**Mismatches between major subhierarchies and semantic tags in SNOMED CT**

Jonathan P. BONA, PhD [1]

Werner CEUSTERS, MD [2]

[1] Department of Biomedical Informatics

College of Medicine

University of Arkansas for Medical Sciences

4301 W. Markham St., #782

Little Rock, AR 72205-7199

Email: jonathanbona@gmail.com

[2] Department of Biomedical Informatics

Jacobs School of Biomedical and Medical Sciences

University at Buffalo

77 Goodell street

Buffalo, NY 14203

USA

Email: ceusters@buffalo.edu

**Conflicts of interest:**

Jonathan P. BONA: none

Werner CEUSTERS: none

**Suggested reviewers:**

| | |
|---|---|
| William Hogan, MD, MS, FACMI: | hoganwr@ufl.edu |
| Michael Chary, MD: | michael.chary@mssm.edu |
| Christian Lovis, MD, MPH, FACMI: | christian.lovis@hcuge.ch |
| Stephan Schulz, MD: | stefan.schulz@medunigraz.at |

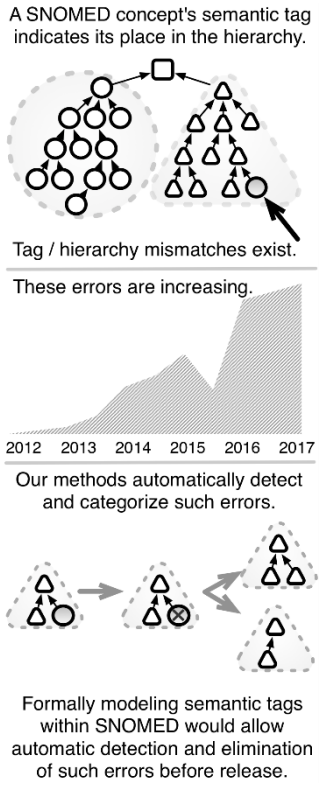
**Abstract:**

The *fully specified name* of a concept in SNOMED CT is formed by a *term* to which in the typical case is added a *semantic tag*. The latter is meant to disambiguate homonymous terms and to indicate in which major subhierarchy of SNOMED CT that concept fits. We have developed a method to determine whether a concept's tag correctly identifies its place in the hierarchy, and applied this method to an analysis of all active concepts in every SNOMED CT release from January 2003 to January 2017. Our results show (1) that there are concepts in almost every release whose semantic tag does not match their placement in the hierarchy, (2) that it is primarily disorder concepts that are involved, and (3) that the number of such mismatches increase since the July 2012 version. Our analysis determined that it is primarily the absence of a mechanism in the SNOMED CT authoring environment to suggest stated relationships for very similar concepts that is responsible for the mismatches. We argue that the SNOMED CT authoring environment should treat the semantic tags as part of the formal structure so that methods can be implemented to keep the sub-hierarchies in sync with the semantic tags.

**Keywords:**

SNOMED CT, semantic tags, quality assurance

## Graphical abstract



## 1 Introduction

SNOMED CT is a large reference terminology for the clinical domain in which what are called 'concepts', claimed to be representations of 'clinical meanings'[1], are linked to 'descriptions' which contain terms indicating various ways of how these clinical meanings are expressed in natural language.

The January 2017 version of SNOMED CT consists of 326,734 active and 214,969 inactive concepts. Concepts are linked to other concepts by means of relationships some of which are grouped so as to form machine-readable logical definitions that can be used for logical inference [2, p757]. SNOMED CT concepts are organized into a hierarchy of 'Is-a' relationships. The top concept, [138875005 | SNOMED CT Concept (SNOMED RT+CTV3)] directly subsumes 19 high level concepts. Most of these concepts are first-order concepts such as [404684003 | Clinical finding (finding)] and [123037004 | Body structure (body structure)] which serve as the root of sub-hierarchies of concepts about entities directly relevant to and within the domain of healthcare. Some of these concepts are second-order concepts that describe the structure of SNOMED CT rather than the structure of what the first-order concepts are about. SNOMED CT comes with a history mechanism that allows for a detailed analysis of how the system has changed over time [3]. The complete SNOMED CT hierarchy for each release is generated by a description logic classifier applied to "stated" definitions and relationships that are created and edited by human authors or editors of the ontology [2, p757].

Every SNOMED CT concept comes with *descriptions* one of which is selected as the *Fully Specified Name* (FSN). For example, the FSN of the concept with unique identifier '35566002' is 'Hematoma (morphologic abnormality)'. This FSN informs us that 'hematoma' – i.e. the part of the FSN that precedes the part written in parentheses – is an acceptable term by means of which concept 35566002 may be expressed in clinical language. An FSN typically ends with a short text surrounded by parentheses that is called the '*semantic tag*'. One function of this tag is to disambiguate the FSN of this concept from the FSNs of other concepts that may be expressed by the same term [2, p41]. It is thus the semantic tag 'morphologic abnormality' which disambiguates the display name of the concept [35566002 | Hematoma (morphologic abnormality)] from the concept [385494008 | Hematoma (disorder)]. This is useful when the user interface of, for example, an electronic healthcare record system returns in response to a search for 'hematoma' all the FSNs of all concepts in which this term appears in at least one of their descriptions without, however, showing the entire hierarchy: without the semantic tag, it would not be possible to determine what the difference in meaning would be between what would be displayed, for example, as [35566002 | Hematoma] and [385494008 | Hematoma].

The semantic tag (now also called the '*hierarchy tag'*) is said to '*identify the hierarchy into which the concept is placed via its Relationships*' [4, p237]. Although the SNOMED CT documentation does not

provide more detail on what this exactly means, our understanding of this is that the directed acyclic graph (DAG) formed by SNOMED CT's complete Is-a hierarchy is intended to be composed out of smaller DAGs, one for each semantic tag. Each one of these smaller DAGs, so we assume, is intended to satisfy the following criteria: (1) it is populated by all concepts whose FSNs contain the same semantic tag, and (2) there is only one concept at the root of this DAG: the 'corresponding concept'. Further, these smaller DAGs may be nested so that, for example, the DAG formed by the concepts with the semantic tag 'finding' includes the DAG formed by the concepts with the semantic tag 'disorder'.

Because semantic tags are substrings added to *names* inside FSNs and are not represented separately as part of SNOMED CT's formal model, it is not easy to determine whether there is for each semantic tag indeed a DAG that satisfies the above mentioned criteria. Moreover, there does not appear to be an official published mapping that lists the semantic tag / concept correspondences for SNOMED CT. In many cases this correspondence may seem obvious to a human observer. For many tags there is indeed a single high-level concept whose semantic tag matches exactly the part of the FSN that precedes the tag. For example, one direct sub-concept of the top SNOMED CT Concept is [71388002 | Procedure (procedure)]. This concept has the semantic tag 'procedure' and its name in the FSN is the word 'Procedure'. In other cases, the correspondence is less obvious. For instance, no direct sub-concept of SNOMED CT's top concept is tagged 'morphologic abnormality', nor is there any concept whose name is exactly 'Morphologic abnormality'. The same holds for the semantic tag 'disorder'. The concept [118956008 | Body structure, altered from its original anatomical structure (morphologic abnormality)] is a child of [123037004 | Body structure (body structure)] and appears to be the highest concept (i.e. closest to the top) tagged with 'morphologic abnormality'. If we are correct in our interpretation, then the concept [35566002 | Hematoma (morphologic abnormality)] should be classified in the sub-hierarchy of morphologic abnormalities and be subsumed by [118956008 | Body structure, altered from its original anatomical structure (morphologic abnormality)] while [385494008 | Hematoma (disorder)] should be classified in the sub-hierarchy of diseases, the highest level concept of this sub-hierarchy being [64572001 | Disease (disorder)].

The exact relationship between SNOMED CT's semantic tags and concepts has thus far not been widely researched. In [3] we explored how the semantic tags of concepts changed over time. We found in total 285 patterns according to which SNOMED CT concepts underwent changes in the semantic tags assigned to them in the collection of SNOMED CT versions studied. This included 43 patterns according to which an FSN *without* a semantic tag was changed into one *with* a semantic tag. There were no patterns with more than 3 changes over time. Changes in semantic tags were found to happen for a variety of reasons. One is a change in SNOMED CT's concept model, for example when in the newer version distinctions were made that did not exist in earlier versions, or when different interpretations were introduced (e.g. the product / substance distinction). Such changes have a global impact on large parts of the ontology. Another reason is

that concepts were in one or other way erroneous and had to be corrected. While doing these analyses, we were nevertheless hampered by the fact that the SNOMED CT documentation available from the IHTSDO provides insufficient information on what the precise set of semantic tags the SNOMED CT editors are working with might be. The information that a semantic tag is that what appears at the end of an FSN between brackets [2, p41] turned out not to be reliable. Historically, FSNs didn't have a semantic tag at all as this was apparently introduced later as witnessed by the many changes in descriptions to that end. It was also found that parsing anything that terminates an FSN between brackets leads to many false positives in older concepts, thus requiring manual inspection for disambiguation.

Furthermore, some FSNs end with more than one parenthesized substring, which makes it look at first glance as if the concepts with such FSNs might have multiple semantic tags. This in turn further confuses the question of what, exactly, counts as a semantic tag. For example, the string "contextual qualifier" appears surrounded in parentheses in 103 FSNs immediately preceding the official semantic tag "qualifier value", as in the concept: [30207005 | Risk of (contextual qualifier) (qualifier value)] and its children. A similar pattern occurs with the quasi-tag "property", as seen in [118597006 | Quantity rate (property) (qualifier value)] and 92 others. This phenomenon is not limited to qualifier values: [110818007 | Bile duct and stomach (combined site) (body structure)] is one of 296 concepts whose FSN ends with '(combined site) (body structure)'. Other examples include terms that appear to be more parenthetical clarifications rather than indicative of an implicit sub-hierarchy among tags: 'less than 2 years' in [4359001 | Early congenital syphilis (less than 2 years) (disorder)] and 'chemical processes, except Petroleum' in [9101001 | Reactor-converter operator (chemical processes, except Petroleum) (occupation)] are two examples.

Throughout this analysis we treat as semantic tags only those parenthesized substrings that occur last in an FSN. The SNOMED CT Editorial Guide supports this interpretation: '*Each FSN term **ends*** [bold emphasis added] *with a 'semantic tag' in parentheses*'[4, p208].

The work presented here assesses the January 31, 2017 International Release of SNOMED CT, including the history information that it contains starting with the January version of 2003, to determine the extent to which SNOMED CT's use of semantic tags is systematic and consistent with its placement of concepts that use those semantic tags within the concept hierarchy.

## 2 Material and methods

The research hypotheses driving this work are:

(1) Within a specific release of SNOMED CT, all semantic tags are intended to be related to the concept system through a one-to-one correspondence between the semantic tag and some unique high-level concept which we call the 'corresponding concept' for that tag.

(2) Every concept that uses a particular semantic tag *t* within a specific SNOMED CT version should be subsumed by that semantic tag's corresponding concept *Ct*, where *Ct* is the highest level concept that uses *t*, within that version. This hypothesis is motivated by the apparent change in terminology from '*semantic tag*' in [2] to '*hierarchy tag*' in [4, p227].

(3) The fact that semantic tags, so we assume, are not part of SNOMED CT's formal model may lead to mismatches: we consider a concept to be 'mismatched' in a specific SNOMED CT version if it has the semantic tag *t* but is not subsumed by that tag's corresponding concept *Ct*.

(4) Where such mismatches exist, they are due to errors in either the concept's placement in the SNOMED CT hierarchy or in its semantic tag. Such errors, when discovered by the SNOMED editors, are corrected in later releases.

To test these hypotheses, we implemented automated procedures (1) to find for each semantic tag its corresponding concept in each release, (2) to identify mismatched concepts, and (3) to group these mismatches in categories based on how mismatched concepts relate to other mismatched concepts.

Because the semantic tag 'disorder' contains the most mismatches in the latest release investigated, semi-automated and manual methods were used to identify possible causes. To that end, we retrieved and analyzed the subsumption hierarchy of all mismatched concepts for the semantic tag 'disorder' as well as their other relationships and we assessed, where possible, how they would be classified following the reference definitions of the Ontology for General Medical Science as published in [5].

### 2.1 Identifying corresponding concepts

We define *the corresponding concept* for a semantic tag *t* in a specific SNOMED CT release as*: the unique concept that uses the tag* t *and is not subsumed by any other concept that uses* t. This definition does not require tags to keep the same corresponding concept across releases.

Based on this we determine the corresponding concept *Ct* for each semantic tag *t* in a SNOMED release by means of the following algorithm:

(1) Calculate the *depth* for each concept *C* as the length of the shortest *Is-a* path from SNOMED CT's top concept, i.e. [138875005 | SNOMED CT Concept (SNOMED RT+CTV3)], to *C*.

(2) For each semantic tag *t*, select from the set of concepts tagged with *t* the concept *Xt* which is the concept with the lowest *depth*,

(3) Let *Ct* = *Xt* if none of *Xt*'s ancestors is tagged with *t*. Otherwise let *Ct* be the ancestor of *Xt* that has the lowest *depth*.

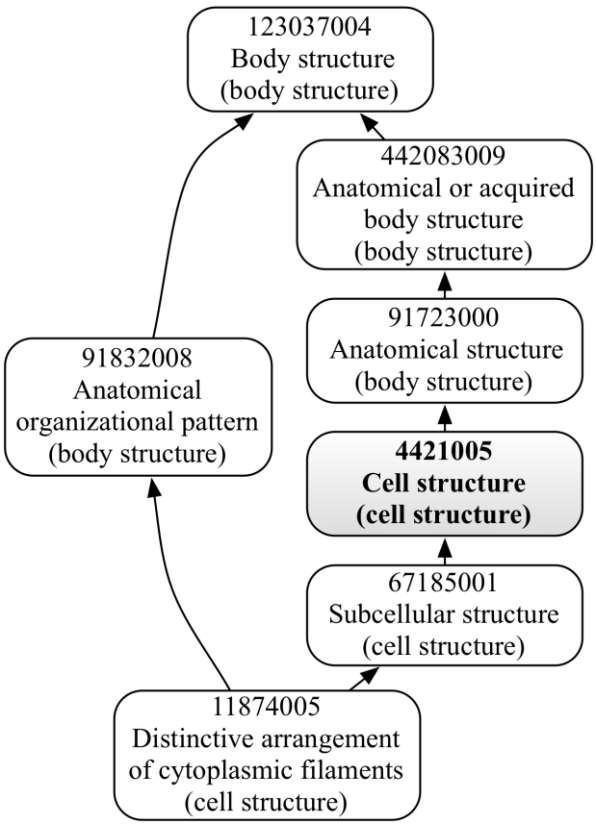



Figure 1: Effect of concept multiple inheritance on SEMANTIC TAG hierarchy.

Step 3 is necessary to handle special cases as exemplified in Figure 1. These cases arise from SNOMED CT's Is-a hierarchy forming a directed acyclic graph with a single root node that has no edges coming into it (i.e. is not subsumed by any other concept), yet allowing for multiple inheritance. Such special cases occur whenever there is a concept with some semantic tag *t* that has the shortest path to the top concept as compared to all other concepts with semantic tag *t*, and at the same time is also subsumed by another concept with semantic tag *t* that has a *longer shortest path* to the top concept. In Figure 1, this is the case for [11874005 | Distinctive arrangements of cytoplasmic filaments] which has as lowest depth '3' and is subsumed by [4421005 | Cell structure (cell structure)] which has lowest depth '4'.

The output of this process is a table with [semantic tag→concept] pairs for each release. This was inspected manually to verify whether the mappings made sense.

## 2.2 Identifying mismatched concepts

Once a corresponding concept has been identified for each semantic tag in each release, mismatched concepts in each release can be found by determining for each SNOMED CT concept whether it is subsumed by the corresponding concept for the semantic tag that it carries.

To facilitate reasoning about, storing, retrieving, and combining historic information about the SNOMED CT hierarchy and the semantic tags assigned to concepts, we developed an RDF model representing a second-order view on SNOMED CT's concept hierarchy and semantic tags, and we developed computational procedures that operate this model. We represent each SNOMED CT concept as an OWL class with separate annotations for its FSN and semantic tag. Each *Is-a* relation between two concepts has a corresponding *rdf:subClassOf* assertion. The identifiers (URIs) used for each concept have a namespace that indicates the release version. For example, *http://ex.com/r20170131#64572001* identifies the concept with concept id 64572001 in the January 31 2017 release (Figure 2).

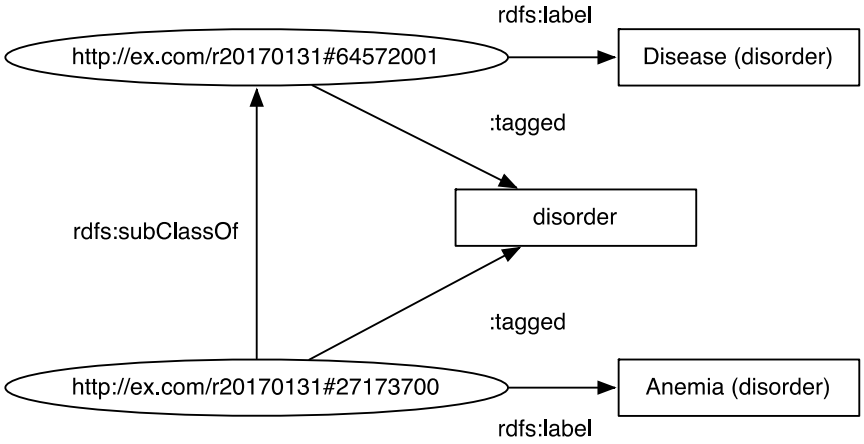



Figure 2: RDF for the concepts Disease (disorder), Anemia (disorder), their tags, and the subsumption relation between them

We produced one such RDF file for each SNOMED CT release. These files were then loaded into a single repository in a triple store database [6] configured to use the relatively lightweight RDFS+ inference rule set. Upon loading each file, the system computes and stores the transitive closure for the subClassOf relation, resulting in pre-computed subsumption information for each release (resulting in a total of roughly 185 million triples). This provides very fast retrieval of subsumption information for all releases using simple SPARQL queries, and allows us to instantly answer questions such as:

- G*iven a release **R** and a tag **t** which concepts are/are not tagged with **t** in **R**?*
- G*iven a release **R** and a concept **C**, which concepts are/are not subsumed by **C** in **R**?, and*
- G*iven a release **R**, a tag **t**, and a concept **C**, which concepts - if any - are tagged with **t** in **R**, but not subsumed by **C** in **R**?*

As an example, the following SPARQL query retrieves the concept URI, label, and semantic tag for every concept in the January 31, 2017 release that uses the tag ‘disorder’ but is not subsumed by **64572001 | Disease (disorder)**:

```
PREFIX corr:  <http://ex.com/r20170131#64572001>
PREFIX tagged:  < http://ex.com/r20170131#tagged>
PREFIX : <http://ex.com/r20170131#>
SELECT   ?conc ?l ?tag
WHERE {
    ?conc rdfs:label ?l .
    ?conc tagged: ?target_tag .
    corr: tagged: ?target_tag .
    ?conc tagged: ?tag .
    FILTER NOT EXISTS {?conc rdfs:subClassOf corr: }
}
```

We use Python scripts to automate the process of running such a query on the triple store for each tag and for each release, and to produce tables of all the concepts that have ever been mismatched in any release for further analysis.

### 2.3 Characterizing mismatched concepts

As a next step, we grouped mismatched concepts into categories based on the presence or absence of other mismatched concepts among their subsumers.

The categories into which concepts were classified were constructed by building up a three-character code where each character serves as a flag indicating whether a certain condition holds of the concept in that release. If a concept is inactive or did not yet exist at a release, then that concept was marked with an ‘empty’ code for that release. The following construction principles were used:

- **Is the concept correctly matched?** The first character is ‘Y’ if the concept is subsumed by its semantic tag’s corresponding concept in this release (i.e. if it is NOT mismatched in the release), and ‘N’ otherwise.
- **Does the concept have any non-mismatched ancestor(s)?** The second character is ‘Y’ if the concept has any ancestor concept that is NOT mismatched in that release. It is ‘N’ if every ancestor of this concept is mismatched.
- **Does the concept have any mismatched ancestor(s)?** The third character is ‘Y’ if the concept has any ancestor concept that IS mismatched in that release. It is ‘N’ if no ancestor of this concept is mismatched.

Combinatorically, this would allow us to code for nine different situations including the inactive concepts. However, given the meanings assigned to these codes, two combinations are impossible: ‘YNN’ and ‘NNN’. Ideally, every active concept in SNOMED would be in the ‘YYN’ category, indicating that the concept is properly matched to its semantic tag’s corresponding concept, as are all of the concepts that subsume it. One possible mismatched concept code is 'NYY', indicating that the mismatched concept has at least two ancestors, one mismatched and one not mismatched; the code 'NYN' indicates that the mismatched concept in question has no mismatched ancestors. Non-mismatched concepts may have either 'YYN' or 'YYY'. The latter indicates a concept that itself is not mismatched, but is subsumed by at least one mismatched concept.

### 2.4 Root-cause analysis for mismatched disorders

Our previous research has determined that, besides human error, there are at least two reasons why SNOMED CT’s structure exhibits mistakes and inconsistencies that persist over time: (1) limitations of what can be logically computed by the description logic classifier (logical issues), and (2) adherence to a concept model that obfuscates important distinctions between certain types of entities (ontological issues) [7, 8].

To assess the possible contribution of logical issues, we used the online SNOMED CT browser (http://browser.ihtsdotools.org/) to collect the stated and inferred relationships of all mismatched disorder concepts (see Table 1 for an example).

| **Relationship** | **Destination** | **Type** |
|---|---|---|
| Is a (attribute) | Elevated liver enzymes level (finding) | Stated |
| Due to (attribute) | Cystic fibrosis (disorder) | Stated |
| Is a (attribute) | Elevated liver enzymes level (finding) | Inferred |
| Due to (attribute) | Cystic fibrosis (disorder) | Inferred |
| Interprets (attribute) | Measurement of liver enzyme (procedure) | Inferred |
| Has interpretation (attribute) | Outside reference range (qualifier value) | Inferred |
| Has interpretation (attribute) | Above reference range (qualifier value) | Inferred |
| Interprets (attribute) | Measurement procedure (procedure) | Inferred |

**Table 1**: Relationships for the concept [707734002 | Elevated liver enzymes level due to cystic fibrosis (disorder)]

Each of the mismatched concepts was then annotated to indicate whether it exhibited the following characteristics:

(1) whether at least one of the concept's relationships suggests that it qualifies to be a disorder, despite not being classified as such. The following SNOMED CT relationships were counted as being suggestive for a disorder:

a) 'Associated morphology (attribute)' with as destination any '(morphologic abnormality)';
b) 'Causative agent (attribute)' with any destination;
c) 'Has interpretation (attribute)' with any destination of the sub-hierarchy '(qualifier value)' that indicates an abnormality, such as 'Abnormal (qualifier value)', 'Decreased (qualifier value)', etc.;
d) 'Pathological process (attribute)' with any destination;
e) 'Due to (attribute)' with any destination that is either in the disease sub-hierarchy, or whose FSN explicitly suggests it is a disease without being classified as one, e.g. 'Systemic disease (finding)';

(2) whether the parents are themselves misclassified concepts with the 'disorder' semantic tag;

To assess the impact of SNOMED CT's concept model on the appearance of mismatched disorders, we classified these concepts, where possible, into the categories of the Ontology for General Medical Science (OGMS) listed in Table 2. OGMS is a realist ontology that makes a clear distinction between, for instance, diseases, disease courses, and disorders. These distinctions are used here to explore whether tag mismatches may be caused by confusions between these distinct categories of entities. Note that SNOMED CT does not make these distinctions, i.e. anything classified by SNOMED CT as a Disease (disorder), as well as several other concepts classified elsewhere, would either fall under a more precisely defined category in OGMS, or contain an ambiguity from the OGMS perspective. The goal of the analysis is not merely to compare the less discriminative ontological structure of SNOMED CT for disorders with the more elaborate one of OGMS, but to detect whether this simplicity may contribute to the mismatches.

| **Term** | **Definition** |
|---|---|
| Disease | A disposition (i) to undergo pathological processes that (ii) exists in an organism because of one or more disorders in that organism. |
| Disorder | A causally relatively isolated combination of physical components that is (a) clinically abnormal and (b) maximal, in the sense that it is not a part of some larger such combination. |
| Disease course | The totality of all processes through which a given disease instance is realized. |
| Diagnosis | A conclusion of an interpretive process that has as input a clinical picture of a given patient and as output an assertion (diagnostic statement) to the effect that the patient has a disease of such and such a type. |
| Etiologic process | A process in an organism that leads to a subsequent disorder. |
| Pathological process | A bodily process that is a manifestation of a disorder. |

**Table 2:** Core categories from the Ontology for General Medical Science.

## 3 Results

### 3.1 Corresponding concept mappings

Table 3 lists the evolution of the 44 semantic tags that ever have been introduced over time. For each semantic tag (first column) the table indicates which was the corresponding concept (concept ID and FSN without the semantic tag) during the period between the releases specified in the last two columns.

| | Corresponding concept | | Release | |
|---|---|---|---|---|
| **Semantic tag** | **Concept ID** | **FSN term** | **First** | **Last** |
| administrative concept | 304813002 | Administrative values | 20030131 | 20100131 |
| assessment scale | 273249006 | Assessment scales | 20030131 | 20170131 |
| attribute | 246061005 | Attribute | 20030131 | 20170131 |
| body structure | 123037004 | Body structure | 20030131 | 20170131 |
| cell | 362837007 | Entire cell | 20030131 | 20170131 |
| cell structure | 4421005 | Cell structure | 20030131 | 20170131 |
| context-dependent category | 243796009 | Context-dependent categories | 20030131 | 20050731 |
| | | Context-dependent category | 20060131 | 20060131 |
| core metadata concept | 900000000000442005 | Core metadata concept | 20030131 | 20170131 |
| disorder | 64572001 | Disease | 20030131 | 20170131 |
| environment | 276339004 | Environments | 20030131 | 20050731 |
| | | Environment | 20060131 | 20170131 |
| environment / location | 308916002 | Environments and geographical locations | 20030131 | 20050731 |
| | | Environment or geographical location | 20060731 | 20170131 |
| ethnic group | 372148003 | Ethnic group | 20030131 | 20170131 |
| event | 272379006 | Events | 20030131 | 20050731 |
| | | Event | 20060131 | 20170131 |
| finding | 246188002 | Finding | 20030131 | 20030731 |
| | 404684003 | Clinical finding | 20040131 | 20170131 |
| foundation metadata concept | 900000000000454005 | Foundation metadata concept | 20030131 | 20170131 |
| geographic location | 223496003 | Geographical and political regions of the world | 20030131 | 20060131 |
| | | Geographical and/or political region of the world | 20060731 | 20170131 |
| inactive concept | 362955004 | Inactive concept | 20030131 | 20170131 |
| life style | 60134006 | Life style | 20030131 | 20170131 |
| link assertion | 416698001 | Link assertion | 20050731 | 20170131 |
| linkage concept | 106237007 | Linkage concept | 20050731 | 20170131 |
| metadata | 900000000000441003 | SNOMED CT Model Component | 20030131 | 20170131 |
| morphologic abnormality | 118956008 | Body structure, altered from its original anatomical structure | 20030131 | 20170131 |
| namespace concept | 370136006 | Namespace concept | 20030131 | 20170131 |
| navigational concept | 363743006 | Navigational concept | 20030131 | 20170131 |
| observable entity | 363787002 | Observable entity | 20030131 | 20170131 |
| occupation | 14679004 | Occupation | 20030131 | 20170131 |
| organism | 257495001 | Organism | 20030131 | 20040131 |
| | 410607006 | Organism | 20040731 | 20170131 |
| person | 125676002 | Person | 20030131 | 20170131 |
| physical force | 78621006 | Physical force | 20030131 | 20170131 |
| physical object | 260787004 | Physical object | 20030131 | 20170131 |
| procedure | 71388002 | Procedure | 20030131 | 20170131 |
| product | 373873005 | Pharmaceutical / biologic product | 20030131 | 20170131 |
| qualifier value | 362981000 | Qualifier value | 20030131 | 20170131 |
| racial group | 415229000 | Racial group | 20050131 | 20170131 |
| record artifact | 419891008 | Record artifact | 20060131 | 20170131 |
| regime/therapy | 243120004 | Regimes and therapies | 20030131 | 20170131 |
| religion/philosophy | 108334009 | Religion / philosophy | 20030131 | 20170131 |
| situation | 243796009 | Situation with explicit context | 20060731 | 20170131 |

| | | | | |
|---|---|---|---|---|
| social concept | 48176007 | Social context | 20030131 | 20170131 |
| special concept | 370115009 | Special concept | 20030131 | 20170131 |
| specimen | 123038009 | Specimen | 20030131 | 20170131 |
| staging scale | 254291000 | Staging and scales | 20030131 | 20170131 |
| substance | 105590001 | Substance | 20030131 | 20170131 |
| tumor staging | 254292007 | Tumor staging | 20030131 | 20170131 |

**Table 3**: Semantic tags and their corresponding concepts in releases January 2003 to January 2017

This mapping is relatively stable across releases, though there are some changes. In the majority of cases, the semantic tag turned out to be identical to the name of the corresponding concept ignoring capitalization and spacing. Differences occur primarily in corresponding concepts of which the term preceding the semantic tag is lengthy, as in 'Pharmaceutical / biologic product' and 'Body structure, altered from its original anatomical structure'. Three tags are absent initially but appear later: 'link assertion', 'linkage concept' and 'situation'. Two were present initially but removed later: 'administrative concept' and 'context-dependent category'. Some corresponding concepts had minor edits made to their FSNs, in all but one a mere lexical change from the plural form to the singular form as from 'events' to 'event'. Finally, only two tags switched their corresponding concepts from one release to the next: the 'finding' tag initially had as its corresponding concept [246188002 | Finding (finding)] but this concept was deactivated in the January 2004 release and the 'finding' corresponding concept changed to [404684003 | Clinical finding (finding)]. The introduction of a new concept is motivated here by the editorial principle that important name changes require de-activation of the concept and introduction of a new one. The second change occurred for the tag 'organism' which saw its corresponding concept changed in the July 2004 release from [257495001 | Organism (organism)] to [410607006 | Organism (organism)]. The latter concept was introduced in that release. The former was deactivated as of that release. A July 2004 Historical Association Reference Set entry (these 'provide links between inactive concepts and their active replacements or equivalents' [2, p508]) asserts that [257495001 | Organism (organism)] POSSIBLY EQUIVALENT TO [410607006 | Organism (organism)].

### 3.2 Mismatched concepts

After identifying all mismatched concepts for every semantic tag in every release, we organized counts of mismatched concepts into a table with one row per semantic tag and one column per release. A total of 466 concepts were found to have been mismatched at least once, a small fraction of the total number of SNOMED CT concepts. These mismatches occurred for only 6 semantic tags out of the 44 tags ever used. Since concepts were found to have been assigned different semantic tags over time – in some cases constituting a violation of the editorial principles when the change is over different subhierarchies [4, p226]

– 3 more semantic tags were involved, however in such a way that concepts were never mismatched while being assigned this tag.

**Table 4** provides a condensed view of the results by eliminating the semantic tags with no mismatches at all. It turns out that most mistakes occurred for the semantic tags ‘disorder’ and ‘regime/therapy’. Whereas problems with the latter disappeared over time, the ‘disorder’ tag seems to become more and more the seat of errors in recent versions. In the January 2017 release there are only four tags with mismatched concepts for a total of 89 mismatches out of the 466 concepts ever mismatched: ‘disorder’ (83), ‘regime/therapy’ (4), ‘product’ (1), and ‘substance’ (1).

| Release | disorder | finding | observable entity | product | regime/ therapy | substance | Total |
|---|---|---|---|---|---|---|---|
| **20030131** | | 4 | | | | 1 | **5** |
| **20030731** | | 1 | | | | | **1** |
| **20040131** | | | | | | | |
| **20040731** | | 5 | 2 | | | | **7** |
| **20050131** | | | | | | | |
| **20050731** | | | | | | | |
| **20060131** | | | | | | | |
| **20060731** | | | | | | | |
| **20070131** | | | | | | | |
| **20070731** | 44 | | | | 37 | | **81** |
| **20080131** | 8 | | | | 259 | | **267** |
| **20080731** | 19 | | | | 123 | | **142** |
| **20090131** | | | | 1 | | | **1** |
| **20090731** | | | | 1 | | | **1** |
| **20100131** | | | | 1 | | | **1** |
| **20100731** | | | | 1 | | | **1** |
| **20110131** | | | | 1 | | | **1** |
| **20110731** | | | | 1 | | | **1** |
| **20120131** | | | | 1 | | | **1** |
| **20120731** | 2 | | | 1 | | | **3** |
| **20130131** | 4 | | | 1 | | | **5** |
| **20130731** | 10 | | | 1 | | | **11** |
| **20140131** | 26 | | | 1 | | | **27** |
| **20140731** | 32 | | | 1 | | | **33** |
| **20150131** | 44 | | | 1 | | | **45** |
| **20150731** | 24 | | | 1 | 3 | | **28** |
| **20160131** | 74 | | | 1 | 3 | | **78** |
| **20160731** | 78 | | | 1 | 3 | | **82** |
| **20170131** | 83 | | | 1 | 4 | 1 | **89** |

**Table 4**: Changes in mismatches per semantic tag over time.

Table 5 provides more detail about the categorization of mismatched concepts by release. Table 6 displays irrespective of version how semantic tag/mismatch categories evolved into each other, including the

activation and de-activation of concepts. Both tables were constructed by retrieving all concepts that are, or have ever been mismatched in any release. Table 7 provides an example restricted to those concepts that were ever mismatched and in at least one version tagged as 'context-dependent entity' to show how the sort of transitions displayed in Table 6 evolve over time. For all 466 ever mismatched concepts, 59 different patterns were found, ranging from minimum 2 to maximum 5 transitions. These 59 patterns are composed out of 61 different sort of transitions between any two phases. Together, these three tables provide indications of where and how things went awry.

Clearly noticeable in Table 5 is the disappearance of the semantic tag 'context-dependent entity' from 13 concepts in July 2006 and the appearance of the tag 'situation' in an equal number. Deeper investigation revealed not only that the very same 13 concepts were involved in the switch, but also that the majority of the 24 concepts ever mismatched that started off as context-dependent entities and the 3 concepts that were activated in the third release as context-dependent entities exhibited a rough trajectory of multiple semantic tag changes involving 'situation', 'observable entity', 'disorder' and 'finding', examples of such tag transitions being context-dependent entity → finding → context-dependent entity → situation → finding, and context-dependent entity → situation → disorder → finding, whereby finally 10 of these 27 concepts were found to be mismatched in the last version (Table 7).

A comparable shift occurred from 'regime/therapy' to 'procedure'. In this case, multiple tag changes were not observed.

A third observation is that during the period from July 2007 to July 2008, mismatches are largely of the NYY-type. In more recent versions however, it is the NYN-type that dominates.

| | context-dependent category | situation | disorder | | | finding | | | observable entity | | procedure | regime/therapy | | | product | substance | | Totals | | | | | %Error |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | YYN | YYN | NYN | NYY | YYN | NYN | NYY | YYN | NYN | NYY | YYN | NYN | NYY | YYN | NYN | NYN | YYN | NYN | NYY | YYN | Active | Observed | |
| **20030131** | 24 | | | | 56 | 3 | 1 | 38 | | | 12 | | | 226 | | 1 | 1 | 4 | 1 | 357 | 362 | 362 | **1.38** |
| **20030731** | 23 | | | | 56 | 1 | | 38 | | | 12 | | | 232 | | | 2 | 1 | | 363 | 364 | 368 | **0.27** |
| **20040131** | 27 | | | | 57 | | | 37 | | | 6 | | | 238 | | | 2 | | | 367 | 367 | 371 | |
| **20040731** | 20 | | | | 57 | 5 | | 37 | 1 | 1 | 1 | | | 244 | | | 2 | 6 | 1 | 361 | 368 | 372 | **1.9** |
| **20050131** | 27 | | | | 57 | | | 37 | | | 1 | | | 246 | | | 2 | | | 370 | 370 | 374 | |
| **20050731** | 27 | | | | 57 | | | 37 | | | 1 | | | 246 | | | 2 | | | 370 | 370 | 374 | |
| **20060131** | 13 | | | | 71 | | | 37 | | | 1 | | | 246 | | | 2 | | | 370 | 370 | 374 | |
| **20060731** | | 13 | | | 74 | | | 34 | | | 1 | | | 246 | | | 2 | | | 370 | 370 | 374 | |
| **20070131** | | 13 | | | 76 | | | 34 | | | | | | 247 | | | 2 | | | 372 | 372 | 376 | |
| **20070731** | | 8 | 18 | 26 | 44 | | | 34 | | | | 18 | 19 | 218 | | | 2 | 36 | 45 | 306 | 387 | 391 | **20.93** |
| **20080131** | | 8 | 6 | 2 | 75 | | | 44 | | | | 39 | 220 | | | | 2 | 45 | 222 | 129 | 396 | 400 | **67.42** |
| **20080731** | | 8 | 1 | 18 | 62 | | | 48 | | | 130 | 6 | 117 | | | | 2 | 7 | 135 | 250 | 392 | 401 | **36.22** |
| **20090131** | | 8 | | | 81 | | | 48 | | | 131 | | | 122 | 1 | | 2 | 1 | | 392 | 393 | 402 | **0.25** |
| **20090731** | | 5 | | | 81 | | | 52 | | | 130 | | | 122 | 1 | | 2 | 1 | | 392 | 393 | 402 | **0.25** |
| **20100131** | | 4 | | | 68 | | | 50 | | | 127 | | | 122 | 1 | | 2 | 1 | | 373 | 374 | 402 | **0.27** |
| **20100731** | | 4 | | | 69 | | | 50 | | | 127 | | | 122 | 1 | | 2 | 1 | | 374 | 375 | 403 | **0.27** |
| **20110131** | | 4 | | | 69 | | | 50 | | | 127 | | | 122 | 1 | | 2 | 1 | | 374 | 375 | 403 | **0.27** |
| **20110731** | | 4 | | | 69 | | | 50 | | | 127 | | | 122 | 1 | | 2 | 1 | | 374 | 375 | 403 | **0.27** |
| **20120131** | | 4 | | | 69 | | | 50 | | | 127 | | | 122 | 1 | | 2 | 1 | | 374 | 375 | 403 | **0.27** |
| **20120731** | | 4 | 2 | | 68 | | | 50 | | | 127 | | | 122 | 1 | | 2 | 3 | | 373 | 376 | 404 | **0.8** |
| **20130131** | | 4 | 4 | | 68 | | | 50 | | | 127 | | | 122 | 1 | | 2 | 5 | | 373 | 378 | 406 | **1.32** |
| **20130731** | | 4 | 10 | | 68 | | | 50 | | | 127 | | | 122 | 1 | | 2 | 11 | | 373 | 384 | 412 | **2.86** |
| **20140131** | | 4 | 17 | 9 | 62 | | | 50 | | | 127 | | | 122 | 1 | | 2 | 18 | 9 | 367 | 394 | 422 | **6.85** |
| **20140731** | | 4 | 22 | 10 | 61 | | | 50 | | | 127 | | | 122 | 1 | | 2 | 23 | 10 | 366 | 399 | 427 | **8.27** |
| **20150131** | | 4 | 34 | 10 | 62 | | | 51 | | | 127 | | | 122 | 1 | | 2 | 35 | 10 | 368 | 413 | 441 | **10.9** |
| **20150731** | | 4 | 24 | | 86 | | | 55 | | | 127 | 3 | | 122 | 1 | | 2 | 28 | | 396 | 424 | 453 | **6.6** |
| **20160131** | | 4 | 65 | 9 | 76 | | | 21 | | | 127 | 3 | | 122 | 1 | | 2 | 69 | 9 | 352 | 430 | 459 | **18.14** |
| **20160731** | | 4 | 67 | 11 | 73 | | | 21 | | | 127 | 3 | | 122 | 1 | | 2 | 71 | 11 | 349 | 431 | 460 | **19.03** |
| **20170131** | | 4 | 69 | 14 | 73 | | | 21 | | | 127 | 4 | | 122 | 1 | 1 | 1 | 75 | 14 | 348 | 437 | 466 | **20.37** |

**Table 5:** Categorization of mismatches over time. Columns for categories not occurring within the realm of a semantic tag are not shown. The 'Observed' column tallies all ever mismatched concepts that since the first version were observed, whether or not de-activated in the meantime. The '%Error' column displays the % of mismatches over all active concepts. Areas of particular interest are shaded. 'NYY': the concept is mismatched and has at least one mismatched ancestor; 'NYY': the concept is mismatched and has no mismatched ancestors; 'YYY': the concept is not mismatched but has at least one mismatched ancestor; 'YYN': the concept is not mismatched and has no mismatched ancestors.

| | | TO | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | context-dependent category | disorder | | | finding | | observable entity | | procedure | product | regime/therapy | | | situation | substance | | Terminal | Deactivated | Totals |
| **FROM** | | **YYN** | **YYN** | **NYY** | **NYN** | **YYN** | **NYN** | **NYY** | **NYN** | **YYN** | **NYN** | **YYN** | **NYY** | **NYN** | **YYN** | **YYN** | **NYN** | | | |
| **context-dependent category** | **YYN** | | 14 | | | | 6 | 1 | 1 | | | | | | 13 | | | | | 35 |
| **disorder** | **YYN** | | | 51 | 27 | | | | | | | | | | | | | 73 | 13 | 164 |
| | **NYY** | | 48 | | | 8 | | | | | | | | | | | | 14 | | 70 |
| | **NYN** | | 38 | | | 10 | | | | | | | | | | | | 69 | 1 | 118 |
| **finding** | **YYN** | | 4 | 2 | 32 | | | | | | | | | | | | | 17 | 2 | 57 |
| | **NYY** | | | | | | | | | | | | | | | | | | 1 | 1 |
| | **NYN** | 6 | | | | | | | | | | | | | | | | | 3 | 9 |
| **observable entity** | **NYY** | 1 | | | | | | | | | | | | | | | | | | 1 |
| | **NYN** | 1 | | | | | | | | | | | | | | | | | | 1 |
| **procedure** | **YYN** | | | | | | | | | | | 13 | | | 1 | | | 126 | 3 | 143 |
| **product** | **NYN** | | | | | | | | | | | | | | | | | 1 | | 1 |
| **regime/therapy** | **YYN** | | | | | | | | | | | | 217 | 30 | | | | 121 | | 368 |
| | **NYY** | | | | | | | | | 96 | | 117 | | 3 | | | | | 5 | 221 |
| | **NYN** | | | | | 1 | | | | 35 | | 5 | 1 | | | | | 4 | | 46 |
| **Situation** | **YYN** | | | 5 | | 4 | | | | | | | | | | | | 4 | 1 | 14 |
| **substance** | **YYN** | | | | | | | | | | | | | | | | 1 | 1 | | 2 |
| | **NYN** | | | | | | | | | | | | | | | 1 | | 1 | | 2 |
| **Inactive** | | 3 | 4 | 12 | 59 | | | | | 1 | 1 | 8 | 3 | 13 | | | | | | 104 |
| **Deactivated** | | | | | | | | | | | | | | | | | | 25 | | 25 |
| **Total** | | 11 | 108 | 70 | 118 | 23 | 6 | 1 | 1 | 132 | 1 | 143 | 221 | 46 | 14 | 1 | 1 | 456 | 29 | 1382 |

**Table 6:** Transitions in semantic tags and mismatch categories. The 'Terminal' column contains counts for the number of concepts that maintained their last state into which they were changed. 'Maintained' is hereby interpreted literally, and excludes 10 concepts that were only active in 1 version. 'NYY': the concept is mismatched and has at least one mismatched ancestor; 'NYY': the concept is mismatched and has no mismatched ancestors; 'YYY': the concept is not mismatched but has at least one mismatched ancestor; 'YYN': the concept is not mismatched and has no mismatched ancestors.

| ID | Fully Specified Name in last version where active | Phase 1 | | Phase 2 | | | Phase 3 | | | Phase 4 | | | Phase 5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ST | MMC | Rel. | ST | MMC | Rel. | ST | MMC | Rel. | ST | MMC | Rel. | ST | MMC |
| 162275003 | No visual symptom (situation) | cdc | YYN | 2 | finding | NYN | 3 | cdc | YYN | 8 | situation | YYN | | | |
| 408311002 | OE - retinopathy (disorder) | InAct | InAct | 3 | cdc | YYN | 7 | disorder | YYN | 28 | disorder | NYN | | | |
| 408312009 | OE - referable retinopathy (disorder) | InAct | InAct | 3 | cdc | YYN | 7 | disorder | YYN | 28 | disorder | NYY | | | |
| 408313004 | OE - non-referable retinopathy (disorder) | InAct | InAct | 3 | cdc | YYN | 7 | disorder | YYN | 28 | disorder | NYY | | | |
| 162649008 | Depth of examination (situation) | cdc | YYN | 4 | observable | NYN | 5 | cdc | YYN | 8 | situation | YYN | | | |
| 272052001 | Patient not understood (situation) | cdc | YYN | 4 | finding | NYN | 5 | cdc | YYN | 8 | situation | YYN | | | |
| 272053006 | Poor witness (finding) | cdc | YYN | 4 | finding | NYN | 5 | cdc | YYN | 8 | situation | YYN | 14 | finding | YYN |
| 272054000 | Poor historian (finding) | cdc | YYN | 4 | finding | NYN | 5 | cdc | YYN | 8 | situation | YYN | 14 | finding | YYN |
| 272055004 | Misleading historian (finding) | cdc | YYN | 4 | finding | NYN | 5 | cdc | YYN | 8 | situation | YYN | 14 | finding | YYN |
| 272056003 | Unreliable witness (finding) | cdc | YYN | 4 | finding | NYN | 5 | cdc | YYN | 8 | situation | YYN | 14 | finding | YYN |
| 162672005 | Depth of examination NOS (situation) | cdc | YYN | 4 | observable | NYY | 5 | cdc | YYN | 8 | situation | YYN | 15 | DeAct | DeAct |
| 312450001 | OE - not dehydrated (finding) | cdc | YYN | 7 | disorder | YYN | 11 | disorder | NYN | 12 | finding | YYN | | | |
| 164200008 | OE - Little's area hyperemic (disorder) | cdc | YYN | 7 | disorder | YYN | 20 | disorder | NYN | 26 | disorder | YYN | | | |
| 274309006 | OE - petechiae on skin (disorder) | cdc | YYN | 7 | disorder | YYN | 23 | disorder | NYY | 26 | disorder | YYN | | | |
| 164582004 | OE - lower leg bone abnormal (disorder) | cdc | YYN | 7 | disorder | YYN | 26 | disorder | NYN | | | | | | |
| 164506003 | OE - joint abnormal (disorder) | cdc | YYN | 7 | disorder | YYN | 27 | disorder | NYN | | | | | | |
| 164508002 | OE - multiple joint abnormal (disorder) | cdc | YYN | 7 | disorder | YYN | 27 | disorder | NYY | | | | | | |
| 164510000 | OE - elbow joint abnormal (disorder) | cdc | YYN | 7 | disorder | YYN | 27 | disorder | NYY | | | | | | |
| 164511001 | OE - wrist joint abnormal (disorder) | cdc | YYN | 7 | disorder | YYN | 27 | disorder | NYY | | | | | | |
| 164513003 | OE - hand joint abnormal (disorder) | cdc | YYN | 7 | disorder | YYN | 27 | disorder | NYY | | | | | | |
| 164520005 | OE - toe joint abnormal (disorder) | cdc | YYN | 7 | disorder | YYN | 27 | disorder | NYY | | | | | | |
| 164521009 | OE - neck joint abnormal (disorder) | cdc | YYN | 7 | disorder | YYN | 27 | disorder | NYY | | | | | | |
| 169580009 | Antenatal care: recurrent aborter (finding) | cdc | YYN | 8 | situation | YYN | 10 | disorder | NYY | 11 | finding | YYN | | | |
| 169581008 | Antenatal care: grand multiparity (finding) | cdc | YYN | 8 | situation | YYN | 10 | disorder | NYY | 11 | finding | YYN | | | |
| 169592009 | Antenatal care: poor home conditions (finding) | cdc | YYN | 8 | situation | YYN | 10 | disorder | NYY | 11 | finding | YYN | | | |
| 169593004 | Antenatal care: poor antenatal attender (finding) | cdc | YYN | 8 | situation | YYN | 10 | disorder | NYY | 11 | finding | YYN | | | |
| 169594005 | Late onset antenatal care (finding) | cdc | YYN | 8 | situation | YYN | 10 | disorder | NYY | 11 | finding | YYN | | | |

Table 7. Evolution of the concepts ever mismatched that in at least one version had the semantic tag 'context-dependent entity' (cdc). Each phase corresponds to a specific semantic tag (ST) /mismatch configuration (MMC) or inactive/de-activated status. 'Rel.' = release, from 1 (January 2003) to 29 (January 2017). Concepts are grouped according to similarity in their phase transitions. The frequently occurring 'On examination' phrase in FSNs is abbreviated to 'OE', the ST 'observable entity is abbreviated to 'observable'. 'NYY': the concept is mismatched and has at least one mismatched ancestor; 'NYY': the concept is mismatched and has no mismatched ancestors; 'YYY': the concept is not mismatched but has at least one mismatched ancestor; 'YYN': the concept is not mismatched and has no mismatched ancestors.

### 3.3 Mismatched disorders

Table 8 shows the distribution of observed problems over the tentative OGMS category (see their definitions in Table 2) the concept would be classified under. The category 'Diagnosis or Disorder' is not a genuine OGMS-category, but is used to categorize the mismatched SNOMED CT concepts of which the fully specified name is ambiguous as to whether a disorder (something 'on the side of the patient' independent of whether a diagnosis about it has been made) is intended or a diagnosis (something 'on the side of the clinician'). An example is the SNOMED CT concept '*On examination - wrist joint abnormal (disorder)*': the semantic tag suggests it to be treated as an OGMS:Disorder while the leading phrase 'On examination' is suggestive for the concept to be referring to a diagnosis. SNOMED CT obfuscates this important distinction by categorizing all disorders as findings, where, obviously, disorders exist whether or not they have been found, and diagnoses stating there to be a disorder of type X can be wrong either because of the absence of a disorder on the side of the patient at all, or because the existing disorder is not of the type suggested by the diagnosis. The category '?' is used where the intended meaning of the concept's fully specified name is obscure to us; it only occurred for 'Sporadic disorder (disorder)': is this concept to be used for disorders that occur rarely (whether or not these disorders are rarely diagnosed) or for diagnoses that are made rarely (whether or not they are about disorders that occur rarely)?

When at least one of the concept's inferred or stated relationships suggests that it qualifies to be a disorder (see section 2.4 for the relationship/target-concept combinations we considered suggestive in this way and Table 9 for the various ways in which disorder indications are realized in SNOMED CT), we marked the concept as carrying a 'disorder indication' (DI+). The label 'DI-' in Table 8 stands for absence of a disorder indication in the concept's relationships. Whereas Table 8 demonstrates that a disorder indication in the relationships does not appear to be discriminative for mismatched concepts with mismatched ancestors, the group of mismatched concepts without mismatched ancestors has twice as many concepts without a disorder indication. Absence of disorder indication is overwhelmingly present in relation to concepts that in OGMS would be classified as disease courses, diseases or pathological processes whereas concepts that would be classified as OGMS:disorders are almost equally distributed.

| | | parent is mismatched disorder | | parent is not a mismatched disorder | | TOTALS |
|---|---|---|---|---|---|---|
| | | **DI+** | **DI-** | **DI+** | **DI-** | |
| **OGMS category** | **Diagnosis or Disorder** | 6 | 2 | 1 | 2 | **11** |
| | **Disease course** | 1 | | 3 | 15 | **19** |
| | **Disease** | | 5 | 1 | 7 | **13** |
| | **Disorder** | | | 16 | 14 | **30** |
| | **Pathological process** | | | 1 | 7 | **8** |
| | **Etiologic process** | | | 1 | | **1** |
| | **?** | | | | 1 | **1** |
| **TOTALS** | | **7** | **7** | **23** | **46** | **83** |

Table 8. Distribution of observed problems over the tentative OGMS category the concept would be classified under.

Table 9 lists the 9 disorder indication patterns that we observed in the mismatched disorder concepts. A pattern is a specific combination of one or more relationship/target-concept pairs indicating that the source concept should be a disorder, for instance 'has-interpretation | Abnormal (qualifier value)'. In Table 9, each row represents a particular combination of disorder indications, the last column indicating how many source concepts were found to have that combination.

| Disorder Indication Pattern | Relationships | | | | | #mismatched |
|---|---|---|---|---|---|---|
| | Associated morphology | causative agent | has interpretation | Pathological process | due to | |
| P1 | | | | | 1 | 2 |
| P2 | | | | 1 | | 3 |
| P3 | | | 1 | | | 9 |
| P4 | | | 1 | | 1 | 2 |
| P5 | | 1 | | | | 4 |
| P6 | | 1 | 1 | | | 2 |
| P7 | 1 | | | | | 6 |
| P8 | 1 | | | | 1 | 1 |
| P9 | 1 | | 1 | | 1 | 1 |

Table 9. Distribution of the 30 mismatched disorder concepts despite disorder indication in the relationships.

## 4 Discussion

### 4.1 Semantic tag / concept correspondence

Our hypothesis that SNOMED CT intends its semantic tags to have a one-to-one correspondence between tags and certain high-level concepts is supported by (1) the very existence of identifiable tag corresponding concepts (a single 'highest' concept for each tag that is close to the top concept and that in each case subsumes the vast majority of concepts that use the tag), (2) the extremely low occurrence of mismatched concepts as compared to the total number of active concepts, and (3) the low number of semantic tags for which mismatches are found.

Errors in semantic tag assignment are non-trivial especially because semantic tags are intended to convey the meaning of concepts to users, specifically under circumstances when the entire hierarchy cannot be visualized. As SNOMED CT sees increasing use in electronic healthcare record (EHR) systems, it is vitally important to eliminate errors which may confuse users into entering wrong information, or into misinterpreting existing information. This may impact not just the usability of EHR data for research, but for patient care as well.

### 4.2 Introduction of mismatches over time

Though tag / hierarchy mismatches are rare, they are persistently present, even in recent releases. In addition, their incidence among the 'disorder'-tagged concepts has been increasing in recent releases, as new errors of this sort are still being introduced.

One example of this in the January 2017 release is the concept [109186003 | Sickle cell test kit (substance)] which is newly mismatched in this release. There were no 'substance' mismatched concepts from 2009 until 2017. In 2017 the Sickle cell test kit concept is mismatched because it is not subsumed by the 'substance' tag's corresponding concept [105590001 | Substance (substance)]. It is directly subsumed by [385387009 | Test kit (physical object)], which has 29 other children that all have the words 'test kit' in their FSN and, in 2017, are correctly tagged with 'physical object'. One example is the concept [1109190001 | Virus test kit (physical object)]. This situation is shown on the left hand side of Figure 3.

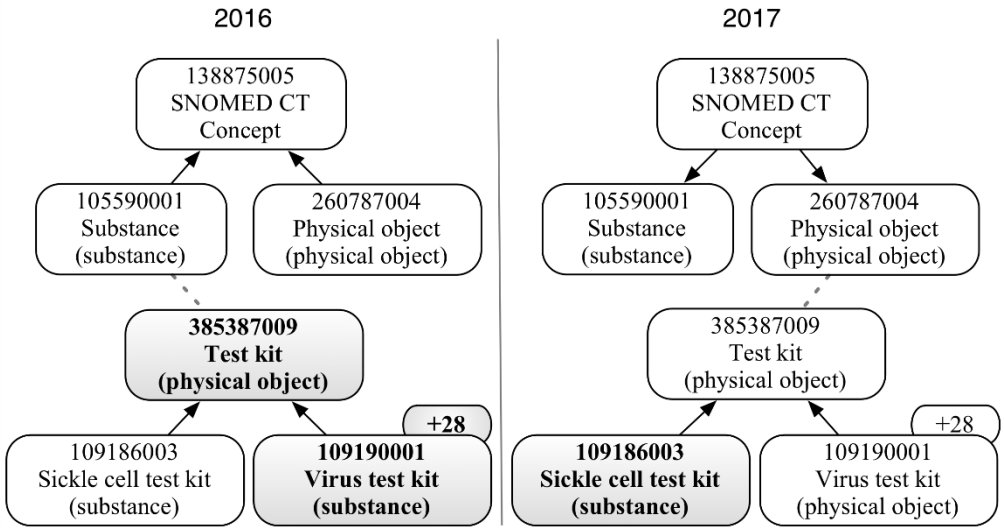



**Fig.3.** Test kit concept changes 2016 – 2017.

There are a number of ways for a mismatched concept to appear in a release, including: 1) the addition of a new concept, 2) re-activation of an old concept, and 3) changes in the concept's subsumption hierarchy, such as the removal of one or more Is-a relations. The addition of an Is-a relation alone cannot cause a concept to become mismatched. The removal of an Is-a relation that severs the path between a concept and its tag corresponding concept is sometimes counteracted by the addition of a new Is-a relation that restores the path.

In the sickle cell test kit case, changes in the hierarchy are responsible for the mismatch. This case also shows an especially interesting transition, as the Sickle cell test kit concept mismatch observable in 2017 is introduced by changes that eliminate other mismatches. In 2016 and earlier, the Sickle cell test kit concept's parent concept [385387009 | Test kit (physical object)] was itself mismatched, being subsumed by [105590001 | Substance (substance)] but, incorrectly, not subsumed by the 'physical object' tag's corresponding concept, [260787004 | Physical object (physical object)]. The test kit concept's children were all as they are now: the sickle cell test kit concept was tagged 'substance' and the rest were tagged 'physical object' (and hence, also mismatched).

In 2017 the test kit concept was (correctly) moved to the physical object hierarchy, and it and 29 of its children went from being mismatched to not mismatched. The move resulted in a net reduction in mismatches but the sickle cell test kit concept became mismatched as a result, as highlighted on the right hand side of Figure 3. Note that most of the child concepts of [385387009 | Test kit (physical object)] are omitted here in the interest of space, as are child concepts of other concepts in the figure.

### 4.3 Limitations of the SNOMED CT authoring environment

Our analysis suggests the existence of two shortcomings in SNOMED CT's authoring environment.

First, SNOMED CT's authoring environment seems to lack a mechanism to verify that each concept's semantic tag matches its placement in the subsumption hierarchy.

Both the test kit subhierarchy example discussed above, and another example involving the abnormal joint subhierarchy (shown in **Figure 4** and discussed in more detail below) illustrate a common occurrence: the introduction or removal of batches of concept mismatches caused by Is-a changes that occur higher up in the hierarchy than the immediate Is-a relations between affected concepts and their parent concepts.

This hints at the likelihood that, despite the documented insistence that semantic tags reflect the hierarchy, the SNOMED CT authoring environment is not equipped to detect and warn editors about the possible effects that stated or inferred hierarchy changes may have on tag assignment correctness; and that it is non-trivial for SNOMED CT editors to predict these effects.

This appears to indicate that, even in the back-end authoring environment, semantic tags are represented only as substrings of FSNs and not in a more formal or structured way that can easily be aligned with classification and other automated processes that are used. We believe that a good first step to address this would be to augment the authoring environment to use a better and formal representation of semantic tags, explicitly representing the tags as separate artifacts on their own, as well as the relation between a concept and its tag. This is the approach followed in the RDF model we are using, as described in Section 2.2. This would facilitate the implementation of simple procedures to verify the tag / hierarchy correspondence in each release. To do this would also involve the creation and use of an official mapping between tags and their corresponding concepts. This mapping should be published as part of the SNOMED CT documentation to clarify for users of SNOMED the exact role of semantic tags as 'hierarchy tags' indicating a concept's place in the hierarchy.

Second, an examination of these mismatched concepts for systematic errors that may reveal some underlying confusion leads us to conclude that there seems to be no mechanism in the SNOMED CT authoring environment to suggest stated relationships for very similar concepts. Table 10 shows stated and inferred relationships for 12 concepts, both mismatched and otherwise, whose FSN contains the pattern 'On examination – X joint abnormal (disorder)'. '

| | | | On examination – X joint abnormal (disorder) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | mismatched | | | | | | | not mismatched | | | | |
| **Attribute Type** | **Destination** | **CharType** | **-** | **multiple** | **elbow** | **wrist** | **hand** | **toe** | **neck** | **hip** | **knee** | **foot** | **ankle** | **shoulder** |
| Finding informer | Performer of method (person) | Stated | x | | | | | | | | | | | |
| Finding informer | Performer of method (person) | Inferred | x | x | x | x | x | x | x | x | x | x | x | x |
| Finding method | Physical examination procedure (procedure) | Stated | x | | | | | | | | | | | |
| Finding method | Physical examination procedure (procedure) | Inferred | x | x | x | x | x | x | x | x | x | x | x | x |
| Finding site | Joint structure (body structure) | Stated | x | | | | | | | | | | | |
| Finding site | Joint structure (body structure) | Inferred | x | x | x | x | x | x | x | | x | | | x |
| Has interpretation | Abnormal (qualifier value) | Stated | x | | | | | | | | | | | |
| Has interpretation | Abnormal (qualifier value) | Inferred | x | x | x | x | x | x | x | | | | | |
| Interprets | Examination of joint (procedure) | Stated | x | | | | | | | | | | | |
| Interprets | Examination of joint (procedure) | Inferred | x | x | x | x | x | x | x | | | | | |
| Is a | Foot joint finding (finding) | Stated | | | | | | | | | | x | | |
| Is a | Foot joint finding (finding) | Inferred | | | | | | | | | | x | | |
| Is a | Arthropathy (disorder) | Stated | | | | | | | | x | x | x | x | x |
| Is a | Arthropathy (disorder) | Inferred | | | | | | | | | | | | x |
| Is a | On examination - joint abnormal (disorder) | Stated | | x | x | x | x | x | x | | | | | |
| Is a | On examination - joint abnormal (disorder) | Inferred | | x | x | x | x | x | x | | | | | |
| Is a | Clinical finding (finding) | Stated | x | | | | | | | | | | | |
| Is a | Joint function disorder (finding) | Stated | | x | | | | | | | | | | |
| Is a | Ankle joint finding (finding) | Stated | | | | | | | | | | | x | |
| Is a | Finding of hand region (finding) | Stated | | | | | x | | | | | | | |
| Is a | Hip joint finding (finding) | Stated | | | | | | | | x | | | | |
| Is a | Knee joint finding (finding) | Stated | | | | | | | | | x | | | |
| Is a | On examination - specified examination finding (finding) | Stated | | | | | | | | x | x | x | x | x |
| Finding site | Ankle joint structure (body structure) | Inferred | | | | | | | | | | | x | |
| Finding site | Foot joint structure (body structure) | Inferred | | | | | | | | | | x | | |
| Finding site | Hand structure (body structure) | Inferred | | | | | x | | | | | | | |
| Finding site | Hip joint structure (body structure) | Inferred | | | | | | | | x | | | | |
| Finding site | Knee joint structure (body structure) | Inferred | | | | | | | | | x | | | |
| Interprets | Joint movement (observable entity) | Inferred | | x | | | | | | | | | | |
| Is a | Abnormal finding on evaluation procedure (finding) | Inferred | x | | | | | | | | | | | |
| Is a | Arthropathy of knee joint (disorder) | Inferred | | | | | | | | | x | | | |
| Is a | Disorder of ankle joint (disorder) | Inferred | | | | , | | | | | | | x | |
| Is a | Disorder of hip joint (disorder) | Inferred | | | | | | | | x | | | | |
| Is a | Disorder of joint of foot (disorder) | Inferred | | | | | | | | | | x | | |
| Is a | Finding of hand region (finding) | Inferred | | | | | x | | | | | | | |
| Is a | On examination - abnormal joint movement (finding) | Inferred | | x | | | | | | | | | | |
| Is a | On examination - joint (finding) | Inferred | x | | | | | | | x | x | x | x | x |
| Is a | On examination - legs (finding) | Inferred | | | | | | | | x | x | x | x | |

Table 10. Stated and inferred relationships for 12 concepts, either mismatched or not mismatched, with an FSN of the pattern ‘On examination – X joint abnormal (disorder)’ where ‘X’ is either absent (indicated by ‘-’ in the header), or one of the words ‘multiple’, ‘elbow’, ‘wrist’, ‘hand’, ‘toe’, ‘neck’, ‘hip’, ‘knee’, ‘foot’, ‘ankle’, or ‘shoulder’.

X' here is either absent or one of: 'multiple', 'elbow', 'wrist', 'hand', 'toe', 'neck', 'hip', 'knee', 'foot', 'ankle', or 'shoulder'. Since these concepts are all about abnormal joints, and all (except 'multiple') mention specific anatomical locations, one might expect all these finding concepts to be linked to the specific relevant body structures through either stated or inferred 'Finding site' relationships. Indeed, some of these 12 are linked in this way. Some, however are not. For example, [164516006 | On examination - knee joint abnormal (disorder)] correctly has as its Finding site [49076000 | Knee joint structure (body structure)], while [164510000 | On examination - elbow joint abnormal (disorder)] has as its Finding site the less precise [39352004 | Joint structure (body structure)], even though an appropriate specific body structure concept [16953009 | Elbow joint structure (body structure)] is available. We have previously observed and described similar inconsistencies in the use of Finding site relationship statements among tumor finding concepts, where some concepts lack appropriate Finding site assertions linking them to the body structure they are about [9].

The abnormal joint concepts in question here are all highly similar because they are all about disorders of joints. This similarity is reflected in the similar syntactic structure of their FSNs, which all have the same patterns of words except for a single word difference that also names a joint body structure concept. For this reason, one might expect these concepts to also have very similar positions in the Is-a hierarchy. For instance, since they are all about joint abnormalities, they arguably should all be subsumed by the concept [164506003 | On examination - joint abnormal (disorder)]. Instead, only six of them ('elbow', 'hand', 'multiple', 'neck', 'toe', and 'wrist') are subsumed by this concept and the rest ('hip', 'knee', 'foot', 'ankle', 'shoulder') are not. However, those joint abnormal concepts that are subsumed by [164506003 | On examination - joint abnormal (disorder)] in 2017 are the same ones that have the less specific Finding sites (using only the inherited Finding site, [39352004 | Joint structure (body structure)]). As discussed more below, these concepts have recently become mismatched due to changes in the hierarchy.

Some mismatched concepts appear together in groups rather than being randomly distributed around the vast SNOMED CT hierarchy. For example, in 2017, the concept [164506003 | On examination - joint abnormal (disorder)] is mismatched because it is not subsumed by the 'disorder' corresponding concept (all of its subsumers are tagged 'finding'). Further, all six of its child concepts are themselves mismatched disorders whose FSNs start with 'On examination …' and end with '… joint abnormal (disorder)' such as [164510000 | On examination – elbow joint abnormal (disorder)]. That is, there is a small subhierarchy rooted at [164506003 | On examination – joint abnormal (disorder)] that is entirely mismatched in 2017. These concepts have been mismatched since the January 2016 release, but they were not mismatched earlier. Prior to 2016, the concept at the root of this small subhierarchy, 164506003, had as one of its parent

concepts [399269003 | Arthropathy (disorder)], a non-mismatched 'disorder' concept that has the 'disorder' corresponding concept among its subsumers. Note that this Arthropathy concept is still among the subsumers of non-mismatched joint abnormal concepts mentioned above, such as [164516006 | On examination - knee joint abnormal (disorder)]. In January 2016, the Is-a link between [164506003 | On examination - joint abnormal (disorder)] and [399269003 | Arthropathy (disorder)] was removed, causing the appearance of the seven mismatches described above. This transition is illustrated in Figure 4, which shows the relevant concepts in 2015 and in 2016.

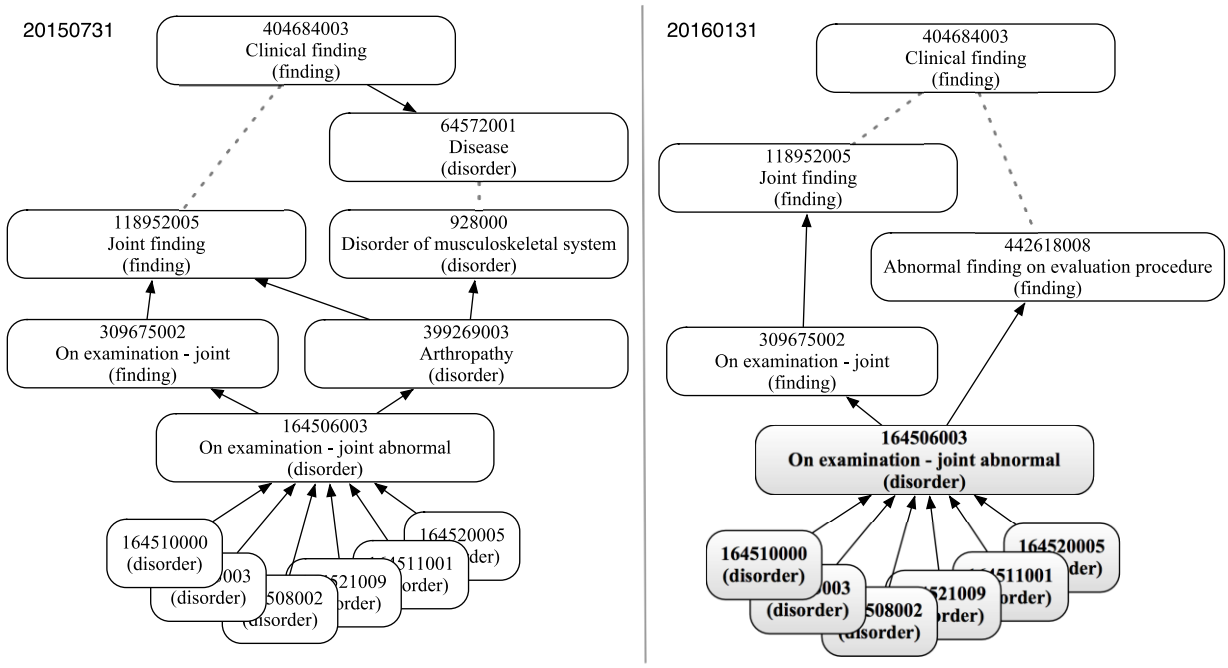



**Figure 4:** Introduction of clustered tag/concept mismatches in the 20160131 release.

In recent releases there is a similar cluster of mismatched concepts subsumed by [106150003 | Cranial nerve finding (finding)], though in that case the picture is more complicated than a single mismatched concept that subsumes only mismatches. Appendix A contains a complete listing of identified mismatched concepts in the January 2017 release of SNOMED CT.

### 4.4 Related work

Many different quality assurance and other evaluation methods have been applied to assess the quality of SNOMED CT's formal structure. Some are focused exclusively on that structure itself, as for example [10] which examines SNOMED's Specimen hierarchy, grouping concepts into Semantic Uniformity Groups based on their relationships, and finds tagging inconsistencies, especially among groups that overlap. [11]

presents a lattice-based structural auditing technique to identify structures in a single version of SNOMED that compromised its well-formedness. Also of this type are abstraction networks which provide a view of a terminology's contents at a higher level than the direct connections between concepts [12]. They have been used as the basis for terminology auditing methods to identify several general types of errors [13], as the basis for a visual auditing tools for SNOMED CT [14], and to identify errors in very complex concepts [15-17].

Some methods focus on changes in SNOMED CT's formal structure over time. [18] introduced a method based on ontological realism to calculate improvements in successive versions of biomedical ontologies. The approach emphasizes the distinction between three levels of reality and argues that changes in an ontology should be explained by their originators in these terms. [19] assesses the adequacy of the history mechanism distributed with SNOMED and recommends the development of an approach based on ontological realism to clarify and record the nature of changes. [20] shows how changes between two SNOMED versions affected a majority of concepts used in a legacy mapped interface terminology, including unexpected effects of structural changes in SNOMED, and argues for a consideration of impact on such implementations as part of terminology development. As discovered in this paper, changes in semantic tag assignment are one unexpected effect of such structural changes.

A third category of approaches includes the linguistic information embedded in SNOMED CT terms and compares linguistic similarities in SNOMED CT terms with expected similarities in the formal structure, for instance [7] and [8]. [21] identifies errors, predominantly missing Is-a relations, with an approach that combines structural information based on discovered subgraphs and lexical patterns in concept descriptions. [22] identifies sets of concepts within the Procedure hierarchy whose descriptions exhibit lexical similarity, but that are modeled inconsistently using the formal (logical) definition.

The work presented here is unique in its approach based on analyzing the connections between the formal SNOMED CT concept hierarchy and its use of semantic tags as parts of concept descriptions that are supposed to match the hierarchy. It is a continuation of earlier efforts in which we examined patterns of semantic tag changes between releases of SNOMED CT and observed that certain change patterns occur frequently among certain subsets of the total set of semantic tags [9]. One such active subset includes the tags 'disorder' and 'finding,' which are two of the main tags we have found here to be assigned to mismatched concepts.

## 5 Future work

We continue ongoing work in the analysis of semantic tags, their changes over time, and their connections to the formal concept hierarchy. Though the work described here doesn't address this specifically, one avenue of exploration examines the shape of the semantic tag hierarchy as encoded in the subsumption

relations that hold between concepts that use the tags in a release. Some tags subsume others in predictable ways (e.g. “finding” concepts commonly subsume “disorder” concepts, as seen in **Figure 4**), but other less predictable connections can also be observed. Multiple inheritance in SNOMED CT complicates this analysis. A closely related question is whether it is the case that all descendants of a corresponding concept have, or should have, the matching semantic tag (or one of its descendants in the tag hierarchy).

We also continue to examine patterns of semantic tag changes over time and, relatedly, patterns of tag mismatch changes between releases of SNOMED CT. Further analysis of the appearance or removal of groups of related mismatches in some releases, including possibly future releases, will help to reveal correctable errors.

## 6 Conclusion

We have successfully implemented algorithms to map semantic tags to corresponding SNOMED CT concepts and to identify and categorize mismatches between a concept’s semantic tag and its placement in SNOMED CT’s hierarchy. The results support our hypothesis that SNOMED CT indeed intends its semantic tags to have a one-to-one correspondence with certain high-level concepts. Nevertheless, mismatches tend to become more prevalent in later releases, specifically in the ‘disorder’ subhierarchy. This is a sign that the SNOMED CT authoring tool is not equipped with a formal mechanism to keep the hierarchy consistent with the semantic tags. It is our recommendation that such mechanism would be implemented and the method developed here might be a good starting point. Since the FSNs of SNOMED CT are intended to be used as interface technology in, for instance, electronic healthcare record systems, mistakes of the sort discovered should not occur.

**Acknowledgements**

none

**Funding**

This work was supported in part by (1) the National Institutes of Health through Clinical and Translational Science Award NIH 1 UL1 TR001412-01, the Department of Veterans Affairs and National Cancer Institute Big Data-Scientist Training Enhancement Program (BD-STEP), and the Department of Veterans Affairs award #80307 ‘Identifying clinical and logical shortcomings in SNOMED CT’, Project #1144333.

**Appendices**

Appendix A: Mismatched SNOMED CT concepts in the January 2017 release

**Supplemental files**

none

## References


1. Ceusters, W., *SNOMED CT's RF2: Is the Future Bright?* Studies in Health Technology and Informatics, 2011. **169**: p. 829-833.
2. IHTSDO, *International Health Terminology Standards Development Organization - SNOMED CT® Technical Implementation Guide - January 2015 International Release (US English)*. 2015. p. 757.
3. Ceusters, W. and J.P. Bona, *Analyzing SNOMED CT's Historical Data: Pitfalls and Possibilities.* AMIA Annual Symposium Proceedings, 2016. **2016**: p. 361-370.
4. IHTSDO. *SNOMED CT Editorial Guide*. 2017 [cited 2017 October 1]; Available from: https://confluence.ihtsdotools.org/display/DOCEG/SNOMED+CT+Editorial+Guide.
5. Scheuermann, R.H., W. Ceusters, and B. Smith, *Toward an Ontological Treatment of Disease and Diagnosis.* AMIA Summit on Translational Bioinformatics, 2009. **2009**: p. 116-120.
6. Bishop, B., et al., *OWLIM: A family of scalable semantic repositories.* Semant. web, 2011. **2**(1): p. 33-42.
7. Ceusters, W., B. Smith, and J. Flanagan. *Ontology and medical terminology: Why description logics are not enough*. in *Towards an Electronic Patient Record (TEPR 2003)*. 2003. San Antonio: Medical Records Institute.
8. Ceusters, W., et al., *Ontology-Based Error Detection in SNOMED-CT®.* MEDINFO, 2004. **11**: p. 482-486.
9. Bona, J.P. and W. Ceusters, *Identifying Missing Finding Site Relations in SNOMED CT.* AMIA Annual Symposium Proceedings, 2016. **2016**: p. 1347.
10. Wei, D., M. Halper, and G. Elhanan, *Using SNOMED semantic concept groupings to enhance semantic-type assignment consistency in the UMLS*, in *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*. 2012, ACM: Miami, Florida, USA. p. 825-830.
11. Zhang, G.-Q. and O. Bodenreider, *Large-scale, Exhaustive Lattice-based Structural Auditing of SNOMED CT.* AMIA Annual Symposium Proceedings, 2010. **2010**: p. 922-926.
12. Halper, M., et al., *Abstraction networks for terminologies: Supporting management of "big knowledge".* Artificial Intelligence in Medicine, 2015. **64**(1): p. 1-16.
13. Min, H., et al., *Auditing as Part of the Terminology Design Life Cycle.* Journal of the American Medical Informatics Association, 2006. **13**(6): p. 676-690.
14. Ochs, C., J.T. Case, and Y. Perl, *Analyzing structural changes in SNOMED CT's Bacterial infectious diseases using a visual semantic delta.* J Biomed Inform, 2017. **67**: p. 101-116.
15. Ochs, C., et al., *Scalable quality assurance for large SNOMED CT hierarchies using subject-based subtaxonomies.* Journal of the American Medical Informatics Association, 2015. **22**(3): p. 507-518.
16. Wang, Y., et al., *Auditing complex concepts of SNOMED using a refined hierarchical abstraction network.* Journal of Biomedical Informatics, 2012. **45**(1): p. 1-14.
17. Geller, J., et al., *New Abstraction Networks and a New Visualization Tool in Support of Auditing the SNOMED CT Content.* AMIA Annual Symposium Proceedings, 2012. **2012**: p. 237-246.
18. Ceusters, W. and B. Smith, *A realism-based approach to the evolution of biomedical ontologies.* AMIA Annual Symposium Proceedings, 2006. **2006**: p. 121-125.
19. Ceusters, W.M., K.A. Spackman, and B. Smith, *Would SNOMED CT benefit from Realism-Based Ontology Evolution?* AMIA Annual Symposium Proceedings, 2007. **2007**: p. 105-109.
20. Wade, G. and S.T. Rosenbloom, *The impact of SNOMED CT revisions on a mapped interface terminology: Terminology development and implementation issues.* Journal of Biomedical Informatics, 2009. **42**(3): p. 490-493.
21. Cui, L., et al., *Mining non-lattice subgraphs for detecting missing hierarchical relations and concepts in SNOMED CT.* Journal of the American Medical Informatics Association, 2017. **24**(4): p. 788-798.

22. Agrawal, A. and G. Elhanan, *Contrasting lexical similarity and formal definitions in SNOMED CT: consistency and implications.* J Biomed Inform, 2014. **47**: p. 192-8.

**Appendix A: Mismatched SNOMED CT concepts in the January 2017 release**



| Concept ID | Fully specified name |
|---|---|
| 109186003 | Sickle cell test kit (substance) |
| 711283001 | Cognitive behavior therapy by unidisciplinary team (regime/therapy) |
| 711284007 | Assessment by uniprofessional team (regime/therapy) |
| 711290006 | Assessment by multiprofessional team (regime/therapy) |
| 722299009 | Step up change in telehealth monitoring (regime/therapy) |
| 440245005 | Dressing medicated with leptospermum honey (product) |
| 5248000 | Supraglottic edema (disorder) |
| 5793009 | Mixed behavior and emotional disorder (disorder) |
| 9412006 | Tongue deviation disorder of twelfth cranial nerve (disorder) |
| 10068001 | Sensory somatic cortical disorder (disorder) |
| 12056002 | Seventh cranial nerve autonomic disorder (disorder) |
| 13343009 | Localized functional disorder (disorder) |
| 20734000 | Psychologic conversion disorder (disorder) |
| 21659007 | Generalized functional disorder (disorder) |
| 28456004 | Subclinical infection (disorder) |
| 31964003 | Lifelong psychologic disorder (disorder) |
| 33832002 | Third division of fifth cranial nerve disorder (disorder) |
| 35141006 | Cochlear nerve disorder (disorder) |
| 38801001 | Vagal gastric disorder (disorder) |
| 40661001 | Joint formation disorder (disorder) |
| 46997001 | Osteoid formation disorder (disorder) |
| 47924005 | Motor cortical disorder (disorder) |

| | |
|---|---|
| 55481000 | Limbic disorder (disorder) |
| 59979003 | Second division of fifth cranial nerve disorder (disorder) |
| 63101008 | Discrimination disorder (disorder) |
| 63864004 | Tongue protrusion disorder of twelfth cranial nerve (disorder) |
| 64089003 | Epiphysis closure disorder (disorder) |
| 65372005 | Cartilage resorption disorder (disorder) |
| 68921007 | Bone resorption disorder (disorder) |
| 72858000 | Speech cortex disorder (disorder) |
| 73466007 | Parasympathetic cardiovascular function disorder (disorder) |
| 80199008 | Sporadic disorder (disorder) |
| 87058001 | Vagus nerve motor disorder (disorder) |
| 89742000 | Vagus nerve autonomic disorder (disorder) |
| 89751008 | Sexual pain disorder (disorder) |
| 90104009 | Epiphysis formation disorder (disorder) |
| 164506003 | On examination - joint abnormal (disorder) |
| 164508002 | On examination - multiple joint abnormal (disorder) |
| 164510000 | On examination - elbow joint abnormal (disorder) |
| 164511001 | On examination - wrist joint abnormal (disorder) |
| 164513003 | On examination - hand joint abnormal (disorder) |
| 164520005 | On examination - toe joint abnormal (disorder) |
| 164521009 | On examination - neck joint abnormal (disorder) |
| 164582004 | On examination - lower leg bone abnormal (disorder) |
| 191798000 | Gender role disorder of adolescent or adult (disorder) |
| 230337001 | Motor tic disorder (disorder) |
| 247384001 | Neurological pain disorder (disorder) |
| 250054005 | Frontal gait disorder (disorder) |
| 265622002 | Equilibration disorder, vestibular nerve (disorder) |
| 312425004 | Infection of blood and lymphatic system (disorder) |

| | |
|---|---|
| 386585008 | Functional disorder (disorder) |
| 408311002 | On examination - retinopathy (disorder) |
| 408312009 | On examination - referable retinopathy (disorder) |
| 408313004 | On examination - non-referable retinopathy (disorder) |
| 431193003 | Infection of bloodstream (disorder) |
| 707734002 | Elevated liver enzymes level due to cystic fibrosis (disorder) |
| 708343006 | Temporomandibular joint popping on opening (disorder) |
| 708484008 | Vesiculoerosive lesion (disorder) |
| 710230000 | Painful os peroneum syndrome (disorder) |
| 711263002 | Pelvic floor dysfunction (disorder) |
| 711285008 | Loss of vertical dimension of occlusion due to worn complete denture (disorder) |
| 711599006 | Entropion of lower eyelid co-occurrent with ectropion of lower eyelid (disorder) |
| 711615005 | Entropion of upper eyelid co-occurrent with ectropion of upper eyelid (disorder) |
| 712734004 | Vertigo due to brain injury (disorder) |
| 713889004 | Atypical odontalgia (disorder) |
| 717889000 | Hypergastrinemia caused by drug (disorder) |
| 718362003 | Functional movement disorder (disorder) |
| 722875003 | Functional dysphagia (disorder) |
| 722878001 | Functional belching disorder (disorder) |
| 722880007 | Functional anorectal disorder (disorder) |
| 3761000119104 | Hypotestosteronism (disorder) |
| 17701000119108 | Noncompliant neuropathic bladder (disorder) |
| 32941000119104 | Ingestion of toxic substance (disorder) |
| 72631000119101 | Human immunodeficiency virus (HIV) II infection category B2 (disorder) |
| 76981000119106 | Human immunodeficiency virus (HIV) infection category B1 (disorder) |
| 76991000119109 | Human immunodeficiency virus (HIV) infection category B2 (disorder) |
| 96531000119109 | Deformity of hand due to rheumatoid arthritis (disorder) |
| 97881000119105 | Adrenal incidentaloma (disorder) |

| | |
|---|---|
| 100211000119106 | Muscle spasm of thoracic back (disorder) |
| 102031000119109 | Paratesticular mass (disorder) |
| 367761000119105 | Oligozoospermia caused by drug therapy (disorder) |
| 367781000119101 | Oligozoospermia co-occurrent and due to obstruction of efferent duct (disorder) |
| 367791000119103 | Oligozoospermia caused by radiation (disorder) |
| 367801000119102 | Oligozoospermia due to systemic disease (disorder) |
| 368311000119105 | Reflex neuropathic bladder (disorder) |
| 1085381000119108 | Cytomegalovirus viremia (disorder) |
| 1085741000119102 | Contour of existing restoration of tooth biologically incompatible with oral health (disorder) |
| 13790001000004101 | Bacteremia caused by Methicillin resistant Staphylococcus aureus (disorder) |
| 29930001000004103 | Intractable low back pain (disorder) |