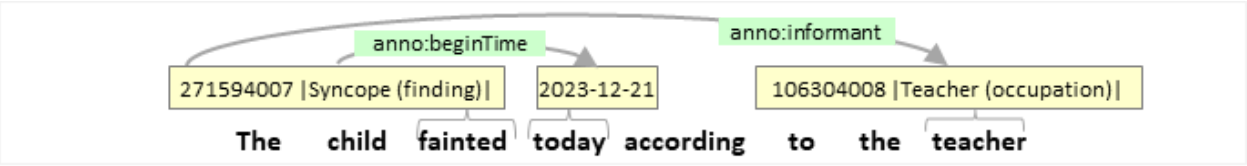
In the family history, the family role is related to the finding, disorder, event or procedure by anno:inFamily, if the family member died from that disease, we add a Boolean value with anno:familyDeath.





5.4.16. Informant

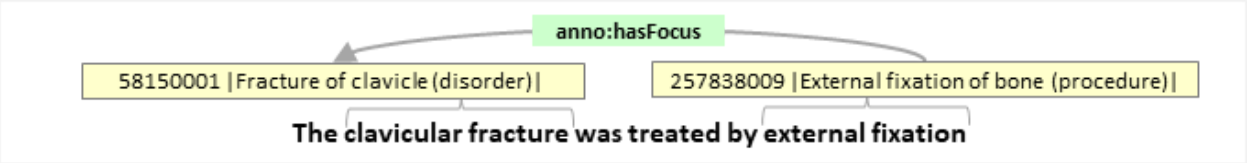
It is sometimes important to specify the person that informs about something that happened.



Occupation role concepts (e.g., Teacher occupation) are accepted as proxies for persons (e.g., teacher).

5.4.17. Treatment focus

With anno:hasFocus a treatment is linked to the disorder it treats.



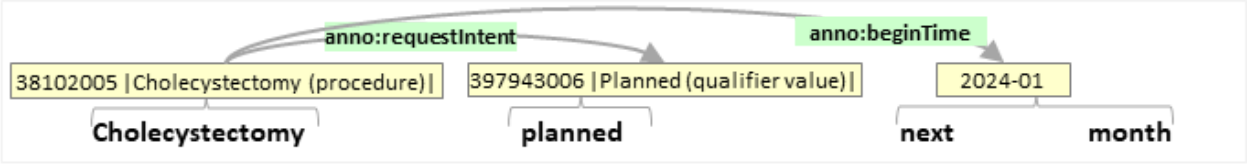
Although there is a more specific concept 448566002 |External fixation of clavicle (procedure)| we do not need to infer this in the annotation process because the anatomy is given by the related disorder concept.

5.4.18. Roles

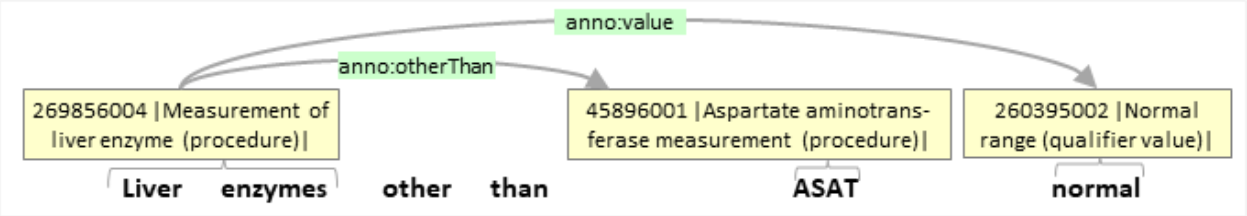
A role that is always implicitly present in a clinical document is that of the patient him/herself. It is the very nature of a clinical document that everything described therein is about exactly one subject of care, with the exceptions of obstetrics documents which refer to two or more individuals.
It is common to refer to "the patient" or to their name in a clinical document. This does not need to be annotated. An exception holds in cases where additional roles are assigned to the patient. Also, if other persons are mentioned in the document, such as accompanying persons or health professionals. The family relationship always points to the patient, an additional link is not needed.

5.4.19. Context specific words

Frequent context specific words in clinical documents refer indirectly to dates ("today", "tomorrow") or to the complement of a list of items ("otherwise"). In the first case we infer the data if the document contains it, otherwise it will be skipped. The following annotation can be done if the document has a creation date in December 2023.



"Otherwise" is expressed by anno:otherThan.

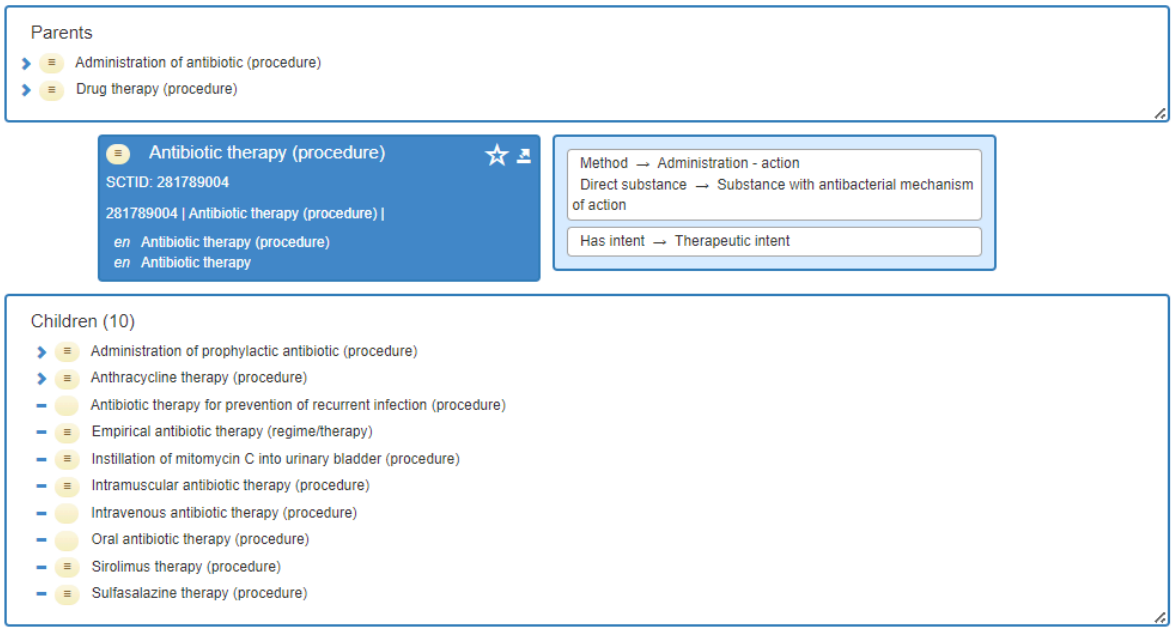


5.4.20. Special elements in text (Headings, enumerations, tables)

(...)

a clear description for the annotation and linkage of headings is necessary. for forms as well as for free text subchapters such as: Medikationen, Antiobiotika. A important hint to the SCT concedpt relations is also necessary

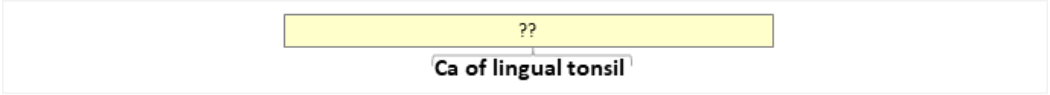
für Medikation muss “Administration of antibiotic” gemacht werden, da “antibiotic therapy” eine relation zu “has therapy intent” hat, was hier nicht definiert ist und die Therapie Prophylaktisch sein kann wie bei Weißheitszahn-OP



5.5. Dealing with ill-defined SNOMED CT content

Numerous SNOMED CT concepts lack clear definitions. Their meaning often can only be derived from the meaning of the parents and childrens, as well as from the FSN and the synonyms. Therefore, several soft criteria are suggested to ensure consistency:

1. **Prefer the more general concept:**



There are two candidates:

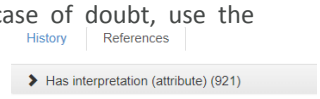
- 363377003 |Malignant tumor of lingual tonsil (disorder)|
- 254423005 |Carcinoma of lingual tonsil (disorder)|

  The upper one is a parent of the lower one, therefore the upper one is preferred

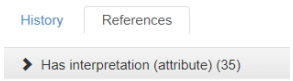
2. **Follow similar cases:**
   In the same case, the lookup of similar terms with “Ca” (e.g., “CA - Cancer of tonsil”, a synonym of 363393007 |Malignant tumor of tonsil (disorder)|) clearly shows that “CA” means neoplasm in general rather than carcinoma.
3. **Stick to most literal translation:**
   If it is not clear whether the German “Wunde” should be matched with “Injury”, “Trauma”, or “Wound”, choose “Wound” as the closest translation.
4. **Analyse the semantics:**
   SNOMED CT offers the siblings:
   260370003 |Decrease (qualifier value)| - 1250004 |Decreased (qualifier value)| -- 260371004 |Decreasing (qualifier value)|
   The first can be considered the more general one (infinitive), so you might apply rule 1, but the choice of the second or third one can be justified when it is clear that something has finished (past participle) or is ongoing (present participle)
5. **Follow the crowd:**

Particularly qualifier values are used in many pre-coordinated definitions. In case of doubt, use the concept that is more often used. Click on “References” in the SNOMED browser.

Example: for 2667000 |Absent (qualifier value)| the browser shows:



for 260385009 |Negative (qualifier value)| the browser shows:

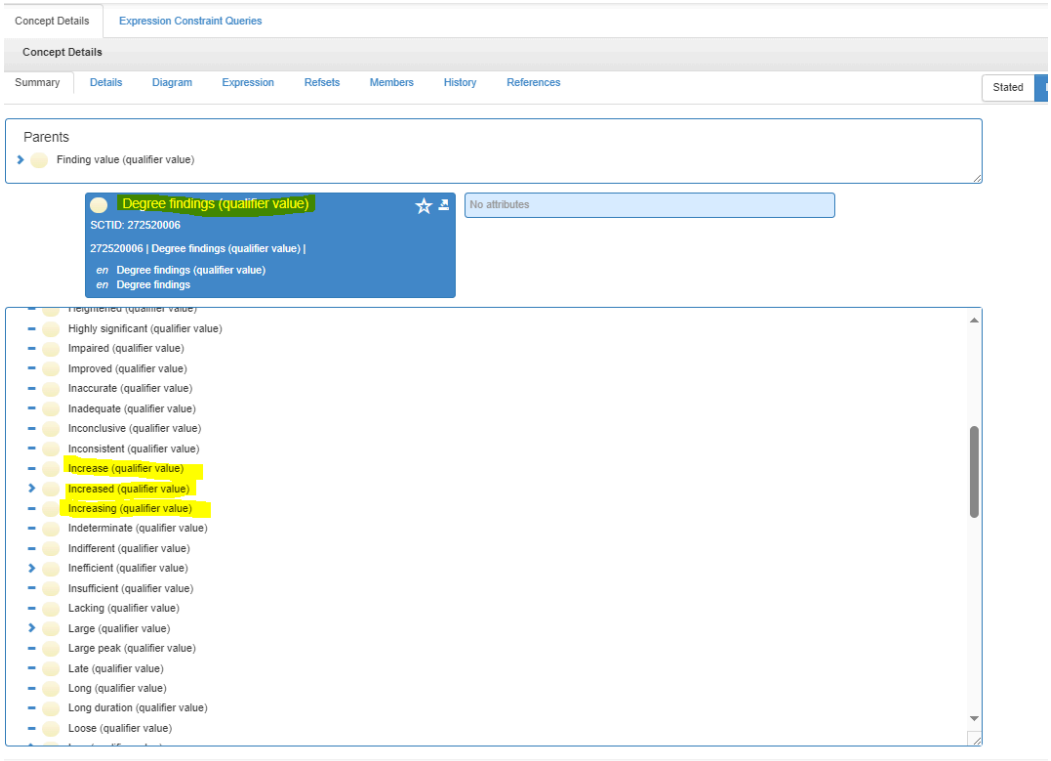


Expanding the list shows where the concept under scrutiny is used, which may facilitate the decision.

How to deal with Lack of terminology precision

Unscharfe / unklare Abgrenzungen von SNOMED-Konzepten → Priorisierungsregeln in Guideline nötig → priorisation rules necessary

Gleichwertige SNOMED-Varianten, denselben Inhalt unterschiedlich zu annotieren



6. Exemplification for specific annotation use cases in AIDAVA

This section shows how the AAG can be instantiated and applied to BC use case and CVD use case by examples. However, new examples encountered during the annotation process can provide valuable insights and help identify areas where the AAG can be enhanced. Therefore, the AAG (including the principles, rules, instructions, and introduced predicates) is requested to be refined and expanded within the project by incorporating more data.

6.1. Assumptions

Several assumptions have been formulated in the AAG that are considered essential to be followed throughout the entire annotation process. These assumptions are outlined below.

- The default subject throughout the entire process is the patient (e.g., the subject of care) and there is no need to annotate it.
- Conditional, hypothetical and imperative expressions, as well as questions, should not be annotated.
- The default value for the presence of a finding is known (yes) and no need to annotate it (the default values of the introduced predicates are shown in the 'default' column of the Table 6).

6.2. Process template

In the following, the instructions mostly related to the BC use case and CVD use case are summarized.

- The steps to annotate data related to situation concepts include:
    a. Identify the subject of the record, which may be the patient or a relevant family member (e.g., mother) and then annotate it only if it refers to the family member (refers to Section 6.4.2).
    b. Identify the clinical finding.
    c. Connect the subject (detected in step a) to the clinical finding using the 'inFamily' predicate.

       If any Protected Health Information (PHI) [14] persists despite the data de-identification procedure outlined in the first version of this guideline (Deliverable D4.1 – Section 4.2), annotators are requested to promptly report such instances to the designated local data steward responsible for the de-identification process[8].

    d. Identify a qualifier value indicating the presence of the clinical finding (i.e., present, absent, or unknown) if applicable (according to Section 6.4.2, if the presence of a finding is already 'known', there is no need to annotate it).
    e. Connect the clinical finding to the qualifier value determined in Step d using the 'verificationStatus' predicate.
    f. Identify a qualifier value indicating the temporal context of the clinical finding (i.e., current, past, current and past, or unknown) if applicable.
    g. Connect the clinical finding to the qualifier value determined in step f using the 'clinicalStatus' predicate.
- To annotate certain findings such as smoking behaviors in detail, the following steps[9] need to be followed:
    a. Annotate the date or duration if there is any mention in the text according to the following Table.

       Table 7: smoking behavior

[8] This ensures that appropriate measures can be taken to address and rectify any remaining PHI to safeguard data privacy and compliance.

[9] However, while it is possible to create general instructions for annotating any clinical finding, they are not applied in the AIDAVA use case as the focus is to utilise SNOMED CT as extensively as possible.

Pattern in Text	SNOMED CT Concept
age at start	Age at starting smoking (observable entity)
age at stop	Age at stopping smoking (observable entity)
start n time-unit ago	Time since started smoking (observable entity) << 258700003 |Non-International System of Units unit of time (qualifier value)|
stop n time-unit ago	Time since stopped smoking (observable entity) << 258700003 |Non-International System of Units unit of time (qualifier value)|
start in the year1	Date of onset (observable entity) **AND** Tobacco use and exposure
stop in the year2	Date ceased smoking (observable entity)
time-unit duration	Total time smoked (observable entity) << 258700003 |Non-International System of Units unit of time (qualifier value)|

b. Annotate the smoking quantity unit according to the following table:

Table 8: Quantity units for smoking behavior

Quantity Unit	SNOMED CT Concept
pack	Pack (physical object)
cigarette	Cigarette (physical object)

c. Annotate the denominator of time regarding the amount of smoking using << 282363004 |Denominators of time (qualifier value)|
d. Annotate the values numbers for duration, years, and smoking quantity
e. Use the introduced relations in Table 5 to establish links among the annotated spans (i.e, unit, value, valueLow, valueHigh, and so on)

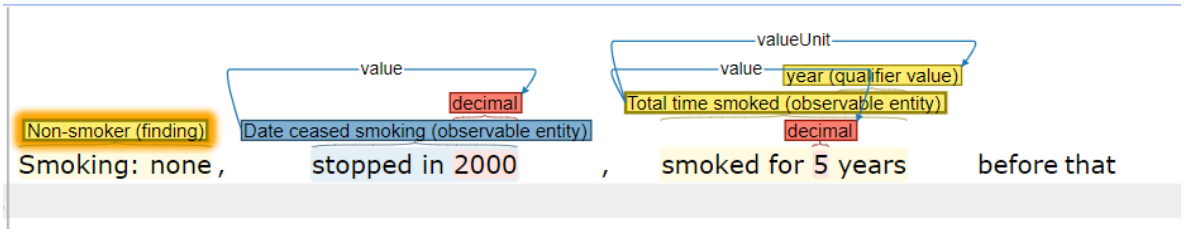
6.3. Annotation examples using INCEpTION

In this section, we will illustrate the annotation process through several examples that demonstrate the application of the guideline. Each subsection will focus on a specific example from CVD or BC and provide a detailed explanation of the approach and results. By following these examples, annotators will gain a better understanding of how to apply the instructions described in Section 6.1. to their own documents. The template used to organise the examples induces the input text, a screenshot of the annotations made using INCEpTION, and a description of how the general manual annotation principles are applied. Examples are an important resource for annotator training. During the annotation process, annotators and trainers will collect more examples, both on English texts (for better mutual discussion) and increasingly in their own language.

6.3.1. Smoking behaviour

Table 10: Example of smoking behaviour

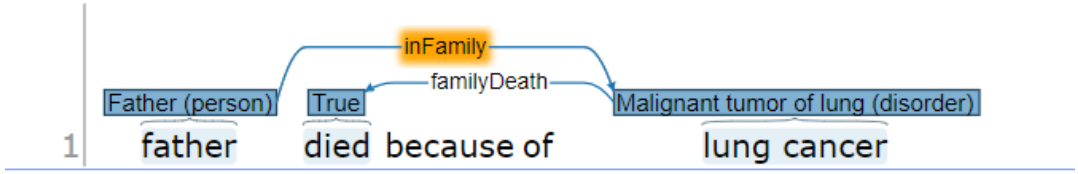
Input	Smoking: none, stopped in 2000, smoked for 5 years before that
INCEpTION screenshot	



Adjudication Description	Two predicted patterns (stop at a specific date, and time duration) in smoking data appear in the input text. Therefore, instructions X and Y are followed to annotate 'stopped in 2000' and 'smoked for 5 years' as 'Date ceased smoking' and 'Total time smoked', respectively. Also, the numbers 2000 and 5 were coded as decimals. Moreover, the time-unit needs to be identified (as instructed in Section 6.2), so 'years' correspond to 'year (qualifier value)'. As shown in Table 5, the predicate 'value' has observable entity and decimal as its domain and range, respectively. Thus, we use the predicate 'value' between the identified observable entities and their corresponding decimals. The same explanation applies to 'unit' between the observable value and the qualifier value (i.e., year).

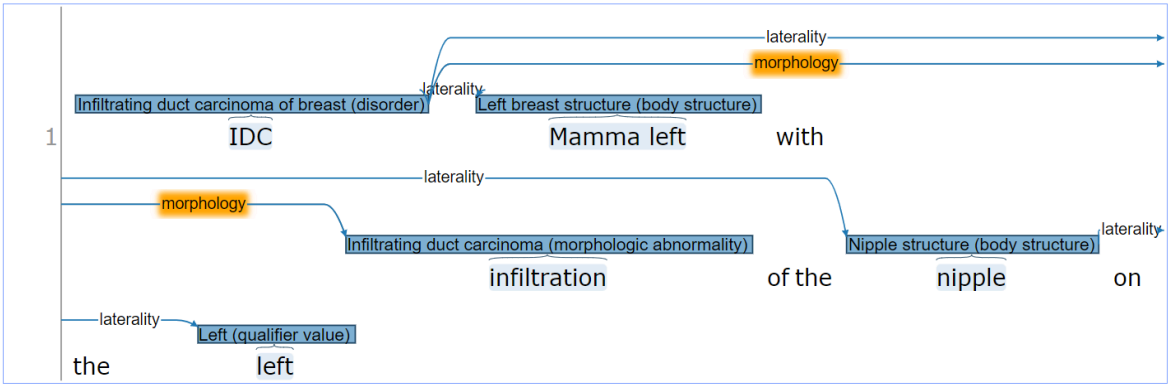
6.3.2. Family history

Table 11: Example of family history

Input	father died because of lung cancer
INCEpTION Screenshot	
Adjudication Description	Since the reference is to a family member (not the main patient), it should be annotated with <<303071001 |Person in the family (person)|, which in this case is the father. Then, the predicate 'inFamily' is used to establish a link between the disorder (i.e., lung cancer) and the person according to Table 5. As the range of the predicate 'familyDeath' is a boolean value (True or False), the term 'died' was marked as True, and then the predicate was established from the lung cancer to True (as per Instruction Family History in Section 5.4.14).

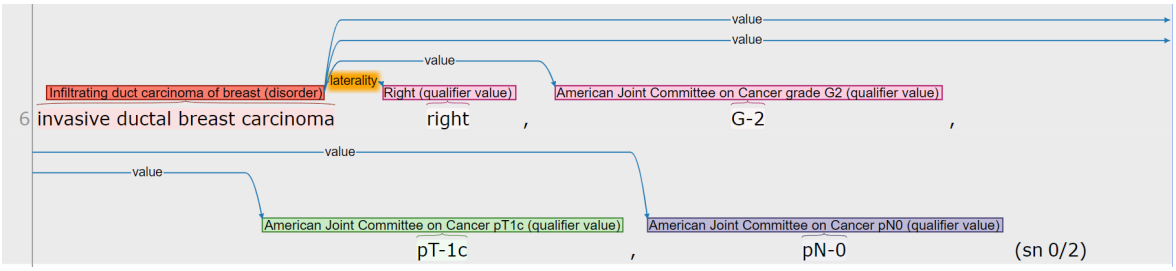
6.3.3. Histology

Table 12: Example of histology

Input	IDC Mamma left with infiltration of the nipple on the left
INCEpTION Screenshot	
Adjudication Description	According to Section 5.3.1 and Section 6.2, the core concepts and their corresponding qualifier values are coded. To establish links among the identified concepts, Table 5 is used to find appropriate predicates between the concepts based on the domain and range of the specified predicates. For example, the predicate 'morphology' is used to connect a clinical finding (i.e., IDC) with a morphologic abnormality (i.e., infiltration).

6.3.4. TNM stage

Table 13: Example of TNM stage

Input	invasive ductal breast carcinoma right, G-2 pT-1c, pN-0(sn 0/2)
INCEpTION Screenshot	
Adjudication Description	The span 'G-2 pT-1c, pN-0(sn 0/2)' is clearly composed by parts that are separated by comma and each part needs to be coded separately (Section 6.2). Thus, G-2, pT-1c, and pN-0 are normalised with the corresponding concepts. According to Table 5, to associate a disorder with a side qualifier value, the predicate 'laterality' should be used. Similarly, the

	predicate 'value' is used to establish a link from the disorder to the qualifier values indicating TNM staging.

Table 14: Example of TNM stage

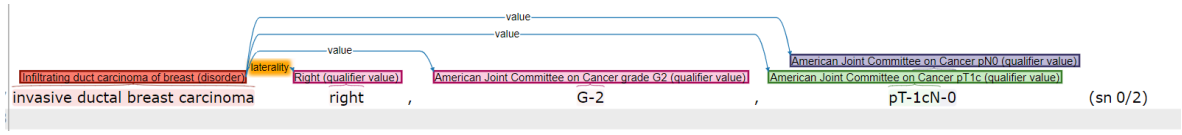
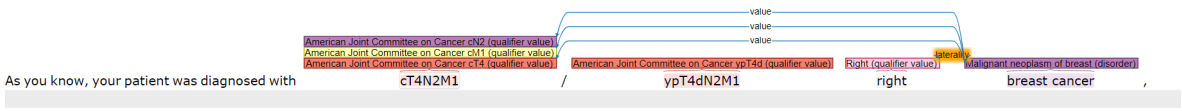
Input	invasive ductal breast carcinoma right, G-2 pT-1cN-0(sn 0/2)
INCEpTION Screenshot	
Adjudication Description	As stated in Section 5.2, annotations on a subword level are permitted. Therefore, 'N-0' is coded with '1229947003 |American Joint Committee on Cancer pN0 (qualifier value)|'. Additionally, the predicate 'value' is employed to establish relationships between a disorder and qualifier values based on Table 5.

Table 15: Example of TNM stage

Input	As you know, your patient was diagnosed with cT4N2M1/ypT4dN2M1 right breast cancer,
INCEpTION Screenshot	
Adjudication Description	Similar to the two previous examples explained, 'CT4N2M1' can be annotated at the subword level (Section 5.2). Furthermore, the predicate 'value' with the domain of clinical finding and the range of qualifier values (Table 5) is used to establish a link between the breast cancer disorder and the identified qualifier values.

6.3.5. Example therapy

Table 16: Example of therapy

Input	for which neo-adjuvant chemotherapy, ablatio and axillary lymph node dissection have already been done.

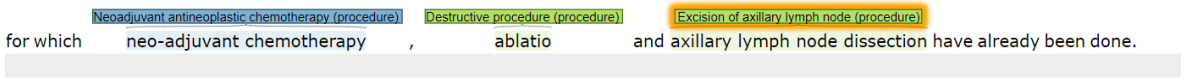
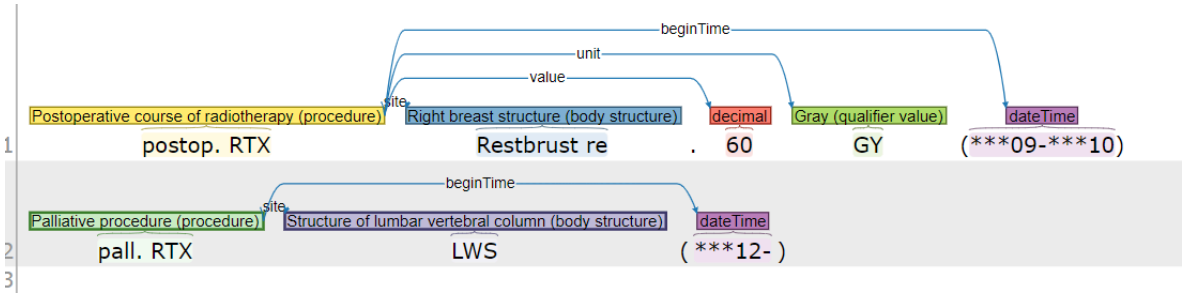
INCEpTION Screenshot	
	
Adjudication Description	According to Table 3, procedures (considered core concepts) and substances (if there is no appropriate product) must be identified and coded.

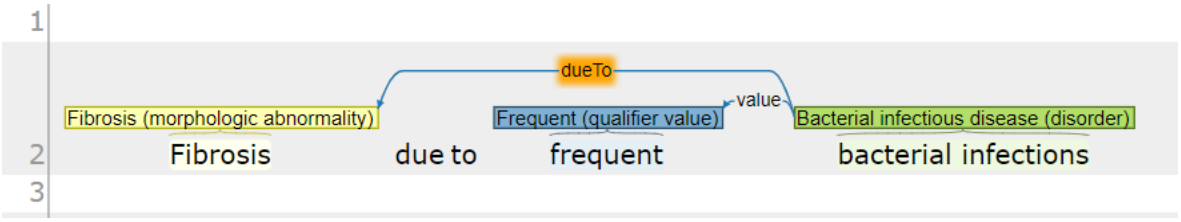
Table 17: Example of therapy

Input	postop. RTX Restbrust re. 60 GY (***09-***10) pall. RTX LWS (***12-)
INCEpTION Screenshot	
Adjudication Description	All abbreviations are annotated using the appropriate concepts (it is recommended to use Google to determine the meaning of an abbreviation). Additionally, Table 5 is referenced to establish predicates between concepts based on the domain and range of the predicates. For instance, 'beginTime' is used to connect a procedure to a dateTime.

6.3.6. Common examples

Table 18: Common example

Input	Fibrosis due to frequent bacterial infections
INCEpTION Screenshot	



Adjudication Description	According to Table 3, the mentioned concepts of core concepts as well as qualifier values are identified and coded. To assert a predicate from a disorder to a morphologic abnormality, the predicate 'deuTo' is allowed to be applied. Moreover, the detected disorder is linked to the qualifier value using the predicate 'value' according to Table 5.

Table 19: Common example

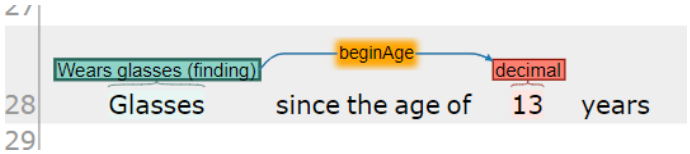
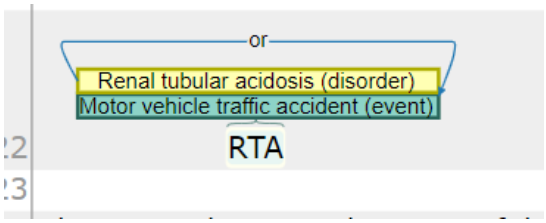
Input	Glasses since the age of 13 years
INCEpTION Screenshot	
Adjudication Description	As stated in Section 5.4.5, to handle ellipses, annotators must utilize the most suitable concepts. Therefore, in this case, 'Glasses' is coded as 'Wears glasses'. The decimal value representing the age is annotated as a decimal, and then the predicate 'beginAge' is employed to connect the finding to the decimal, as specified in Table 5.

Table 20: Common example

Input	RTA
INCEpTION Screenshot	

Adjudication Description	As discussed in Section 5.4.3, in cases of ambiguities, despite the principles and expertise of the annotators, the predicate 'OR' is applied to connect candidate concepts of the ambiguous span.

Table 21: Common example

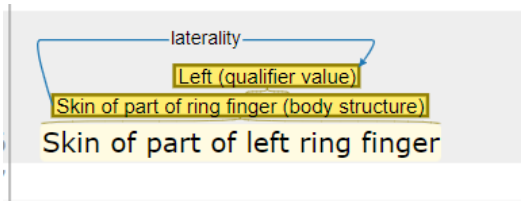
Input	Skin of part of left ring finger
INCEpTION Screenshot	
Adjudication Description	Annotators must consider the most detailed annotations throughout the entire process, as explained in Section 5.4.1. Therefore, in this example, the most detailed concept is used to annotate 'Skin of part of the ring finger', and then the predicate 'Left' is employed to assert a predicate between the detected body structure and its corresponding side, which is the left side.

Table 22: Common example

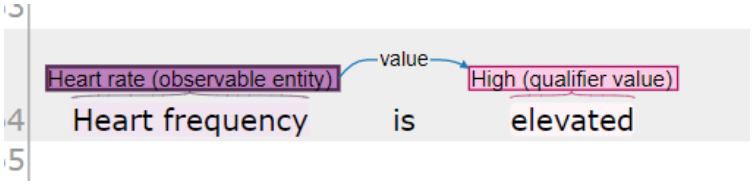
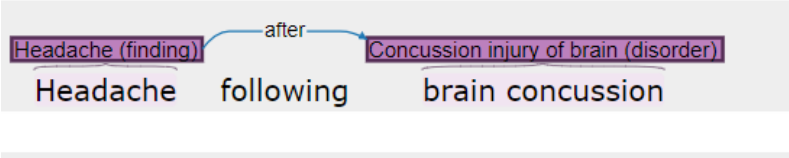
Input	Heart frequency is elevated
INCEpTION Screenshot	
Adjudication Description	According to Table 5, the predicate 'value' is used with the domain and range of observable entities and qualifier values to assert a predicate between the detected concepts.

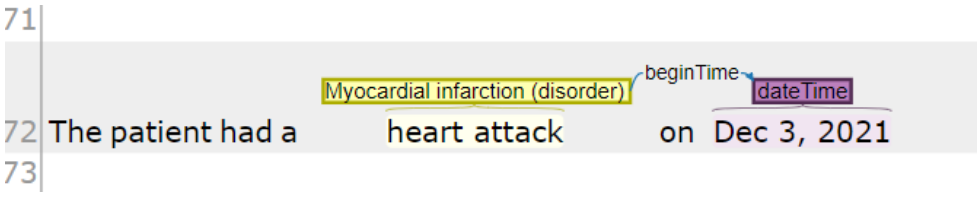
Table 23: Common example

Input	Headache following brain concussion
INCEpTION Screenshot	



Adjudication Description	As discussed in Section 5.2, annotators are not permitted to infer causality. Therefore, in this example, the only applicable predicate is a 'after'.

Table 24: Common example

Input	The patient had a heart attack on Dec 3, 2021
INCEpTION Screenshot	
Adjudication Description	According to the assumption in Section 6.1, there is no need to annotate the patient. Furthermore, the 'beginTime' is used to establish a link between the disorder and dateTime, as stated in Table 5.

7. Annex 1 - Concept tag triples

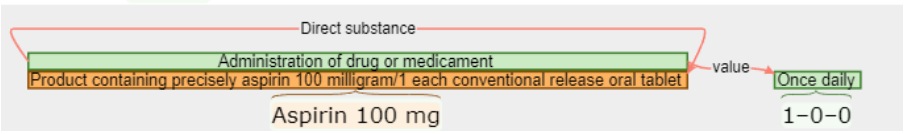
Domain concept tag	Relation	Range concept tag	Frequency in axioms
'procedure'	'Access'	'qualifier value'	1708
'disorder'	'After'	'disorder'	859
'disorder'	'After'	'finding'	21
'disorder'	'After'	'procedure'	1491
'disorder'	'After'	'regime/therapy'	8
'finding'	'After'	'procedure'	122
'situation'	'Associated finding'	'disorder'	1591
'situation'	'Associated finding'	'event'	79
'situation'	'Associated finding'	'finding'	660

'disorder'	'Associated morphology'	'cell'	77
'disorder'	'Associated morphology'	'morphologic abnormality'	66589
'finding'	'Associated morphology'	'cell'	12
'finding'	'Associated morphology'	'morphologic abnormality'	2405
'situation'	'Associated procedure'	'procedure'	2396
'disorder'	'Associated with'	'disorder'	485
'disorder'	'Associated with'	'finding'	77
'disorder'	'Associated with'	'physical object'	523
'disorder'	'Causative agent'	'organism'	7192
'disorder'	'Causative agent'	'physical force'	754
'disorder'	'Causative agent'	'physical object'	298
'disorder'	'Causative agent'	'substance'	7929
'event'	'Causative agent'	'physical force'	61
'event'	'Causative agent'	'substance'	21
'finding'	'Causative agent'	'organism'	73
'finding'	'Causative agent'	'physical force'	31
'finding'	'Causative agent'	'physical object'	13
'finding'	'Causative agent'	'substance'	1398
'procedure'	'Component'	'cell structure'	12
'procedure'	'Component'	'cell'	193
'procedure'	'Component'	'organism'	205
'procedure'	'Component'	'substance'	7934
'procedure'	'Direct device'	'physical object'	4374
'regime/therapy'	'Direct device'	'physical object'	29
'procedure'	'Direct morphology'	'morphologic abnormality'	8752
'procedure'	'Direct substance'	'substance'	4991
'regime/therapy'	'Direct substance'	'substance'	355
'disorder'	'Due to'	'disorder'	4564
'disorder'	'Due to'	'finding'	372
'disorder'	'Finding site'	'body structure'	79195
'disorder'	'Finding site'	'cell structure'	1132
'disorder'	'Finding site'	'cell'	394
'disorder'	'Finding site'	'morphologic abnormality'	432
'finding'	'Finding site'	'body structure'	12647
'finding'	'Finding site'	'morphologic abnormality'	44
'medicinal product form'	'Has active ingredient'	'substance'	8541
'medicinal product'	'Has active ingredient'	'substance'	10879
'product'	'Has active ingredient'	'substance'	1160
'procedure'	'Has focus'	'disorder'	1431
'procedure'	'Has focus'	'finding'	581
'procedure'	'Has focus'	'procedure'	1139
'procedure'	'Has focus'	'regime/therapy'	921
'regime/therapy'	'Has focus'	'disorder'	236
'regime/therapy'	'Has focus'	'finding'	284
'regime/therapy'	'Has focus'	'procedure'	132
'regime/therapy'	'Has focus'	'regime/therapy'	71

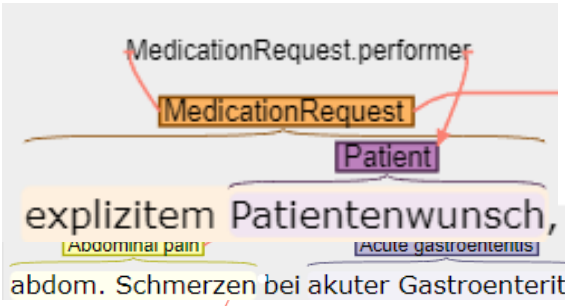
'procedure'	'Has intent'	'qualifier value'	3885
'regime/therapy'	'Has intent'	'qualifier value'	1158
'finding'	'Has interpretation'	'qualifier value'	8886
'clinical drug'	'Has manufactured dose form'	'dose form'	7677
'procedure'	'Has specimen'	'specimen'	3062
'regime/therapy'	'Has specimen'	'specimen'	1
'procedure'	'Indirect morphology'	'morphologic abnormality'	593
'disorder'	'Interprets'	'observable entity'	6277
'finding'	'Interprets'	'observable entity'	22873
'finding'	'Interprets'	'procedure'	5766
'body structure'	'Laterality'	'qualifier value'	19437
'cell structure'	'Laterality'	'qualifier value'	15
'cell'	'Laterality'	'qualifier value'	17
'morphologic abnormality'	'Laterality'	'qualifier value'	7
'procedure'	'Measurement method'	'procedure'	15
'procedure'	'Method'	'qualifier value'	63769
'regime/therapy'	'Method'	'qualifier value'	1237
'disorder'	'Occurrence'	'qualifier value'	19349
'event'	'Occurrence'	'qualifier value'	32
'finding'	'Occurrence'	'qualifier value'	1560
'disorder'	'Pathological process'	'qualifier value'	23699
'procedure'	'Priority'	'qualifier value'	218
'regime/therapy'	'Priority'	'qualifier value'	3
'procedure'	'Procedure device'	'physical object'	254
'procedure'	'Procedure morphology'	'morphologic abnormality'	149
'procedure'	'Procedure site - Direct'	'body structure'	35683
'procedure'	'Procedure site - Indirect'	'body structure'	8538
'procedure'	'Procedure site'	'body structure'	3055
'regime/therapy'	'Procedure site'	'body structure'	195
'procedure'	'Property'	'qualifier value'	229
'procedure'	'Recipient category'	'occupation'	1
'procedure'	'Recipient category'	'person'	36
'procedure'	'Recipient category'	'social concept'	40
'regime/therapy'	'Recipient category'	'person'	6
'regime/therapy'	'Recipient category'	'social concept'	51
'procedure'	'Revision status'	'qualifier value'	1309
'procedure'	'Scale type'	'qualifier value'	222
'specimen'	'Specimen procedure'	'procedure'	983
'specimen'	'Specimen source identity'	'person'	13
'specimen'	'Specimen source identity'	'physical object'	32
'specimen'	'Specimen source morphology'	'morphologic abnormality'	163
'specimen'	'Specimen source topography'	'body structure'	1395
'specimen'	'Specimen source topography'	'cell'	17
'specimen'	'Specimen substance'	'substance'	946
'situation'	'Subject relationship context'	'person'	4798

8. Annex 2 - German examples in inception

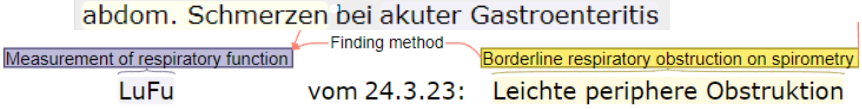
[Regime] if there is a administration regime like 1-0-0 or 0-0-1 mark the term as in “Layer” and copy and paste a SNOMED ID (like Once daily (qualifier value)|). Connect it with the medication concept with “Relation” and define it with the FHIR ID “ Quantity.value” short “value”.



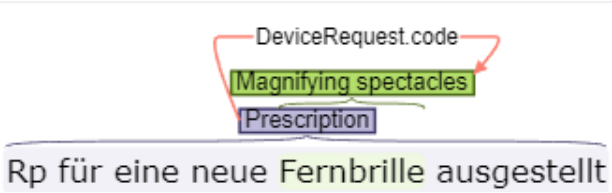
[Medication] whenever there is a SNOMED-ID with administration use it. If there is just the substance, mark the word and add a Layer. In the second layer you copy and paste 386359008 |Administration of drug or medicament via oral route (procedure)| and in the first the SNOMED ID for the substance. Connect both with “relation” and define it with “363701004 |Direct substance (attribute)|”.



[Finding method]- if there is a finding and its finding method (procedure)- mark the finding as in [Layer] and copy and paste the SNOMED ID. Then mark the finding method and copy and paste the SNOMED ID. Connect the two concepts with “Relation” and define it with the SNOMED ID “finding method”.



[DeviceRequest.code] if there is a device request e.g. “a device request for glasses”, mark the device as in [Layer] and mark the request as in [Layer]. Connect both concepts with [Relation] and define it with “DeviceRequest” for the short term and with the FHIR ID “DeviceRequest.code” for the long term.



[ServiceRequest.category] if there is a service request e.g. “the service of an otorhinolaryngologist was requested (Überweisung) “ mark the requested service as in [Layer] and mark the request e.g. “germ: Überweisung” as in [Layer] e.g. “103320006 |Request for (contextual qualifier) (qualifier value)|”. Connect both concepts with [Relation] and define it with “ServiceRequest” for the short and with the FHIR ID “ServiceRequest.category” for the long term.

[PlanDefinition.action.code] - geplante Prozeduren
[goal.description.CodeableConcept.coding] Empfehlungen für Lifestyle (Gewicht, Rauchen, Sport) laut FHIR wir auch Empfehlung

[MedicationRequest.performer] if there is a person with a request for a medication- mark the medication request and use the FHIR ID “MedicationRequest” and mark the person who is requesting

use the SNOMED ID as in Layer. Connect both concepts with “Relation” and define it with the FHIR ID “MedicationRequest.performer”. Connect the “MedicationRequest” with the medication which is requested. Use “Relation” and define it with
“363701004 |Direct substance (attribute)|”.

9. Annex 3 Limitations and workarounds

9.1. Planned procedure:

“NTx geplant”
“NTx” annotated with “70536003 |Transplant of kidney (procedure)|”
“Geplant” annotated with “405613005 |Planned procedure (situation)|”
“70536003 |Transplant of kidney (procedure)|” --- [PlanDefinition.action.code] --- > “405613005 |Planned procedure (situation)|”

Potentially refine by purpose (indication) and goal (intended state after the action)

9.2. Conditional recommendations (like in clinical guidelines):

“conditional” : if… then … else
Not represented
goals vs. plan:

Clinical narrative does not reveal all background discussions / decisions: therefore our baseline is:
Goal: if a specified state/condition of the patient shall be achieved
Plan: if a specified intervention (diagnostic / therapeutic) is planned / scheduled (PlanDefinition.action)

9.3. Ambiguities:

Example “patient was recommended to seek therapy by community surgery service”
PlanDefinition.action vs goal
Better: ServiceRequest.category -> FHIR points to surgical procedure in SNOMED ?
275146006 |Refashioning of ingrowing toenail (procedure)|
Rule of thumb: choose the FHIR resources that require the least that you have to take decisions not grounded in the text
“Lesen vom Papier wurde vom Facharzt zu Übungszwecken empfohlen” ? (Schielaug S 11 Z17)

10. Annex 5

If possible, only those SNOMED relations should be used that also occur in SNOMED concept definitions (below *Concept model object attribute*). In case of doubt about which relation to choose, look up similar concepts in SNOMED CT and follow the pattern they follow. For a complete list with domain and range restrictions see Annex A. Apart from SNOMED relations, we recommend the use of the FHIR relational elements as follows
(tentative mappings to SNOMED)

Note that here the way of how these attributes are used in annotations correspond to a more complex representation in FHIR or SNOMED:

Example:

↙ verificationStatus ↖	
refuted	22298006 |Myocardial infarction (disorder)|
No	**myocardial infarction**

Or

↙ 408729009 |Finding context (attribute)| ↖	
410594000 |Definitely NOT present (qualifier value)||	22298006 |Myocardial infarction (disorder)|
No	**myocardial infarction**

Both can be translated into:

SNOMED CT concept model

413350009 |Finding with explicit context (situation)| :
{ 246090004 |Associated finding (attribute)| = 22298006 |Myocardial infarction (disorder)|,
408729009 |Finding context (attribute)| = 410516002 |Known absent (qualifier value)|}

FHIR

Condition.verificationStatus = Refuted
Condition.code = 22298006 |Myocardial infarction (disorder)

11. Annex 5 - Abstract submitted to SNOMED EXPO 2023:

May 3, 2023: abstract submitted to SNOMED EXPO 2023:

The need for standards-based annotation guidelines in times of large language models

1 -- Scope

Most EHR content is in narrative form and uses a highly compact, idiosyncratic language [1]. Document retrieval, information extraction and summarization about patients and cohorts is still an unsolved problem, despite efforts and advances in Natural Language Processing (NLP).
With the recent progress of artificial intelligence (AI) and the advent of large language models such as GPT-x [2], there is an increased expectation that the gap between human language and computational representation will eventually be bridged. This also includes the mapping from text to standardized representations such as supported by SNOMED CT and FHIR.

It is more urgent than ever that evaluation scenarios be devised, to assess the quality of AI-powered information extraction. This requires benchmarks that have proved their quality in inter-annotator studies. Because EHRs use the official language of their jurisdiction, such benchmarks need to be created for many natural languages. The expressiveness of human language poses problems to the target representation, which should ideally be identical for all linguistic renderings of a given clinical state of affairs .

We present a standards- and ontology-based annotation guideline [3] for clinical narratives. Its goal is to support the creation of annotated corpora for training and evaluation of clinical NLP systems. It is currently undergoing formative evaluation in the EU project AIDAVA [4] and will be adapted to the German annotation initiative GemTeX [5].

2 -- How SNOMED CT is used in the work

Our Annotation Guideline is committed to the creation of a canonical form of representing clinical narratives, which follows as much as possible existing specifications for structured EHR content. This means a strong commitment to SNOMED CT as a reference ontology, which provides identifiers for all types of entities referred to by clinical narratives, together with FHIR as an information model which provides standardized context for these entities.
The use of these two standards requires the agreement on high-level annotation principles, a selection of which is presented here:

- The granularity of annotation spans is not given by a named entity recognition step prior to annotation, yielding entity types, such as “disorder” and “body part”, in the case of “fracture of skull”. Instead, the principle of longest match is followed and, as a consequence, pre-coordinated concepts are used whenever possible.

- The annotation strategy is descriptive and not interpretative. This means that annotators annotate only what they read, without interpretation. An exception is the disambiguation of acronyms as long as their meaning can be derived from the context.

- Ambiguities in terminology content are mitigated by giving preference to the “core” hierarchies Clinical finding, Event, Observable entity, Pharmaceutical / biologic product, Procedure, Specimen. E.g., “Hodgkin's disease (disorder)” is given preference over “Hodgkin lymphoma (morphologic abnormality)”.

- Pre-coordinated content of the hierarchy “Situation in specific context” is not used, because FHIR has shown to be more granular, actively maintained and frequently used to represent context.

- Entities are linked by a predefined set of binary relations. To this end, a set of predicates was introduced, mapped to (i)

SNOMED CT linkage concepts or chains thereof, (ii) to relational chains of FHIR elements or (iii) both. E.g., the predicate "site" between a SNOMED CT clinical finding and a body structure, is mapped to the linkage concept “finding site” as well as the concatenation of the inverse of the FHIR element "Condition.code" with "Condition.bodySite".

- SNOMED CT mappings to HL7 value sets are proposed. E.g., in FHIR Condition, the value “Recurrence” is mapped to 255227004 |Recurrent (qualifier value)| in SNOMED CT, and the value “Refuted” is mapped to both SNOMED CT concepts 410594000 |Definitely NOT present (qualifier value)| and 410516002 |Known absent (qualifier value)|.

3 -- Why SNOMED CT was selected for this work
Past clinical annotation projects were often based on UMLS CUIs [6], as freely accessible concept identifiers. In other cases, annotations were limited to entity types, such as “disorder” and “body part”, with a focus on relations [7]. Our choice of SNOMED CT is its international acceptance as a standard, its availability to the research community, as well as its scope and granularity, and particularly its logical underfitting, which facilitates the bridging between pre-coordinated and post-coordinated expressions.

However, our approach also meets with reservations. It is argued that SNOMED CT is little used in routine, particularly in continental Europe, that current licenses exclude important countries, and translations are still missing. We reply that the status quo in clinical terminologies, with national ICD versions, national procedure classifications and drug catalogs, does not offer a convincing interoperability perspective without SNOMED CT.

One limitation is the still unresolved management of overlap between SNOMED CT, FHIR and related value sets. A continuation of the Terminfo work [8] in the light of FHIR would be desirable. Another limitation are SNOMED CT concepts that lack formal and textual definitions, and which pose challenges to annotators particularly with texts in languages for which no official translation exists.

We are convinced that in times where 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, standardized representations. Such representations as a gold standard for clinical content representation need to be elaborated and refined. We understand the proposed annotation guideline as a step in this direction.

4 -- References:
[1] Dash S et al. Big data in healthcare: management, analysis and future prospects. J Big Data. 2019, 19;6(1):54.
[2] Patel SB et al. ChatGPT: Friend or Foe. Lancet Digit. Health 5 (2023): e102.
[3] Schulz S. et al. Annotation guideline for semantic annotations of clinical narratives https://docs.google.com/document/d/1BQPL8sNIMorRb9qdvsZL0ckpmx2DILsZF6bewRduvWI/edit
[4] AIDAVA - AI-powered Data Curation https://www.aidava.eu/
[5] Boeker M. GeMTeX - German Medical Text Corpus (2022) https://www.gmds.de/fileadmin/user_upload/AG_MTK/2022-08-25_MP-GeMTeX_GMDS-AG-Terminologie__Boeker_.pdf
[6] Annotation Guidelines for Clinical Entity Normalization. Based on the annotation guidelines for ShARe/CLEF eHealth 2013 Shared Task
[7] Lohr C et al. Evolutionary Approach to the Annotation of Discharge Summaries. Stud Health Technol Inform. 2020, 16;270:28-32.
[8] TermInfo Project. http://www.hl7.org/Special/committees/terminfo/

Popular qualifier values:

2575	263654008 |Abnormal (qualifier value)|
2205	260379002 |Impaired (qualifier value)|
1967	90734009 |Chronic (qualifier value)|
1964	24028007 |Right (qualifier value)|
1949	7771000 |Left (qualifier value)|
1863	424124008 |Sudden onset AND/OR short duration (qualifier value)|

1810 410515003 |Known present (qualifier value)|
1754 410513005 |In the past (qualifier value)|
1687 281300000 |Below reference range (qualifier value)|
1490 281302008 |Above reference range (qualifier value)|
1241 385658003 |Done (qualifier value)|
1211 410511007 |Current or past (actual) (qualifier value)|
1048 123029007 |Single point in time (qualifier value)|
953 35105006 |Increased (qualifier value)|
923 2667000 |Absent (qualifier value)|
869 1250004 |Decreased (qualifier value)|
814 255407002 |Neonatal (qualifier value)|
771 371150009 |Able (qualifier value)|
692 371157007 |Able with difficulty (qualifier value)|
678 52101004 |Present (qualifier value)|
670 371151008 |Unable (qualifier value)|
660 17621005 |Normal (qualifier value)|
645 385640009 |Does (qualifier value)|
607 255314001 |Progressive (qualifier value)|
591 717896003 |Does not (qualifier value)|
516 262202000 |Therapeutic intent (qualifier value)|
473 30766002 |Quantitative (qualifier value)|
444 410516002 |Known absent (qualifier value)|
410 255227004 |Recurrent (qualifier value)|
395 261424001 |Primary operation (qualifier value)|
347 123027009 |24 hours (qualifier value)|
318 260373001 |Detected (qualifier value)|
306 255410009 |Maternal postpartum (qualifier value)|
299 281301001 |Within reference range (qualifier value)|
284 443390004 |Refused (qualifier value)|
241 360156006 |Screening - procedure intent (qualifier value)|
230 129428001 |Preventive - procedure intent (qualifier value)|
214 15240007 |Current (qualifier value)|
202 18307000 |Altered (qualifier value)|
200 14803004 |Transitory (qualifier value)|
199 1505281000004101 |Direct local invasion (qualifier value)|
193 41847000 |Adulthood (qualifier value)|
185 415684004 |Suspected (qualifier value)|
184 385315009 |Sudden onset (qualifier value)|
151 394844007 |Outside reference range (qualifier value)|
146 260415000 |Not detected (qualifier value)|
143 31874001 |True (qualifier value)|
142 25876001 |Emergency (qualifier value)|
130 385660001 |Not done (qualifier value)|
128 260378005 |Excessive (qualifier value)|
126 260392004 |Non-patent (qualifier value)|
119 360271000 |Prophylaxis - procedure intent (qualifier value)|
111 64100000 |False (qualifier value)|
108 19939008 |Subacute (qualifier value)|
97 255361000 |Slow (qualifier value)|
94 261004008 |Diagnostic intent (qualifier value)|
90 7087005 |Intermittent (qualifier value)|
89 410536001 |Contraindicated (qualifier value)|
86 260372006 |Deficient (qualifier value)|
85 75540009 |High (qualifier value)|
84 385651009 |In progress (qualifier value)|
76 71978007 |Inadequate (qualifier value)|
65 410589000 |All times past (qualifier value)|
61 385644000 |Requested (qualifier value)|

61 10828004 |Positive (qualifier value)|
56 56116003 |Patent (qualifier value)|
54 41277001 |Lacking (qualifier value)|
54 410534003 |Not indicated (qualifier value)|
54 263782009 |Inaccurate (qualifier value)|
51 255212004 |Acute-on-chronic (qualifier value)|
47 62459000 |Chronic persistent (qualifier value)|
47 26716007 |Qualitative (qualifier value)|
43 397943006 |Planned (qualifier value)|
42 385652002 |Started (qualifier value)|
40 42425007 |Equivocal (qualifier value)|
39 736678006 |Solid (state of matter)|
39 702322003 |Non-progressive (qualifier value)|
38 410545000 |Stopped before completion (qualifier value)|
36 50811001 |Routine (qualifier value)|
36 441808003 |Delayed priority (qualifier value)|
36 410546004 |Discontinued (qualifier value)|
36 410523001 |Post-starting action status (qualifier value)|
35 260385009 |Negative (qualifier value)|
33 255318003 |Relapsing course (qualifier value)|
32 410528005 |Not wanted (qualifier value)|
31 62482003 |Low (qualifier value)|
31 255228009 |Recurrent acute (qualifier value)|
30 385643006 |To be done (qualifier value)|
29 46651001 |Isolated (qualifier value)|
28 103390000 |Elective (qualifier value)|
26 274392008 |Examination under anesthesia (qualifier value)|
26 263675000 |Antenatal (qualifier value)|
24 371154000 |Dependent (qualifier value)|
22 385425000 |Improved (qualifier value)|
22 371153006 |Independent (qualifier value)|
22 260400001 |Reduced (qualifier value)|
21 371879000 |Abnormally high (qualifier value)|
20 264887000 |Not isolated (qualifier value)|
19 410535002 |Indicated (qualifier value)|
18 371880002 |Abnormally low (qualifier value)|
18 1156040003 |Self reported (qualifier value)|
17 410587003 |Past - time specified (qualifier value)|
16 18043004 |Thin (qualifier value)|
15 88694003 |Immediate (qualifier value)|
15 281304009 |Within therapeutic range (qualifier value)|
15 281303003 |Above therapeutic range (qualifier value)|
15 255507004 |Small (qualifier value)|
14 425323003 |Sudden onset AND short duration (qualifier value)|
14 371152001 |Assisted (qualifier value)|
14 281306006 |Below therapeutic range (qualifier value)|
14 278499009 |Episodic (qualifier value)|
13 44180009 |Cyclic (qualifier value)|
13 26593000 |Paroxysmal (qualifier value)|
13 21864008 |Seasonal course (qualifier value)|
13 18131002 |Acute fulminating (qualifier value)|
12 713152004 |Early childhood (qualifier value)|
12 363676003 |Palliative - procedure intent (qualifier value)|
12 180625006 |Transperitoneal approach to spine (qualifier value)|
12 134223000 |Narrow (qualifier value)|
11 788800008 |Delayed onset (qualifier value)|
11 447295008 |Forensic intent (qualifier value)|
11 410605003 |Confirmed present (qualifier value)|

10 36692007 |Known (qualifier value)|
9 260377000 |Exaggerated (qualifier value)|
8 443942000 |Requested by recipient (qualifier value)|
8 428263003 |NOT suspected (qualifier value)|
8 410537005 |Action status unknown (qualifier value)|
8 410525008 |Needed (qualifier value)|
8 276986009 |Antepartum (qualifier value)|
8 260381000 |Inefficient (qualifier value)|
8 1255665007 |Moderate (qualifier value)|
8 1156075003 |Broken (qualifier value)|
7 733985002 |Reported (qualifier value)|
7 47501007 |Chronic active (qualifier value)|
7 43261007 |Abnormal presence of (qualifier value)|
7 39187007 |Bent (qualifier value)|
7 261665006 |Unknown (qualifier value)|
7 257805000 |Medial displacement (qualifier value)|
6 74551000 |Circumference (qualifier value)|
6 47492008 |Not seen (qualifier value)|
6 410590009 |Known possible (qualifier value)|
6 371155004 |Able to and does (qualifier value)|
6 28017001 |Daytime (qualifier value)|
6 263821009 |Obstructed (qualifier value)|
6 262459003 |Low dose (qualifier value)|
6 260405006 |Trace (qualifier value)|
6 260376009 |Enlarged (qualifier value)|
6 257821005 |Manual expression (qualifier value)|
6 2546009 |Night time (qualifier value)|
6 228922002 |Gram/meal (qualifier value)|
5 708353007 |Since last encounter (qualifier value)|
5 260380004 |Inconsistent (qualifier value)|
5 255599008 |Incomplete (qualifier value)|
5 255509001 |Large (qualifier value)|
4 897015005 |Recommended (qualifier value)|
4 82334004 |Indeterminate (qualifier value)|
4 724073007 |Refused by caregiver of subject (qualifier value)|
4 419984006 |Inconclusive (qualifier value)|
4 376161000221102 |Adult population (qualifier value)|
4 261425000 |Second revision (qualifier value)|
4 260399008 |Raised (qualifier value)|
4 260370003 |Decrease (qualifier value)|
4 260350009 |Present ++++ out of ++++ (qualifier value)|
4 260349009 |Present +++ out of ++++ (qualifier value)|
4 260348001 |Present ++ out of ++++ (qualifier value)|
4 260347006 |Present + out of ++++ (qualifier value)|
3 897016006 |Not recommended (qualifier value)|
3 6493001 |Recent (qualifier value)|
3 423437008 |Insufficient (qualifier value)|
3 385653007 |Not to be stopped (qualifier value)|
3 385650005 |Organized (qualifier value)|
3 255594003 |Complete (qualifier value)|
3 255319006 |Remitting (qualifier value)|
3 14497002 |Weekly (qualifier value)|

3 103391001 |Urgency (qualifier value)|

Annex 6 - open discussion (originally as comment)

Sareh Aghaei

1:42 PM Dec 21

indicates rateQuantity?

Sareh Aghaei

1:46 PM Dec 21

amount of medication per time unit

Andrea Riedel

1:47 PM Dec 21

Rate = Amount of medication per unit of time
dose = Amount of medication per dose
rate refers to infusions or things like that, dose to tablets etc

Andrea Riedel

1:48 PM Dec 21

@steschu@gmail.com shouldn't you name it doseAndRate because it depends on the kind of medication and e.g. in our german project we use ratequantity for the same thing

Fen Natthanaphop

8:25 AM Dec 27

@steschu@gmail.com @andreariedelukerlangen@gmail.com

Should doseQuantity be FHIR: Element Id Dosage.doseAndRate.dose[x] and Rate of dose be FHIR: Element Id Dosage.doseAndRate.rate[x]?

Stefan Schulz

11:22 AM Dec 30

The domain of all relations that point to the **use** of a drug must be of the type procedure. Could you check, whether it is correct now?

Stefan Schulz

May 13, 2023

Could you give an example where this is necessary?

I could imagine also organisms in the range of value, e.g. in

71219004 |Bacterial resistance, function (observable entity)|

(although it is not 100% clear what it means and the combination of observable and function is rather obscure)

Show less

Sareh Aghaei

May 15, 2023

In the CVD data available in Estonia and the Netherlands, smoking behavior is expressed using different quantity units, such as packs and cigarettes, and time units, such as per day or year. As there is no unit for the number of cigarettes in SNOMED CT, we may consider using "cigarette" as a physical object... (?!)

Show less

Stefan Schulz

Aug 13, 2023

There should be an average number of cigarettes in a pack, so that we can deduce pack years. I am however not sure how we can resolve it at the annotation level. The best would be having something like "cigarette years" analogously to "pack years" .

Sareh Aghaei

Aug 14, 2023

We discussed it before with Kristian, the number of packs per year is something that is mostly available in their narratives. Moreover, there was an argument why annators have to do an extra calculation in annoation process

Stefan Schulz

11:49 AM Today

The conversion of units (including pseudo-units such as cigarettes or packs) should not be done by the annotators. Correctly, they are dimensionless numbers. A link to the related physical object should be in the definition of the observable concept (to do for SNOMED Intl.)

Show less

Stefan Schulz

4:15 PM Feb 27

"Definitely not present"

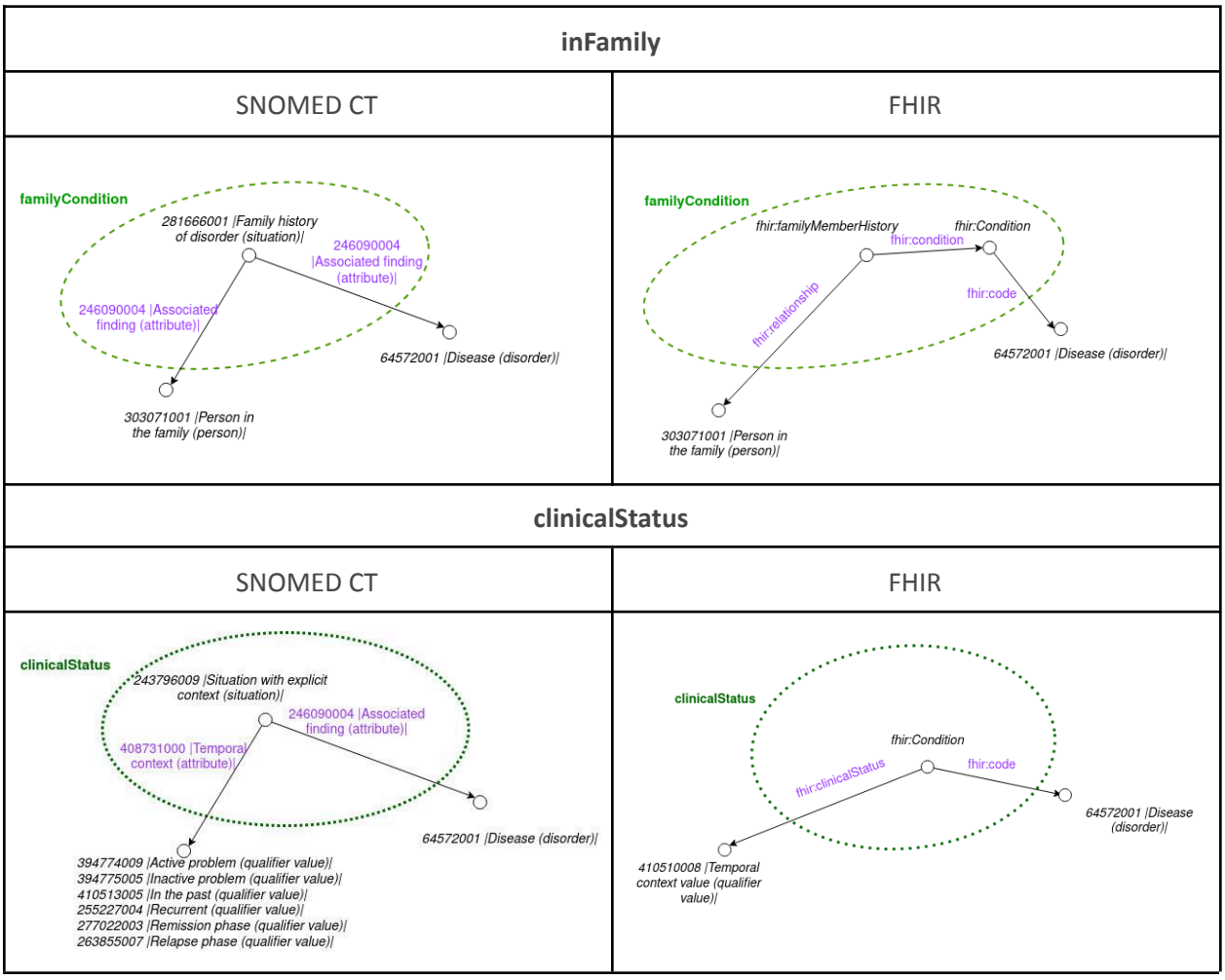
only use if emphasized negation

Stefan Schulz

5:47 PM Mar 5

don't use

(it is never used in any SNOMED axiom)



Andrea Riedel

2:42 PM Dec 18

today I saw a few examples where it is more difficult to decide, wether they are full or inferred coverage depending on the point of view: from guideline side or from snomed ct side. if the guideline says you need to use a code like product only

containing "medication xy" even if you dont necessary can read that in the text is it still full coverage? The same with "(not) detected" or "not pathologic diagnosis" because we have many different expressions that are sometimes not 100% covered. Last examples should be included in guideline for normal/intakt etc. Or "closed atlas fracture"?

Show less

Andrea Riedel

Nov 21, 2023

In Erlangen and Murcia we use 1193546000 |Map source to map target correlation (foundation metadata concept)| to describe, if the mapping is exact, partial, broader, narrower or if there is no mapping etc. It is relevant for the quality and the comparison of annotations.

Stefan Schulz

Nov 30, 2023

Andrea, if you have time, could you summarize our use of metadata here?
@andreariedelukerlangen@gmail.com

Stefan Schulz

Nov 30, 2023

@andreariedelukerlangen@gmail.co