9

‘Pain’ in SNOMED CT: Is There an Anesthetic?

Werner Ceusters and Jonathan P. Bona

1 Introduction

For sure, an unpleasant odor of feet may be categorized as an unpleasant odor. Though foot fetishists (De Block and Adriaens 2013), podiatrists and manufacturers of washing machines might disagree (Question Everything 2015), most people would contend such odor to be classified as an offensive odor. No clinician would have qualms in classifying an unpleasant odor of feet as an odor of feet nor as a foot finding. Perhaps some, in particular the ontology-savvy ones who are therefore able to detect ambiguities in natural language phrases, might doubt a foot odor to be a limb finding or a body odor. But who would argue it to be a finding

W. Ceusters (✉)
AU1 University at Buffalo, State University of New York, Buffalo, NY, USA

J.P. Bona
Division of Biomedical Ontology, Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, State University of New York, Buffalo, NY, USA

L. Zaibert (ed.), *The Theory and Practice of Ontology*,
DOI 10.1057/978-1-137-55278-5_9

of sense of smell, i.e. a neurological finding, or a duplicate, and therefore, inactive concept? This question smells of *The Systematized Nomenclature of Medicine—Clinical Terms* (*SNOMED CT*) (Donnelly 2006) all over, does it not? Indeed, the former is what its authors argued to be the case until 2003 when they discovered that there are actually two different sorts of smelly feet: "duplicate concept" ones and "offensive body odor" ones. What this discovery means for patients about whom SNOMED CT-based smelly feet assertions were made in their electronic healthcare records (EHR) is rather unclear. Some might, upon chart inspection, be surprised to find their smelly feet to have changed from a neurological finding to an offensive body odor. Others might find their smelly feet to have become inactive concepts and perhaps therefore conclude they are documented as being successfully treated. But in both cases their feet themselves are as stinky as before.

The problem we are dealing with here is the underestimated complexity of representing the evolution of terminologies and ontologies over time, in this case, of SNOMED CT. SNOMED CT is developed by the International Health Terminology Standards Development Organization (IHTSDO) and is the largest healthcare terminology currently available. It is supported by a concept-based ontology which can formally be represented by means of a description logic. It is worthwhile pointing out that SNOMED CT's authors have thus far not satisfactorily acted upon the confusions around what the word "concept" might denote (Smith 2004). Although "concept" in the SNOMED CT documentation is defined as "a clinical idea to which a unique concept identifier has been assigned", the term is also homonym for the concept identifier as well as for "the real-world referent(s) of the concept identifier, that is, the class of entities in reality that the concept identifier represents" (IHTSDO 2015, 725). To avoid any confusion, we will perceive for the purposes of this paper a version of SNOMED CT as an information content entity (ICE) of which concretizations exist as information artifacts in the form of, for example, data structures that can be rendered as tables on a computer screen by using appropriate software. We will use the term "SNOMED CT concept"—or "concept" for short—exclusively to denote any smaller information artifact which is part of such concretization and in which

inheres an information quality which is about or intended to be about some portion of reality (Smith and Ceusters 2015).

Roughly 400,000 SNOMED CT concepts are classified under several hierarchies, of which the top classes roughly correspond either to the types of entities clinicians encounter instances of during their work (body parts, organisms, diseases, substances, procedures, etc.) or to types instantiated by descriptive components of SNOMED CT as an ICE itself, for example those denoted by the terms "inactive concept", "navigational concept", and "metadata".

SNOMED CT is regularly updated (Ceusters 2011), not only to correct mistakes (Geller et al. 2012; Ochs et al. 2015), but also to represent better and in more detail how the entities in reality denoted by SNOMED CT concepts relate to each other. Updates are also made to account for changes in biomedical reality itself as well as in our scientific knowledge about biomedical reality (Ceusters 2010).

The work described here is part of a larger endeavor intended to find out whether it would be possible to use the growing number of historic relationships and other changes documented in SNOMED CT release files as an information source to detect mistakes that have not been discovered thus far. Since neither author of this paper suffers from smelly feet, but rather of an occasional headache while dealing with biomedical and other ontologies, we were interested to see how SNOMED CT concepts related to *pain* evolved throughout different versions and whether certain patterns of errors could be detected.

2 The Distribution of SNOMED CT Versions

International versions of SNOMED CT are biannually distributed by the IHTSDO in January and July as a set of release files designed to be loaded into healthcare software applications such as electronic healthcare record systems. Certain countries endorsing the use of SNOMED CT transform the international version into local adaptations. In the USA, it is the National Library of Medicine (NLM) that develops US versions as an extension of the international versions usually within 3 months of the international releases. The NLM makes both the international and

US versions available to authorized users as part of the Unified Medical Language System (UMLS) (Fung et al. 2005).

Whereas prior to July 2011 all releases were in a format now known as "Release Format 1 (RF1)" current releases are also available in the newer RF2-format. Core files included in both formats are (1) the *concepts table*, (2) the *descriptions table* containing terms associated with concepts, and (3) the *relationships table* which contains information on how the concepts relate to each other. Entries, i.e. rows, in each of these tables are called "*components*".

RF1 releases include also a *component history table* in which any changes such as additions and inactivations introduced in the concepts and descriptions tables—but not the relationships table!—over subsequent versions are logged. Since July 2008 RF1 releases come also with a *references table* which contains references from inactive components to other equivalent or related components that were current in the release version in which that component was inactivated.

In the RF2 format these changes are tracked in a uniform manner in the core files themselves, including the relationships table, but not for changes that occurred prior to 2002. A more extensive change history can only be computed on the basis of the original RF1 releases prior to July 2011 in addition to the RF1 (or with no additional advantage for the work described here, RF2) releases since July 2011.

RF1 versions consist of several tables, five of which are important for the work described here.

The *concepts table* of any version in RF1 includes for each concept (1) a SNOMED CT internally unique concept identifier, (2) whether it is in active use in the current version and, if not, the reason for withdrawal, and (3) whether the concept is primitive or fully defined in terms of the description logic used. Examples of two distinct concepts are "60932006: Buttock pain (finding)" and "279043006: Pain in buttock (finding)".

The *descriptions table* contains for each concept a varying number of description records each of which consists of the following data elements:

1. a unique identifier for the description,
2. a status marker indicating whether it is in active use and, if not, the reason for withdrawal from current use,

3. the unique identifier of the associated concept,
4. a term used to describe the associated concept and,
5. an indication of whether this specific term for the concept to which this description applies is:

- the Fully Specified Name (FSN), e.g. "Backache (finding)",
- the preferred term, e.g. "Backache", or
- a synonym, e.g. "Back pain" and "Pain in back".

Each FSN term ends with a "semantic tag" in parentheses "which indicates the semantic category to which the concept belongs" (e.g. clinical finding, disorder, etc.) and which "helps to disambiguate the different concepts which may be referred to by the same commonly used word or phrase' (IHTSDO 2015, 41). Examples of semantic tags are provided in Table 9.1. Although most semantic tags correspond each to some unique SNOMED CT concept, their taxonomic structure does not follow the taxonomic structure of the concepts.

The *relationships table* contains relations that obtain between SNOMED CT concepts. These relationships are expressed by means of existentially restricted triples of the form "source concept–relationship–target concept"—note that in the citation that follows "concept" is to be understood in the SNOMED sense—whereby each triple "implies that there is some instance of that relationship from each instance of the source concept to any instance of the target concept" (IHTSDO 2015, 678). For example, a triple of the form "*x* partOf *y*" is to be understood as: forall *x*: instance-of (*x, X*) ➞ exists *y*: instance-of (*y, Y*) and partOf(*x,y*).

Two types of such relationships are included in the release files. The first ones are called "*stated relationships*" and are relationships that are directly edited in the formal terminology management system by SNOMED CT's authors. Examples, leaving out the concept unique identifiers, are:

No genitourinary pain (situation): (E1)
Is a (attribute) = Clinical finding absent (situation),
Temporal context (attribute) = Current or specified time (qualifier value),
Associated finding (attribute) = Genitourinary pain (finding),
Finding context (attribute) = Known absent (qualifier value),
Subject relationship context (attribute) = Subject of record (person)

Table 9.1 Examples of SNOMED CT concepts related to pain

Semantic tag[a]	Leaf example[b]	Non-leaf example
Disorder	Phantom pain following amputation of penis	Disorder characterized by pain
Finding	Complaining of a headache	Pain
Situation	Pain behavior present	No genitourinary pain
Procedure	Pain relief	Pain management
Observable entity	Brief pain coping inventory score	Characteristic of pain at anatomical site
Product	Aromatic analgesic	Drugs used in neuropathic pain
Regime/therapy	Back pain prevention education	—[c]
Navigational concept	Analgesics and non-steroidal anti-inflammatory drug allergy	Additional pain and sensation observations
Substance	—	Analgesic
Physical object	Pain management medication delivery system pump	Anesthesia equipment
Qualifier value	Painless	Pain management service
Assessment scale	Pain coping strategies questionnaire	—
Environment	Pain clinic	—
Occupation	Pain management specialist	—
Attribute	Character of pain	—
Context-dependent category	On examination—in pain	—
Event	[X] Pain due to internal orthopedic prosthesis	Analgesic and/or antipyretic and anti-rheumatic drug poisoning
Staging scale	Chest pain rating	—

Notes

[a]The semantic tags are ranked in descending order of occurrence of pain-related SNOMED CT concepts.

[b]The column "Leaf examples", in contradistinction to "non-Leaf examples", exhibits SNOMED CT concepts that do not subsume other concepts.

[c]Empty slots indicate that for this category no occurrences were found in any of the versions studied

and,

Adnexal tenderness absent (situation): (E2)
Is a (attribute) = Clinical finding absent (situation)
Associated finding (attribute) = Adnexal tenderness (finding),
Temporal context (attribute) = Current or specified time (qualifier value),
Finding context (attribute) = Known absent (qualifier value),
Subject relationship context (attribute) = Subject of record (person)

The second type of relationships—"*inferred relationships*"—are obtained through inference by applying the EL++ description logic classifier which is part of SNOMED CT's ontology authoring system on the stated relationships (Dentler et al. 2011). An example is (E3) which is obtained by inference on the basis of (E1) and (E2):

Adnexal tenderness absent (situation): (E3)
Is a (attribute) = No abdominal pain (situation)
Is a (attribute) = No genitourinary pain (situation)
Is a (attribute) = Tenderness absent (situation)
Associated finding (attribute) = Adnexal tenderness (finding),
Temporal context (attribute) = Current or specified time (qualifier value),
Finding context (attribute) = Known absent (qualifier value),
Subject relationship context (attribute) = Subject of record (person)

Additional so called "*historic relationships*" are found in the *references table* where each such reference indicates the nature of the relationship between the inactive and persistent component.

Examples (E4)–(E14) in Table 9.2 indicate that in versions prior to and including the version in which these historic relationships appeared, there were five distinct SNOMED CT concepts that represented one or more types of entities in reality that clinicians colloquially would refer to by means of the words "back pain" or "backache". Three of these concepts—the ones with the identifiers 373644009, 399079008, and 419258005—were named "Back pain (finding)"; a fourth one—16986008—carried the FSN "Back pain (disorder)"—and the fifth one—161891005—was

Table 9.2 Examples of historic relationships

373644009: Back pain (finding): SAME AS (attribute) = 399079008: Back pain (finding)	(E4)
161891005: Backache (finding): SAME AS (attribute) = 399079008: Back pain (finding)	(E5)
161891005: Backache (finding): SAME AS (attribute) = 373644009: Back pain (finding)	(E6)
419258005: Back pain (finding): SAME AS (attribute) = 161891005: Backache (finding)	(E7)
399079008: Back pain (finding): MAY BE A (attribute) = 419258005: Back pain (finding)	(E8)
399079008: Back pain (finding): MAY BE A (attribute) = 161891005: Backache (finding)	(E9)
16986008: Back pain (disorder): MAY BE A (attribute) = 398997008: Vertebrogenic pain syndrome (disorder)	(E10)
16986008: Back pain (disorder): MAY BE A (attribute) = 399194009: Disorder characterized by back pain (disorder)	(E11)
16986008: Back pain (disorder): MAY BE A (attribute) = 419258005: Back pain (finding)	(E12)
16986008: Back pain (disorder): MAY BE A (attribute) = 399079008: Back pain (finding)	(E13)
16986008: Back pain (disorder): MAY BE A (attribute) = 161891005: Backache (finding)	(E14)
267984001: Backache, unspecified (finding): WAS A (attribute) = 161891005: Backache (finding)	(E15)
15941001: Brachialgia (disorder): REPLACED BY (attribute) = 102556003: Pain in upper limb (finding)	(E16)

named "Backache (finding)". A series of relational assertions were then made for various purposes:

1. to eliminate redundancies—"*x* SAME AS *y*" asserts that the SNOMED CT concepts *x* and *y* denote the same entity in reality, whereby starting with the version in which this relationship appears y would not anymore be used as an active SNOMED CT concept;
2. to eliminate erroneous or ill-defined concepts while keeping track of how they were classified in previous versions (*x* WAS A *y*) and what they were replaced by (*x* REPLACED BY *y*), if by anything at all, and,
3. to indicate which concepts were found to be ambiguous and, whenever they would have been used to annotate patient data, which concepts should be considered as unambiguous alternatives (*x* MAYBE A *y*).

The *component history table* contains for each changed description or concept (1) the unique identifier for the changed component, (2) the version of SNOMED CT in which this change was made, each version being represented using the format YYYMMDD, e.g. "20040731", (3) an indication of the nature of the change such as "added" or "status change", and (4) the status of the component after the change examples being "current", "retired", "duplicate", etc.

The RF1 tables contain more information than described above, though not relevant for the work reported on here.

3 Methodology

For our analysis, we used the concepts, descriptions, component history and references tables in the RF1 release of the US adaptation released March 31, 2015, which includes the international version released by the IHTSDO January 31, 2015. These files allowed us to compute changes that occurred at the level of concepts and descriptions. To track changes in the relationships, we used the RF1 relationship files of *all* international versions from January 2002 to January 2015, as well as all US adaptations since 2011.

3.1 Generation of Intermediate Tables

Several intermediate tables had to be constructed for the intended analyses. As a first step, a *Historic Relationships Table* (HISREL) was constructed which provides a complete history for each relationship that has appeared in any SNOMED CT release, one per row. It was created by merging the relationships tables of each SNOMED CT version into a single table wherein the existing columns were preserved and an additional column for each release date added. Each row represents a single relationship, and is marked in each date column to indicate whether that relationship was part of the release on that date.

HISREL was further reduced into the *Historic Subsumption Table* (HST) by retaining only those rows containing one of the following relationships: Is a (SNOMED CT's formal subsumption relation), ISA (mapping), MAY BE A, REPLACED BY, SAME AS, and WAS A. We will use the term "historically subsumed" as in "*x* is historically subsumed by *y*" whenever we refer to any of these relations holding between *x* and *y*.

The *Pain Terms Table* (PTT) is a manually curated list of SNOMED CT concepts that are about or mention pain in one or other form. An initial version was generated from a search for descriptions in SNOMED CT's description table containing any of the following substrings: "dynia", "algesia", "algaesia", "dolor", "algia", "algic", "esthesia", "esthaesia", "hyperpathia", "hyperpathic", "hypopathia", "hypopathic", "pain", "nocicept", "noxious", "hurt", "ache", "aching", "sore", "soring", "tender", and "throb". This list of terms was then manually filtered through several passes to exclude false matches such as "Paint (substance)". We also filtered it to exclude entries that were causing the subsumer table to be polluted because of (apparent) cycles caused by collapsing the history (e.g. rheumatism).

The *PTT Subsumer Table* (PTTST) is a list of all the concepts which historically subsume at least one of the concepts in the PTT. That is, it contains every concept "taxonomically above" any concept in the PTT. This table was further reduced to a *Leaf Nodes Table* (LEAFS) containing all and only leaf nodes from PTTST, i.e. those concepts in PTTST that do not relate to any other concepts via one of the selected relations for inclusion in HST.

A *DATES* table was constructed to capture all release version dates used in SNOMED CT, 37 in total, including 11 release dates that preceded SNOMED CT's first official release (January 31, 2002) and of which traces were found in SNOMED CT's component history table. Indeed, SNOMED CT was created in 2001 by merging what was then known as "SNOMED RT" with the UK Clinical Terms project (Wang et al. 2001). To keep track of which terms came from where, including which were already active or retired *before the* merger and which were duplicates in the first release *because of* the merger, two dates—20020129 and 20020130, corresponding respectively to the 10th and 11th date in the DATES table—were artificially created without actually corresponding to a physical release.

The *Pain Graph Nodes Table* (PGN) contains a taxonomy constructed from the bottom up, starting with the concepts in LEAFS. PGN includes every concept in LEAFS, as well as every concept that historically subsumed any of those concepts. That is, it contains every concept that lies along a path from a concept in LEAFS to the SNOMED root concept via any of the historic subsumption relations in HST, collapsing the relations from all release versions into a single graph. It also accounts for concepts that were replaced or considered as alternatives for inactivated versions by making use of "Replaced By" and "Alternative" entries as collected from SNOMED CT's references tables. PGN was then used as a filter on the concept table of SNOMED CT's last version used in this endeavor thereby copying into a new table (PCONC) only those records about concepts which were are also in PGN.

Similarly, PGN was used to filter the *Pain Historic Subsumption Table* (PHST) out of HST and the *Pain Descriptions Table* (PDT) out of the most recent SNOMED CT Descriptions table. For PHST this was achieved by retaining all and only records from HST expressing a relationship in which one or both of the relata are in the PGN table. To capture "Replaced by" and "Alternative" relations between concepts, PHST also includes additional entries for each concept in PGN that is the subject of a historic reference entry in the references tables.

The *Historic Pain Concept Table* (HPCONC) is built from the concept tables of each processed release. For each concept in PCONC, HPCONC contains one row for each release in which that concept was included in

the concept table. For instance, a pain concept in PCONC that appears only in the US National concept tables for releases 20140301, 20140901, and 20150301, would have three rows in the HPCONC table, one per release. HPCONC has columns for the concept ID, primitiveness, and release date in which the concept appears. Although one could argue that there is no need to combine the concept tables of each version as the last version should contain the total history, we preferred to take the safe way in light of the IHTSDO's motivations to develop RF2 because of several inconsistencies in RF1 releases (IHTSDO 2015, 663–664).

The *Pain Component Table* (PCT) contains all records from SNOMED CT's most recent Component History table for which the component identifier corresponds either to a concept identifier in PGN or PDT, or to a relationship identifier in PHST.

The *Pain History Table* (PHIS), finally, brings together information contained in the tables described above into a single structure with historic and taxonomic information about pain concepts and related concepts.

Table 9.3 contains the historic information about three concepts with the FSN "Pain (finding)". Two of them (367206007 and 366981002) are annotated as being retired in the (fictitious) 11th version as a result of the merger. The other one (22253000) existed in SNOMED RT prior to the merger as indicated by the "1" in the CUR field which corresponds with the earliest date of which a trace was found: January 1, 1994. The table shows also the various SNOMED CT concepts that subsume—see the "Is a (attribute)" in the INFOTYPE column—this concept, including one with the FSN "Pain finding (finding)" during the period covered by the first four versions.

3.2 Data Analysis

Several types of analysis have been—and are still being—carried out. The ones we report on here involve the changes in and evolution of the semantic tags in the FSNs. To that end we retrieved from PHIS all records indicating a change in the FSN, whether or not including a change in the semantic tag. For example, the following four records from PHIS show that in the 17th version the semantic tag for "Pain in lumbar spine" was

Table 9.3 Historic information about three SNOMED CT concepts for "pain" accumulated in the PHIS table

Concept ID	L[a]	Infotype	Component	Label	CIN[b]	CUR[c]	DUP[d]	RET[e]	Relation present in version[f]
22253000	0	Concept status	22253000			1			
22253000	0	Is primitive	22253000			1			
22253000	0	Synonym	2162242017	Pain observations		12			
22253000	0	Preferred term	37361011	Pain		12			
22253000	0	Synonym	37362016	Pain, NOS				12	
22253000	0	Synonym	37363014	Dolor		12			
22253000	0	Synonym	481278012	Painful		12			
22253000	0	Synonym	481279016	Part hurts		12			
22253000	0	FSN	751640015	**Pain (finding)**		12			
22253000	0	Is a (attribute)	19019007	Symptom (finding)					11111111000000000000000000000000000000
22253000	0	Is a (attribute)	279075009	Pain finding (finding)					11110000000000000000000000000000000000
22253000	0	Is a (attribute)	421833007	Finding of pain sense (finding)					00000000011100000000000000000000000000
22253000	0	Is a (attribute)	106147001	Sensory nervous system finding (finding)					11110000000000000000000000000000000000
22253000	0	Is a (attribute)	276435006	Pain / sensation finding (finding)					00001111100011111111111111111111111111
367206007	1	Concept status	367206007			10	11		
367206007	1	Is primitive	367206007			10			
367206007	1	FSN	2744729017	**Pain (finding)**	25				

(*continued*)

Table 9.3 (continued)

Concept ID	L[a]	Infotype	Component	Label	CIN[b]	CUR[c]	DUP[d]	RET[e]	Relation present in version[f]
367206007	1	Synonym	490781010	Pain	25				
367206007	1	Preferred term	773821016	Pain				25	
367206007	1	SAME AS (attribute)	22253000	Pain (finding)					011111111111111111111111111111111111
367206007	1	Is a (attribute)	363662004	Duplicate concept (inactive concept)					111111111111111111111111111111111111
366981002	1	Concept status	366981002			10	11		
366981002	1	Is primitive	366981002			10			
366981002	1	FSN	2734046010	**Pain (finding)**	25				
366981002	1	Synonym	490539018	Pain	25				
366981002	1	Preferred term	773572019	Pain				25	
366981002	1	Is a (attribute)	363662004	Duplicate concept (inactive concept)					111111111111111111111111111111111111
366981002	1	SAME AS (attribute)	22253000	Pain (finding)					011111111111111111111111111111111111

Notes

[a]**L**: whether the concept is present (1) or not (0) in the *LEAFS* table

[b]**CIN**: Concept Inactivated

[c]**CUR**: CURrent

[d]**DUP**: DUPlicated

[e]**RET**=RETired. Numericals in CIN, CUR, DUP and RET stand for the version in which the component obtained the status.

[f]**Relation present in version**: indicates whether the relationship was present (1) or not (0) in the SNOMED CT versions, starting with date 20020131 (12th data in the *DATES* table), each version represented by one digit, organized chronologically.

changed from "disorder" to "finding", as indicated by "17" in both the RET(ired) column and CUR(rent) column.

Concept ID	Label	CUR	RET
267982002	Pain in lumbar spine (disorder)		17
267982002	Pain in lumbar spine (finding)	17	
43116000	Eczema (disorder)	13	
43116000	Eczema [Ambiguous] (disorder)		13

Similarly, for the SNOMED CT concept denoting eczema, an FSN name change was introduced in the 13th version by dropping the modifier "[Ambiguous]" without changing, however, the semantic tag.

We then annotated changes for these concepts as "disorder → finding" and "disorder → disorder" respectively. An exploratory statistical analysis was conducted to assess the extent to which changes of this sort were distributed significantly differently over concepts which directly mention "pain" or a lexical variant thereof as collected in the PTT table (all of which are also included in the PHIS table), versus those concepts in the PHIS table which are not in PTT. As a result of the methodology described in the section "Generation of Intermediate Tables" above, these concepts are either historical subsumers of the PTT concepts themselves, or descendants of PTT concepts with their historical subsumers. As an example, Figure ure 9.1 shows the historical subsumption taxonomy of the concept "405154001: Level of suffering (observable entity)". This concept is included in PHIS because it is historically subsumed by the concept "405161002: Pain level (observable entity)", i.e. from January 2004 until July 2005. As a consequence, also all other concepts displayed in Figure ure 9.1 are contained in PHIS.

3.3 Results

The historical subgraph of SNOMED CT extracted for our research includes 7,673 concepts (1.83 %) out of a total of 420,221 concepts that ever have been introduced up to the US national version of March 2015. They have been extracted on the basis of 2,164 concepts (28 % of 7,673) which directly mention "pain" in one or other form. They are historically

related by means of 26,511 relationships, 4,028 of which (15.2 %) being based on "was a", "maybe a", etc.

These 7,673 concepts were annotated by a total of 8,829 FSNs which include a semantic tag. Semantic tags were not always used in SNOMED CT's predecessors, so there are FSNs of inactivated concepts that do not have one, where, obviously, some concepts have more than one FSN.

Table 9.4 provides an overview of the various semantic tags that were initially assigned to the concepts in our extracted graph. It makes clear that the majority of the tags were not changed: only 809 FSNs, in comparison to 4,974 that remained active throughout SNOMED CT's history without any change, or that were inactivated without involving any change. Within the group of FSNs whose semantic tag was changed, nearly half (49.5 %) involved those annotated as "disorder" while another 31.8 % is accounted for by what originally was qualified as "finding". Three categories disappeared completely: "context-dependent category", "environment/location" and "function". An important number of changes—inactivations and semantic tag changes—can be noted for the large groups of "situation" (90.7 % of the 248 "situation" FSNs), "procedures" (31.6 % of 291 FSNs) as well as for "disorder" (31.4 %) and "finding" (30.4 %).

Table 9.5 gives an insight in what specific semantic tags were changed into, thereby excluding from the counts in Table 9.4 those changes with less than five occurrences in order to keep the table readable. Target tags excluded include "environment", "environment/location", "event", "linkage concept", "physical object", "qualifier value", and "substance". As a result of this elimination, some original tags were to be removed, resulting in a more compact table. What Table 9.5 tells us is that changes occurred in certain clusters. Notable are the reciprocal switches between "procedure" and "regime/therapy", and "disorder" and "finding". FSNs with semantic tag "context-dependent entity" were later in the first place classified under "finding" (67.8 %) and for the remainder mainly under "situation".

Table 9.6, finally, demonstrates that for almost all semantic tag changes, the FSNs directly mentioning "pain" in one or other lexical form as collected in the PTT table changed in different ways than in the FSNs of concepts that are not directly related to pain. Significantly more

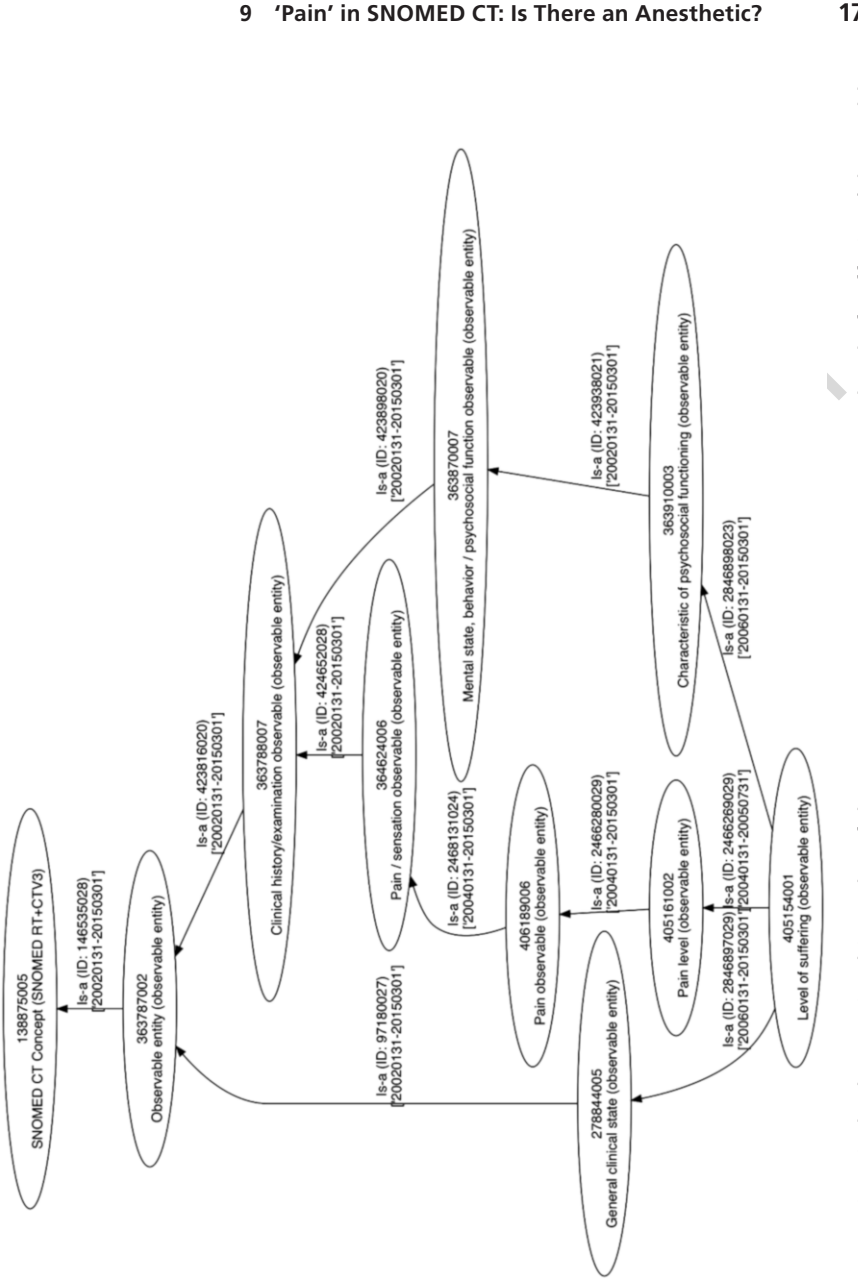

Figure 9.1 Historical subsumption hierarchy of the SNOMED CT concept "405154001: level of suffering (observable entity)". Relationships are annotated with the period during which they were active

Table 9.4 Changes in semantic tags over different SNOMED CT versions

	Unchanged active	Inactivated	Type change	Total	% of total	% changed	% type changes
	A	*B*	*C*	*D = A+B+C*	*E = D/* SUM(*D*)	*F = (B+C)/D*	*G =C/(B+C)*
Assessment scale	22	3	7	32	0.4	31.3	70.0
Attribute	11	0	2	13	0.2	15.4	100.0
Context-dependent category	0	163	273	436	5.5	100.0	62.6
Disorder	2,697	940	292	3,929	49.5	31.4	23.7
Environment	14	0	1	15	0.2	6.7	100.0
Environment/location	0	0	1	1	0.0	100.0	100.0
Event	2	2	1	5	0.1	60.0	33.3
Finding	1,754	678	89	2,521	31.8	30.4	11.6
Function	0	16	8	24	0.3	100.0	33.3
Observable entity	154	29	7	190	2.4	18.9	19.4
Physical object	31	1	3	35	0.4	11.4	75.0
Procedure	199	45	47	291	3.7	31.6	51.1
Product	21	11	2	34	0.4	38.2	15.4
Qualifier value	17	19	15	51	0.6	66.7	44.1
Regime/therapy	2	13	20	35	0.4	94.3	60.6
Situation	23	215	10	248	3.1	90.7	4.4
Social concept	1	0	1	2	0.0	50.0	100.0
Substance	26	16	30	72	0.9	63.9	65.2
Totals	**4,974**	**2,151**	**809**	**7,934**	**100.0**	**37.3**	**27.3**

Table 9.5 Evolution of semantic tags over different SNOMED CT versions

	Semantic tag							
TO:	Disorder	Finding	Navigational concept	Observable entity	Procedure	Product	Regime/ therapy	Situation
FROM:								
Assessment scale			10					
Attribute			5					
Context-dependent category	3.7	67.8	0.4	0.7	0.4			27.1
Disorder	75.7	15.8	8.2					0.3
Finding	32.6	50.6	6.7	1.1				9.0
Function				10				
Observable entity			14.3	71.4				
Procedure			2.1		19.1		76.6	
Product						10		
Qualifier value			6.7	53.3				
Regime/therapy					65.0		35.0	
Situation								10
Substance			3.3			86.7		
TOTAL FRACTION	32.1	34.1	5.3	3.0	2.8	3.5	5.3	11.5

Note

Changes FROM semantic tag *x* TO semantic tag *y* are quantified as the percentage of $x \rightarrow y$ changes over the totality of all $x \rightarrow$ changes. For example: of all the changes that occurred for SNOMED CT concepts that had the original semantic tag of "context-dependent category", 67.8 % were changed to "finding". Changes from $x \rightarrow x$ indicate FSN name changes without change in the semantic tag.

Table 9.6 Semantic tag changes in SNOMED CT concepts in whose FSN a pain term (or lexical variant thereof) is directly used, compared to semantic tag changes in the SNOMED CT concepts contained in the PHIS table *not* directly related to pain

Semantic tag change		Observed in PHIS[a]			Expected in PHIS		Chi square		Chi test
FROM	TO	+PTT	–PTT	Total	+PTT	–PTT	+PTT	–PTT	"p"
Context-dependent category	Disorder	3	7	10	2.26	7.74	0.24	0.07	0.5738
	Finding	**64**	**121**	**185**	**41.74**	**143.26**	**11.86**	**3.46**	**0.0001**
	Situation	**45**	**29**	**74**	**16.70**	**57.30**	**47.97**	**13.98**	**0.0000**
Disorder	*Disorder*	*4*	*217*	*221*	*49.87*	*171.13*	*42.19*	*12.29*	*0.0000*
	Finding	**29**	**17**	**46**	**10.38**	**35.62**	**33.40**	**9.73**	**0.0000**
	Navigational concept	*0*	*24*	*24*	*5.42*	*18.58*	*5.42*	*1.58*	*0.0082*
Finding	*Disorder*	*1*	*28*	*29*	*6.54*	*22.46*	*4.70*	*1.37*	*0.0138*
	Finding	**18**	**27**	**45**	**10.15**	**34.85**	**6.06**	**1.77**	**0.0051**
Procedure	*Regime/therapy*	*3*	*33*	*36*	*8.12*	*27.88*	*3.23*	*0.94*	*0.0411*
Regime/therapy	Procedure	1	12	13	2.93	10.07	1.27	0.37	0.1996
Situation	Situation	3	7	10	2.26	7.74	0.24	0.07	0.5738
Substance	*Product*	*0*	*26*	*26*	*5.87*	*20.13*	*5.87*	*1.71*	*0.0059*
Other changes[b]		**12**	**80**	**92**	**20.76**	**71.24**	**3.70**	**1.08**	**0.0289**

Notes

[a] "Observed in PHIS": actual counts of changes observed in relation to concepts in the PHIS table, thereby differentiating between concepts directly mentioning "pain" in one or other lexical form as collected in the PTT table (+PTT), versus those in PHIS not in the PTT table (–PTT).

[b] "Other changes" include all changes for which the total count in PHIS was <10.

Bold: statistically significant; *italics*: significant "with caution" because <5 occurrences in observed cells.

"pain terms" than statistically expected became classified as "findings" where they used to be under "context-dependent category" or "disorder". On the other hand, significantly more "pain terms" remained classified as "finding" in comparison to "non-pain terms" that were classified as "findings" but were then reclassified as "disorder". Remarkable also is that nearly three times more than expected (45 observed versus 16.7 expected) "pain terms" that were tagged as "context-dependent category" became later tagged as "situation". The differences are undoubtedly statistically significant for those changes printed bold in Table 9.6, while some statisticians prefer to remain cautious when individual counts are lower than 5, despite a Chi test result < 0.05. We have indicated these cases in italics.

4 Discussion

Many efforts have been made to measure the amount and type of changes occurring between SNOMED versions. Spackman (2005) categorized changes and measured the rate of changes in SNOMED over a three-year span (2002–2005), finding that the most change activity during that span was occurring among relationships, and in particular subsumption relationships, and concluding that implementers must "carefully examine mechanisms for handling this degree of change". Lee et al. (2011) examined changes in SNOMED over three years as recorded in the Component History and Concept Model, with a focus on the sub-set of concepts in the NLM CORE Problem List. Mortensen et al. (2014) identified errors and patterns of errors in the CORE Problem List sub-set of a single version of SNOMED by focusing on inferred "Is a" relations. Of the studied relations 19.5 % exhibited errors, many of which were not caught on the first pass by human domain experts. Tao et al. (2015) present an approach and analysis, using it to identify relation reversals (a particularly dramatic type of structural change) in the evolution of SNOMED, finding 48 such reversals since 2009.

To our best knowledge, no research has thus far been done on SNOMED CT's semantic tags. Semantic tags are claimed to have been introduced in SNOMED CT "to help disambiguate different concepts which may be referred to by the same commonly used word or phrase".

For example, “Hematoma (morphologic abnormality)” is “the FSN of the concept that represents the hematoma that a pathologist sees at the tissue level. In contrast, ‘Hematoma (disorder)’ is the FSN of the concept that represents the clinical diagnosis that a clinician makes when they decide that a person has a hematoma” (IHTSDO 2015, 41). Semantic tags are not part of the formal taxonomic structure of SNOMED CT, although most of them are closely related to one or other taxonomic category. The tag “finding”, for instance, appears prominently—perhaps exclusively, we did not investigate this thus far—in the FSN of concepts subsumed by the concept “Clinical finding (finding)”. So, is the tag “situation” part of the FSNs of concepts subsumed by the concept “Situation with explicit context (situation)”? The concept “Clinical finding (finding)” subsumes, inter alia, the concepts “Disease (disorder)” and “Deformity (finding)” which in its turn subsumes, inter alia, the concepts “Deformed pupil (finding)”, “Corneal deformity (disorder)”, and, astonishingly, also “Complaining of a deformity (finding)” thereby thus implying that complaining of a deformity is a special kind of deformity in its own right. Also amazing is that not all concepts with the semantic tag “disorder” are subsumed by the concept “Disease (disorder)”.

If, at this point, it becomes hard to understand, then that is because, in our opinion, it is not understandable at all. One could indeed wonder why there is not a taxonomic category “Disorder (disorder)” which subsumes all “Disease (disorders)” plus those under other hierarchies. The absence of such a category is even more astonishing in light of some reflections we find in IHTSDO (2015, 275–276):

> Clinical findings have been defined as observations, judgments or assessments about patients. The problem with the terms finding and observation is that they seem to refer to the judgment of the observer rather than to the actual state of the body. Examples of clinical findings include: difficulty swallowing, nose bleed, diabetes, headache, and so forth. More precise and reproducible definitions of clinical findings, and the precise boundaries between findings and events, between findings and observables, between findings and situations, and the distinction between finding and disorder, remain ongoing challenges at the margins. The distinction between a disorder and an observation has proven to be difficult to define in a reproduc-

> ible manner across the tens of thousands of concepts included under clinical findings. *Nevertheless, there are several reliable characteristics of each sub-category (disorders and findings)*. [emphasis added]

Yet, there are no subcategory disorders at all, there is only the semantic tag "disorder"!

It is clear that SNOMED CT suffered—and still does suffer—dramatically from the adherence to concepts such as "Clinical finding (finding)" and "Observable entity (observable entity)". Clinical findings are stated "to represent the result of a clinical observation, assessment or judgment, and include both normal and abnormal clinical states" (IHTSDO 2015, 275). Observable entities, so we are told, "can be thought of as representing a question or procedure which can produce an answer or a result". Observables are considered to be partial observation results, where there is a defined part of the observation missing. In many cases, what is missing is a numeric value, or a numeric value with units. "In other cases, the observable is like a question, and what is missing can be regarded as the answer" (IHTSDO 2015, 316). This explains why, for example, "Pain threshold (observable entity)" carries the semantic tag "observable entity" and "Decreased pain threshold (finding)" the tag "finding". It fails to explain why "Threshold (qualifier value)" does not carry the tag "observable entity".

5 Conclusion

SNOMED CT has undoubtedly come a very long way since its original conception as a mere nomenclature for pathology (Major et al. 1978; Sommers 1967). The IHTSDO has been working very hard on developing editorial and technical principles for updating SNOMED CT and on training its terminologists in applying the principles faithfully. The significantly larger number of changes introduced in pain-related terms compared to non-pain terms as observed in our research are most likely the result of bringing order in what once was chaos; chaos not only created because of the inherent complexity of pain as a clinical entity—pain indicates that some abnormality is present, yet it is not necessarily abnor-

mal itself—(Smith et al. 2011) from which other terminologies than SNOMED CT are suffering as well (Ceusters 2014), but also because of the misplaced focus on observations and findings, thereby confusing existing entities on the side of the patient on the one hand, with processes of observing these entities and representations/communications about what is believed to be observed on the other hand.

In this light it is encouraging to read that slowly, very slowly, some principles of the Open Biomedical Ontology Foundry (Smith et al. 2007) and ontological categories from the Basic Formal Ontology (Arp et al. 2015; Smith et al. 2005) are trickling down into SNOMED CT's concept model (IHTSDO 2015, 322). It would be even better if this model were to be based on the Ontology of General Medical Science (OGMS) (Scheuermann et al. 2009) which separates first-order entities (e.g. diseases, disorders, bodily features, processes of measuring and observing) clearly from second-order entities (diagnoses, representations).

So is there an anesthetic for the pains caused by SNOMED CT's concept model? We believe there is: the OGMS. Whether IHTSDO will believe there is a need for an anesthetic remains doubtful. After all, from stated relationship (E1), see the section "The Distribution of SNOMED CT Versions", the SNOMED CT's description logic infers:

No pain (situation):
Temporal context (attribute) = Current or specified time (qualifier value),
Associated finding (attribute) = Genitourinary pain (finding),
Finding context (attribute) = Known absent (qualifier value),
Subject relationship context (attribute) = Subject of record (person)

Thus when "no genitourinary pain" is the case, there is supposed to be no pain at all. Since neither author of this paper suffers either from stinky feet or genitourinary pain, it is according to SNOMED CT not possible that we would suffer from headache. If it were all that simple!

6 Epilogue

One can still wonder what smelly feet have to do with pain other than being emotionally painful when it is pointed out by one's environment. Figure 9.2 explains how the SNOMED CT concepts "unpleasant odor of feet" became part of the intermediate files that were constructed to arrive at our final collection for inspection. Before releasing the January 2004 version of SNOMED CT, it was discovered that the bodily feature of foot odor was represented twice: once by means of the SNOMED concept 102597005 and once by 394643003. During the entire period covered by the analysis presented here, at least one of these SNOMED concepts was subsumed by "Body odor finding" which itself was subsumed by "Finding of sense of smell" from January 2002 until (and including) January 2003. It is after that version, that "Finding of smell" became, in two versions, subsumed by "Pain/sensation finding", one of the "*x* or *y*" type of classes by which many biomedical classification systems, terminologies and ontologies are infested. Although this SNOMED concept—unfortunately—still exists, there is no version in which it actually subsumes either "unpleasant odor of feet" concept. These were pulled in in our analysis sets to be able to draw graphs such as partially drawn in Figure 9.2, which provides a view on the evolution of SNOMED CT concepts in a historical context.

Acknowledgements This work was supported in part by Clinical and Translational Science Award NIH 1 UL1 TR001412-01 from the National Institutes of Health, by grant R21LM009824 from the National Library of Medicine (NLM), and by grant 1R01DE021917-01A1 from the National Institute of Dental and Craniofacial Research (NIDCR). The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the NIDCR, the NLM or the National Institutes of Health.

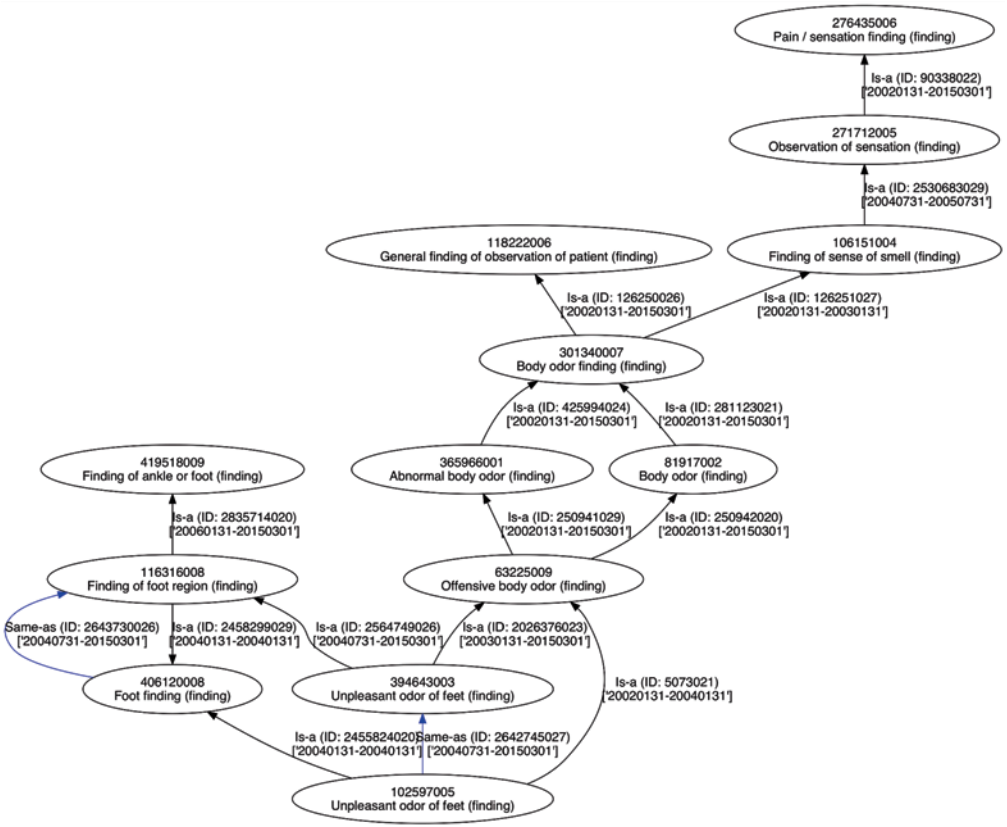

Figure 9.2 Partial historical subsumption hierarchy of the SNOMED CT concept "102597005: unpleasant odor of feet (finding)"

References

Arp, Robert, Barry Smith, and Andrew D. Spear. 2015. "Building ontologies with basic formal ontology." In Cambridge, Massachusetts: The MIT Press. ebrary http://site.ebrary.com.gate.lib.buffalo.edu/lib/buffalo/Doc?id=11085695

Ceusters, W. 2010. "Applying Evolutionary Terminology Auditing to SNOMED CT." *AMIA Annu Symp Proc* 2010:96–100.

Ceusters, W. 2011. "SNOMED CT revisions and coded data repositories: when to upgrade?" *AMIA Annu Symp Proc* 2011:197–206.

Ceusters, W. 2014. "Pain assessment terminology in the NCBO BioPortal: evaluation and recommendations." In *Proceedings of the International Conference on Biomedical Ontology 2014*, 1–6. Houston, TX.

De Block, A., and P. R. Adriaens. 2013. "Pathologizing sexual deviance: a history." *J Sex Res* 50 (3–4):276–98. doi:10.1080/00224499.2012.738259.

Dentler, Kathrin, Ronald Cornet, Annette ten Teije, and Nicolette de Keizer. 2011. "Comparison of reasoners for large ontologies in the OWL 2 EL profile." *Semantic Web* 2 (2):71–87.

Donnelly, K. 2006. "SNOMED CT: The Advanced Terminology and Coding System for eHealth." In *Studies in Health Technology and Informatics—Medical and Care Compunetics 3. Vol 121*, edited by Bos L, Roa L, Yogesan K, O'Connell B, Marsh A and Blobel B, 279–290. Amsterdam: IOS Press.

Fung, K. W., W. T. Hole, S. J. Nelson, S. Srinivasan, T. Powell, and L. Roth. 2005. "Integrating SNOMED CT into the UMLS: an exploration of different views of synonymy and quality of editing." *J Am Med Inform Assoc* 12 (4):486–94. doi:10.1197/jamia.M1767.

Geller, J., C. Ochs, Y. Perl, and J. Xu. 2012. "New abstraction networks and a new visualization tool in support of auditing the SNOMED CT content." *AMIA Annu Symp Proc* 2012:237–46.

IHTSDO. 2015. International Health Terminology Standards Development Organization - SNOMED CT® Technical Implementation Guide - January 2015 International Release (US English).

Lee, Dennis, Ronald Cornet, and Francis Lau. 2011. "Implications of SNOMED CT versioning." *International Journal of Medical Informatics* 80 (6):442–453. doi:10.1016/j.ijmedinf.2011.02.006.

Major, P., B. J. Kostrewski, and J. Anderson. 1978. "Analysis of the semantic structures of medical reference languages: part 2. Analysis of the semantic power of MeSH, ICD and SNOMED." *Med Inform (Lond)* 3 (4):269–81.

Mortensen, Jonathan M., Mark A. Musen, and Natalya F. Noy. 2014. "An empirically derived taxonomy of errors in SNOMED CT." *AMIA Annual Symposium Proceedings* 2014:899–906.

Ochs, C., J. Geller, Y. Perl, Y. Chen, J. Xu, H. Min, J. T. Case, and Z. Wei. 2015. "Scalable quality assurance for large SNOMED CT hierarchies using subject-based subtaxonomies." *J Am Med Inform Assoc* 22 (3):507–18. doi:10.1136/amiajnl-2014-003151.

Question Everything. 2015. Is it normal to like smelly feet? https://www.youtube.com/watch?v=gw-rWFadU7M

Scheuermann, R. H., W. Ceusters, and B. Smith. 2009. "Toward an ontological treatment of disease and diagnosis." *Summit on Translat Bioinforma* 2009:116–20.

Smith B, Ceusters W, Goldberg LJ, and Ohrbach R. 2011. "Towards an Ontology of Pain." In *Proceedings of the Conference on Logic and Ontology*, edited by Mitsu Okada, 23–32. Tokyo: Keio University Press.

Smith, B., M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall, O. B. I. Consortium, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S. A. Sansone, R. H. Scheuermann, N. Shah, P. L. Whetzel, and S. Lewis. 2007. "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration." *Nat Biotechnol* 25 (11):1251–5. doi: 10.1038/nbt1346.

Smith, B., W. Ceusters, B. Klagges, J. Kohler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, A. L. Rector, and C. Rosse. 2005. "Relations in biomedical ontologies." *Genome Biology* 6 (5). doi: Artn R46 Doi 10.1186/Gb-2005-6-5-R46.

Smith, Barry. 2004. "Beyond concepts: ontology as reality representation." In *Proceedings of the third international conference on formal ontology in information systems (FOIS 2004)*, 73–84. Amsterdam: IOS Press.

Smith, Barry, and Werner Ceusters. 2015. Aboutness: Towards Foundations for the Information Artifact Ontology. International Conference on Biomedical Ontology, Lisbon, Portugal, July 27–30.

Sommers, S. C. 1967. "Systematized nomenclature of pathology." *Pathol Microbiol (Basel)* 30 (5):826–7.

Spackman, Kent A. 2005. "Rates of Change in a Large Clinical Terminology: Three Years Experience with SNOMED Clinical Terms." *AMIA Annual Symposium Proceedings* 2005:714–718.

Tao, Shiqiang, Licong Cui, Wei Zhu, Mengmeng Sun, Olivier Bodenreider, and Guo-Qiang Zhang. 2015. "Mining Relation Reversals in the Evolution of SNOMED CT Using MapReduce." *AMIA Summits on Translational Science Proceedings* 2015:46–50.

Wang, A. Y., J. W. Barrett, T. Bentley, D. Markwell, C. Price, K. A. Spackman, and M. Q. Stearns. 2001. "Mapping between SNOMED RT and Clinical terms version 3: a key component of the SNOMED CT development process." *Proc AMIA Symp*:741–5.

Author Queries

Chapter No.: 9 0002829782

Queries	Details Required	Author's Response
AU1	Please check if the affiliations are presented correctly. Also first author has been treated as corresponding author. Please provide email address for the corresponding author.	
AU2	Four entries of "10", and one of "5" are given as integers, not as, e.g., 10.0. Please confirm this is OK. If possible should be 10.0 for consistency	
AU3	No semantic tag for thi last row in Table 9.5. Should it be "Other changes"?	

AU1: email corresponding author = wceusters@gmail.com
The affiliation is for both authors the same. It should be for both how it is written out for Bona.
AU2: You may change '10' and '5' into '10.0' and '5.0' resp.
AU3: No, it should be 'TOTAL FRACTION'. Add a sentence in the table's legend stating: 'TOTAL FRACTION indicates the percentage of FSNs of which the semantic tag changed to the semantic tag in the corresponding 'TO' column, irrespective of what the original semantic tag was'.