

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/44804148>

Construction of an Interface Terminology on SNOMED CT Generic Approach and Its Application in Intensive...

Article in *Methods of Information in Medicine* · June 2010

DOI: 10.3414/ME09-01-0057 · Source: PubMed

CITATIONS

15

READS

88

5 authors, including:



Leila Ahmadian

Kerman University of Medical Sciences

20 PUBLICATIONS 121 CITATIONS

SEE PROFILE



Ronald Cornet

Academisch Medisch Centrum Universiteit va...

115 PUBLICATIONS 1,987 CITATIONS

SEE PROFILE



Evert de Jonge

Leiden University Medical Centre

265 PUBLICATIONS 6,941 CITATIONS

SEE PROFILE



Nicolette F de Keizer

Academisch Medisch Centrum Universiteit va...

229 PUBLICATIONS 4,003 CITATIONS

SEE PROFILE

Construction of an Interface Terminology on SNOMED CT

Generic Approach and Its Application in Intensive Care

F. Bakhshi-Raiez¹; L. Ahmadian^{1, 2}; R. Cornet¹; E. de Jonge³; N. F. de Keizer¹

¹Department of Medical Informatics, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands;

²Department of Medical Records, Kerman University of Medical Sciences, Kerman, Iran;

³Department of Intensive Care, Leiden University Medical Centre, Leiden, The Netherlands

Keywords

Terminological system, interface terminology, SNOMED CT, subsetting, APACHE IV

Summary

Objective: To provide a generic approach for developing a domain-specific interface terminology on SNOMED CT and to apply this approach to the domain of intensive care.

Methods: The process of developing an interface terminology on SNOMED CT can be regarded as six sequential phases: domain analysis, mapping from the domain concepts to SNOMED CT concepts, creating the SNOMED CT subset guided by the mapping, extending the subset with non-covered concepts, constraining the subset by removing irrelevant content, and deploying the subset in a terminology server.

Results: The APACHE IV classification, a standard in the intensive care with 445 diagnostic categories, served as the starting point for designing the interface terminology. The majority (89.2%) of the diagnostic categories from APACHE IV could be mapped to SNOMED CT

concepts and for the remaining concepts a partial match was identified. The resulting initial set of mapped concepts consisted of 404 SNOMED CT concepts. This set could be extended to 83,125 concepts if all taxonomic children of these concepts were included. Also including all concepts that are referred to in the definition of other concepts lead to a subset of 233,782 concepts. An evaluation of the interface terminology should reveal what level of detail in the subset is suitable for the intensive care domain and whether parts need further constraining. In the final phase, the interface terminology is implemented in the intensive care in a locally developed terminology server to collect the reasons for intensive care admission.

Conclusions: We provide a structure for the process of identifying a domain-specific interface terminology on SNOMED CT. We use this approach to design an interface terminology on SNOMED CT for the intensive care domain. This work is of value for other researchers who intend to build a domain-specific interface terminology on SNOMED CT.

terminologies, and interface terminologies, each used for different purposes and each serving different requirements regarding their intended use and domain [2–4]. Aggregate terminologies, such as ICD-9 CM, are used to categorize disease encounters into disjoint classes [3, 5, 6]. The lack of structure and formal semantic definitions for classes in most aggregate terminologies however results in shortcomings when aiming for data re-use [7]. Reference terminologies, such as SNOMED CT, are meant to address these re-use issues by providing precise meaning by formal concept definitions (i.e. semantic definitions), required for consistent and computer-readable coding and storage of clinical data [8, 9]. Finally, interface terminologies are used for actual data entry into electronic medical records, facilitating display and collection of clinical data in a simple way while simultaneously linking user's own descriptions to structured data elements in a reference terminology or aggregate terminology [8].

SNOMED CT is regarded the most comprehensive terminological system for coding clinical information [10, 11]. Its wide coverage and semantic structure aim at making it applicable as a reference terminology while the mappings to existing aggregate terminologies such as ICD-9 CM enable aggregate terminology features. SNOMED CT also includes several terms for each concept that can be used in an interface terminology for data entry. However, its use as an interface terminology in a clinical setting is the subject of discussion in many studies (e.g. [8, 12, 13]). In case all of SNOMED CT is provided to users for systematic data collection, its comprehensiveness forms an impediment, as it contains large amounts of concepts that are irrelevant for most clinical domains. Fur-

Correspondence to:

Ferishta Bakhshi-Raiez
Academic Medical Centre
University of Amsterdam
Dept. of Medical Informatics
P.O. Box 22700
1100 DE Amsterdam
The Netherlands
E-mail: f.raiez@amc.uva.nl

Methods Inf Med 2010; 49: 349–359

doi: 10.3414/ME09-01-0057

received: June 29, 2009

accepted: March 3, 2010

prepublished: June 22, 2010

1. Introduction

Systematic recording of clinical data is central to the use, exchange, analysis, and interpretation of this data. To this end, numerous terminological systems have been de-

veloped. A terminological system interrelates concepts of a particular domain and provides concepts with related terms and possibly definitions and codes [1]. Terminological systems can be distinguished as aggregate terminologies, reference ter-

thermore, the extensiveness of SNOMED CT, covering all kinds of medical domains, does not guarantee that SNOMED CT completely covers the detailed information necessary for data collection in a specific clinical setting [8, 12–14]. Accordingly, instead of providing all of SNOMED CT to the users, an interface terminology on SNOMED CT is proposed to provide easier adoption to the user requirements, i.e. the non-relevant content of SNOMED CT is excluded and the relevant concepts and terms which are not present in SNOMED CT are added to the interface terminology [15]. Despite its advantages for structured data entry, no formal methods to develop a domain-specific interface terminology on SNOMED CT have been described. The present medical informatics literature has dealt with different parts of this process, such as mapping from domain concepts to SNOMED CT concepts [16], terminology modularization techniques [17]. However, thus far no papers concern the methodology underlying the process of designing an interface terminology on SNOMED CT as a whole. To fill this gap, the objective of this paper is to provide a generic approach for developing a domain-specific interface terminology on SNOMED CT. An application to the domain of intensive care is used to illustrate this process.

The paper is organized as follows. Section 2 provides an overview of the sub-processes of developing a domain-specific interface terminology on SNOMED CT. Section 3 presents an application of this process to the domain of intensive care. Section 4 discusses the merits and limitations of the study. Section 5 concludes the paper.

2. Methods

2.1 SNOMED CT

SNOMED CT is the world's largest concept-based terminological system. The January 2008 release, which was used in this study, contains 284,777 active concepts associated with 737,695 active terms and interrelated by 860,865 active relationships, which can be hierarchical (i.e. Is-A relationships) or non-hierarchical (i.e. any of about 60 attribute relationship types such

as “finding site” or “causative agent”). SNOMED CT is a compositional terminology, i.e. it supports post-coordination, the use of composite expressions of concepts to define and refine (new) concepts [18–21].

2.2 Interface Terminology Development Process

We propose that the process of developing an interface terminology on SNOMED CT involves six sequential phases: domain analysis, mapping from domain concepts to SNOMED CT concepts, creating the SNOMED CT subset based on the mapping results, extending the subset with non-covered concepts (i.e. user-defined concepts, relationships, and terms which are not included in SNOMED CT), constraining the subset by eliminating irrelevant content, and deploying the subset in a terminology server. In the following sections each of these phases is briefly described. An overview of the process and the output of each phase is provided in ► Figure 1.

2.2.1 Domain Analysis

It is important that the interface terminology provides the level of granularity needed by the users. Therefore, to develop a domain-specific interface terminology on SNOMED CT, the relevant parts of SNOMED CT need to be extracted. Domain analysis is defined as the knowledge acquisition process by which the relevant information for a particular interface terminology is identified [22]. To this end, different resources, such as historical data in medical files, medical domain expert knowledge, and/or existing standards or domain ontologies can be used [15, 16, 23]. Using historical data is intricate and time-consuming; medical files, especially paper-based files, are often unstructured and there is a large variation in terminology used among care providers. Semi-automatic or fully automatic processing tools can facilitate the process of terminology acquisition from historical data [16, 24]. However, these tools are usually domain and language-specific, and, if not available, costly to build. Gathering the knowledge from medical domain experts is also not straight-

forward since these experts cannot be brought together frequently for knowledge elicitation sessions and they can overlook important concepts [22]. Existing standards or terminologies within the domain of interest can provide an ample starting point. These are usually based on expert agreement and provide a good overview of the knowledge needed within a certain domain.

2.2.2 Mapping from Domain Concepts to SNOMED CT Concepts

Mapping is defined as creating a link between the content of two systems, in this case the domain concepts derived in the domain analysis and the target-terminological system such as SNOMED CT, through semantic correspondence [6, 25]. Mapping can be done with different methods comprising (semi-)automated or manual procedures [25–27]. Manual mapping of large amounts of source concepts (e.g. when creating a mapping between a system like ICD-10-CM and SNOMED CT) is time-consuming and automated tools might be used to assist this task [26, 28]. Automated mapping tools are usually domain- and language-specific, and, if not available, costly to build. The level of confidence in automated mapping can vary from application to application [29, 30]. In all cases, manual review is required to validate the output of the automated mapping and to map the portions that failed the automated mapping [27, 29, 30].

In case of manual mapping, more than one reviewer is needed and agreement statistics need to be calculated to determine interrater reliability.

2.2.3 SNOMED CT Subset Creation

A SNOMED CT subset is a group of concepts, descriptions and/or relationships relevant for use in a given domain. To create the subset, in a database for instance, terminology modularization or segmentation techniques such as hierarchy traversal can be used [16, 17, 31].

In extraction by traversal, the SNOMED CT hierarchy is considered as a rooted directed acyclic graph and the subset is extracted by starting at a target concept and following its links to other concepts to in-

clude them in the subset [17]. Hereby, the structure of the SNOMED CT hierarchy remains intact in the subset. In downwards traversal, the algorithm only goes down the Is-A hierarchy from the target concepts, including all their subordinates. In upwards traversal, the algorithm only goes up the Is-A hierarchy from the target concepts, including all their superordinates. ▶ Figure 2 gives an illustration of these hierarchy traversal algorithms [17]. In downwards traversal or upwards traversal from at-

tributes, also the attributes of the target concepts are traversed.

The hierarchy traversal mechanism that is used for subsetting depends on the purpose for which the interface terminology will be used. In case the interface terminology is going to be used to record detailed information in medical files, the subordinates (downwards traversal) and the related attributes (downwards traversal from attributes) of the target concepts should be included.

2.2.4 Extending the Subset with Local Concepts, Relationships, and Terms

The advantages of creating an interface terminology on SNOMED CT are among others the possibilities to include concepts, relationships, and terms which are not present in SNOMED CT required by the user [15]. Such extensions can be placed into one of the three categories [32]: 1) Adding a new interface term, i.e. a term that does not

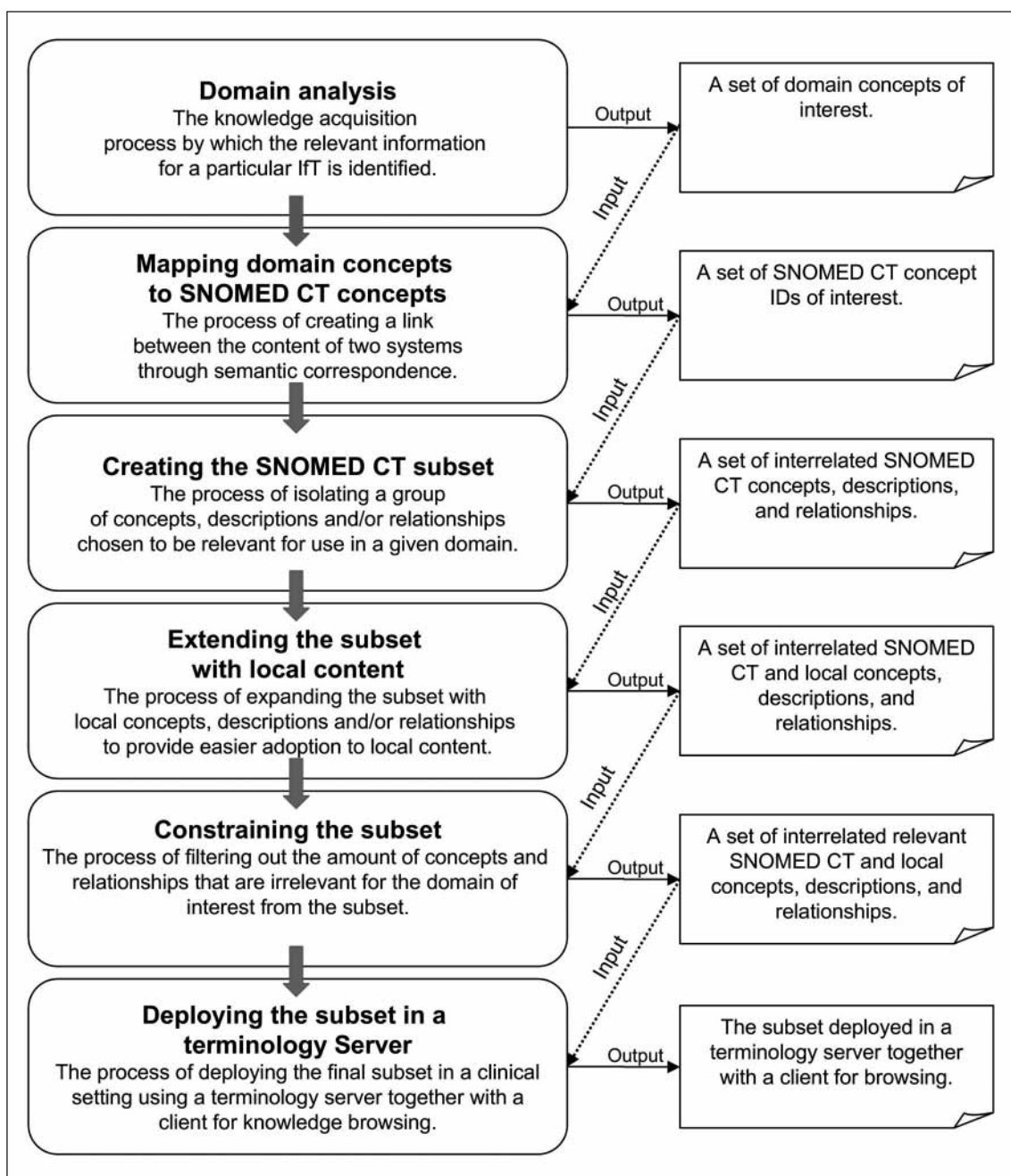


Fig. 1 Interface terminology development process

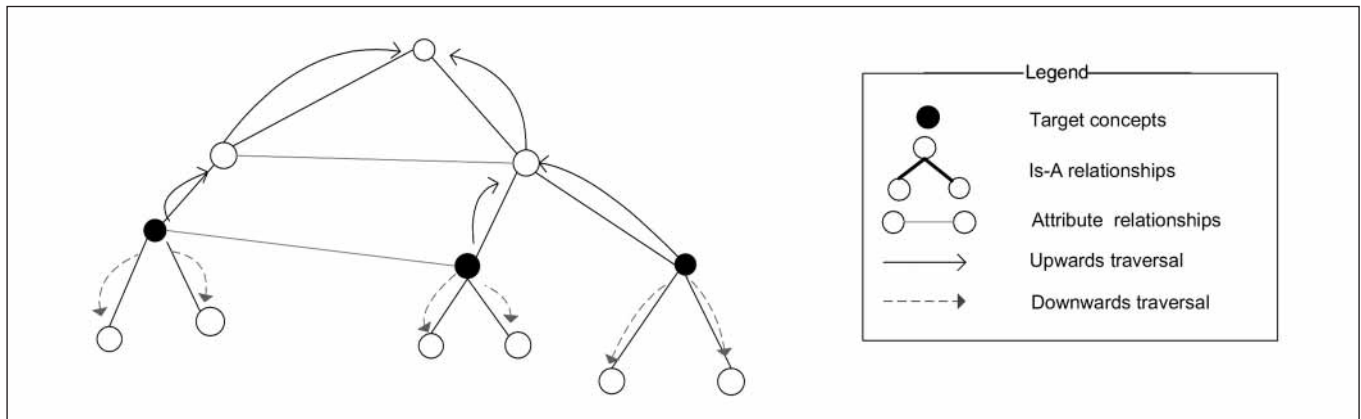


Fig. 2 Examples of hierarchy traversal algorithms: in downwards traversal the algorithm goes down the Is-A hierarchy from the target concepts, including all their subordinates, and in upwards traversal the algorithm goes up the Is-A hierarchy from the target concepts, including all their superordinates.

exist in SNOMED CT, but that describes an already existing SNOMED CT concept. 2) Adding a new leaf concept as a subordinate of an existing leaf concept, i.e. generation of an entirely new concept in the SNOMED CT hierarchy, along with its preferred and interface terms and relationships (► Fig. 3). 3) Adding a new node concept as a superordinate of existing concept(s) or leaf concept(s), i.e., the concept is absent, and the missing concept is a parent of one or more existing SNOMED CT concepts. In order to add the concept to the hierarchy, it should be “grafted” into some branches of the SNOMED CT tree (► Fig. 3).

2.2.5 Constraining the Subset

Subsetting SNOMED CT and subsequently extending the subset result in a (smaller) set of interrelated SNOMED CT and local concepts, descriptions and relationships. However, this subset may still contain a large amount of concepts and relationships that

are irrelevant and may require further constraining. This can be done by systematically filtering irrelevant content from the subset. The included non-human concepts in SNOMED CT, for instance, might be irrelevant for clinical domains and can be excluded from the subset. When one develops a subset of diagnoses one might restrict to concepts coming from the subtype hierarchy disease (disorder). Another possibility is limiting the graph depth, by terminating the downwards traversal when some criterion is reached, relating either to the depth of the graph or to the size of some other property of the target concept [17, 31]. It is also possible to clean the subset manually by removing the irrelevant and anomalous parts after an evaluation with the end users.

2.2.6 Deploying the Subset in a Terminology Server

Interface terminologies are designed to support interactions between humans and

structured medical information systems. To achieve this goal, human interfaces to computer applications, i.e. terminology servers, are needed to provide the content of an interface terminology to its users [33]. Together with a client for knowledge browsing, terminology servers can be used at the point of care to enter patient observations, findings, and events into the computerized patient records.

3. Case Study: An Interface Terminology on SNOMED CT for Intensive Care

In intensive care, diagnostic information is often recorded in different systems in free text or using a specific classification system resulting in registration insufficiency. For instance, calculation of case-mix-adjusted mortality risks in the Acute Physiology and Chronic Health Evaluation (APACHE) IV

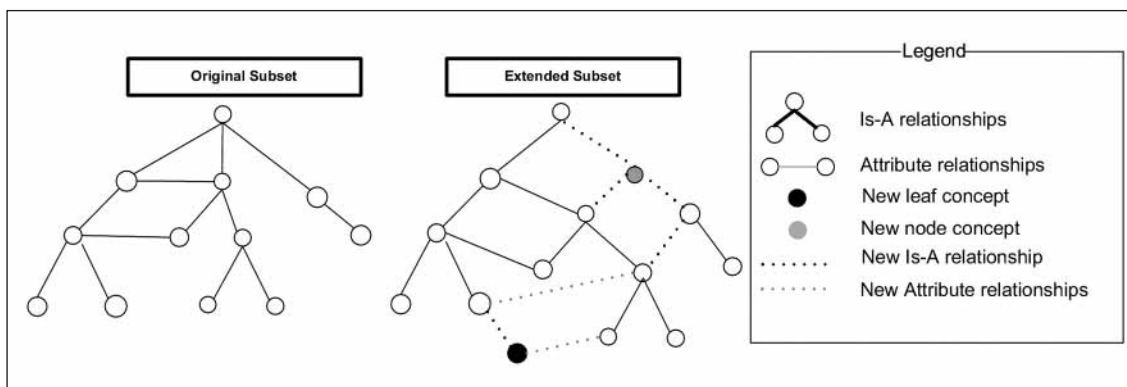


Fig. 3 Examples of subset extension: a new leaf concept and a new node concept are added to the hierarchy together with their relationships.

prognostic model requires a variety of patient information, such as physiological parameters, comorbidities and reason for ICU admission. The reason for ICU admission is captured using the APACHE IV classification [34]. However, the diagnostic categories in this APACHE IV classification lack the detail and structure needed for an unambiguous description of health problems. Therefore, the same information about the reason for intensive care admission is often separately registered in medical files and discharge letters, but with more detail and in free text. Hence, there is a need for a standard terminological system to support multiple use of single recorded diagnostic information. To facilitate this goal, the use of an interface terminology on SNOMED CT is proposed. For our case study this interface terminology should enable recording of detailed and structured reasons for intensive care admission in daily care practice, from which the corresponding APACHE IV category can be determined automatically. But it should also be useful for other purposes, e.g. aggregation of patient groups in research, automated discharge letters and triggering decision rules.

3.1 Domain Analysis

Using historical data for terminology acquisition was difficult as diagnostic information in intensive care is mostly recorded in free text. Automatic processing of this data was not feasible as there were no tools available to translate the Dutch intensive care free text entries into English terms or into terms from another language for which a SNOMED CT translation exists.

A few standards are available for the domain of intensive care. For instance, the reasons for intensive care admission required for the calculation of the APACHE II and APACHE IV prognostic models are collected using APACHE II reasons for admission classification including 54 diagnoses and the APACHE IV reasons for intensive care admission classification including 445 diagnoses, respectively. In general, the prognostic models serve the purpose of calculating case-mix-adjusted mortality rates to measure the quality of care

[35, 36]. However, the corresponding classifications are also used for aggregation of patient groups for retrospective research.

Within this case study we will use the APACHE IV classification to start identifying the intensive-care-specific SNOMED CT subset, as the APACHE IV classification is the most extended and widely used classification within the domain of intensive care [34]. In spite of its extensiveness and wide use, the use of the APACHE IV classification as an interface terminology is limited, due to the fact that no synonyms are used to describe the diagnostic categories, and because of its strict monohierarchical structure. In the APACHE IV classification, each diagnostic category is first classified as non-operative or post-operative, representing acute conditions leading to ICU admission and procedures after which patients are monitored on the ICU, respectively. Next, categories are distinguished by body system (e.g. cardiovascular, neurologic, respiratory) or a transplant- or trauma-related category, and then by diagnosis (e.g. “coronary artery bypass graft”, “bone marrow transplant”, “face injury”). A residual “other” category is used for unlisted diagnoses within each body system, transplant, and trauma category (e.g. “non-operative gastrointestinal disorder – other”).

3.2 Mapping from APACHE IV to SNOMED CT

Each of the 445 APACHE IV diagnostic categories was mapped to one or more SNOMED CT concepts. Given the relatively small amount of categories to be mapped, the mapping was performed manually.

Composite APACHE IV diagnostic categories (i.e. containing more than one diagnosis such as “chest trauma with spinal trauma”) were first split into atomic diagnoses (i.e. “chest trauma” plus “spinal trauma”). These atomic diagnoses were then each mapped to a SNOMED CT concept. Finally, the composite categories were represented as the conjunction of the mapped SNOMED CT concepts. Some atomic diagnoses are used in more than one composite diagnostic category. The atomic category

“chest trauma”, for instance, is part of 16 APACHE IV diagnostic categories in the trauma group. Each of such atomic diagnoses needed to be mapped only once to a SNOMED CT concept.

The concepts in SNOMED CT were navigated using the CliniClue^a browser (version 2006.2.30), a look-up engine for SNOMED CT concepts. Mapping consisted of three consecutive activities:

1. Interpreting and analyzing the meaning of the APACHE IV diagnostic categories. To resolve ambiguities in the diagnostic categories, the researchers consulted an intensivist. This was done for 10% (n = 45) of the diagnostic categories. For instance, from the structure of the APACHE IV classification it was not clear what the exact meaning of the diagnostic category “non operative heart transplant” was. It could be mapped to the SNOMED CT concept “planned operative procedure for heart transplantation” which is a preoperative concept or to “cardiac transplant disorder” which is a concept meant to describe possible complications of a previous heart transplantation. Intensivist consultation revealed that both are possible and therefore, both matches are included in the current mapping. The ambiguous terms encountered in all classes of the APACHE IV classification with a majority (12 categories) in the transplant class.
2. Matching one atomic APACHE IV diagnostic category to one or more active SNOMED CT concept(s). The diagnostic categories were first mapped to pre-coordinated concepts. The following three rules were applied to search for the right APACHE IV diagnostic categories in the SNOMED CT hierarchy: I) selection of the correct SNOMED CT category, e.g. non-operative APACHE IV diagnostic categories were searched for in the “disorder” subtype hierarchy of SNOMED CT and post-operative APACHE IV diagnostic categories were searched for in the “procedure” subtype hierarchy of SNOMED CT, II) selection of the appropriate search term, i.e. start

^a The Clinical Information Consultancy Ltd. <http://clininfo.cliniclue.com/home>

Table 1 Examples of the mapping types

	APACHE IV diagnosis category	SNOMED CT concept(s)
Complete match pre-coordinated	Colon cancer or rectal cancer	Malignant tumor of colon (disorder) Malignant tumor of rectum (disorder)
Complete match post-coordinated	Repair of ventricular aneurysm	<i>Is-A</i> : Repair of aneurysm <i>Procedure site – Direct</i> : cardiac ventricular structure
Partial match to a superordinate concept	Other gastrointestinal cancer	Malignant neoplasm of gastrointestinal tract (disorder)

with the APACHE IV term and in case of unsatisfactory results, continue the search with the corresponding synonyms and abbreviations, and III) verify the final mapping by examining the relationships of the SNOMED CT concept to ensure a correct concept-based mapping. In case no pre-coordinated match was available, a post-coordinated match was searched for. Post-coordination was based on the concept model of SNOMED CT and was not restricted to the functionalities of the CliniClue interface. To this end, post-coordination instructions in the SNOMED CT guides were followed [19, 20, 37]. In case an exact match could not be found in

SNOMED CT, the APACHE IV diagnostic category was mapped to an appropriate superordinate concept.

- Assessing each matched category-concept pair on how well they matched by marking each category-concept pair as “complete match” (i.e. an APACHE IV category matches to a semantically equivalent SNOMED CT concept), “non-match” (i.e. no semantically equivalent SNOMED CT concept is available) or “partial match” (i.e. matches to superordinate concepts).

The mapping process was performed independently by two medical informatics researchers (FBR and LA), both practiced in

coding medical data and with knowledge of the APACHE IV classification. Furthermore, the researchers were educated in SNOMED CT with extensive knowledge of the SNOMED CT guidelines beforehand. Before the mapping, there was a general agreement between the two researchers on the mapping activities, including how the quality of the matches would be defined and how post-coordinated matches would be represented. The final match was based on consensus between the two researchers. In case of disagreement, a third researcher (RC), a SNOMED CT expert, was involved. The interrater reliability, i.e. the percentage of the APACHE IV diagnostic categories that were similarly mapped to one or more SNOMED CT concepts by the two researchers, was 90%. For the remaining diagnostic categories, in half of the cases consensus was reached and half of the diagnostic categories were also searched for by the third researcher to achieve consensus.

At the end, the diagnostic categories in the APACHE IV classification were mapped to 404 atomic SNOMED CT concepts in the disorder and procedure subtype hierarchies. Additionally, 67 concepts were needed to post-coordinate concepts. ▶Table 1 provides some examples of the different mapping types. ▶Table 2 provides the results of the mapping between the APACHE IV classification and SNOMED CT. SNOMED CT provided complete matches for 89.2% of the diagnostic categories. There were no non-matches. For 10.8% of the diagnostic categories a partial match was found. Partial matches were all related to the diagnostic categories including the word “Other” (e.g. “Non-operative cardiovascular disorder – Other”). The use of post-coordination supported the good matching scores: 39.1% of all complete matches and 31.4% of all partial matches were realized through post-coordination.

3.3 Subset Creation in SNOMED CT

As mentioned the interface terminology will be used to record clinical data for daily care processes, automatic generation of discharge letters, and for calculation of APACHE IV mortality risks. Therefore, we

Table 2 Results of the mapping from APACHE IV classification to SNOMED CT

APACHE IV classes	Number of diagnostic categories	Number of complete match (%)		Number of partial match (%)		Number of “Other categories” in partial matches
		Pre	Post	Pre	Post	
Cardiovascular	101	63 (62.0)	25 (25.0)	10 (10.0)	3 (3.0)	13
Gastrointestinal	57	32 (56.0)	18 (32.0)	7 (12.0)	0 (0.0)	7
Genitourinary	35	18 (51.5)	13 (37.4)	0 (0.0)	4 (11.1)	4
Hematology	18	13 (72.0)	2 (11.0)	3 (17.0)	0 (0.0)	3
Metabolic	18	15 (83.0)	1 (6.0)	0 (0.0)	2 (11.0)	2
Musculoskeletal	20	12 (60.0)	5 (25.0)	3 (15.0)	0 (0.0)	3
Neurological	54	37 (69.0)	11 (20.0)	4 (7.0)	2 (4.0)	6
Respiratory	46	31 (67.0)	9 (20.0)	4 (9.0)	2 (4.0)	6
Transplant	24	16 (67.0)	6 (25.0)	2 (8.0)	0 (0.0)	2
Trauma	72	5 (7.0)	65 (90.0)	0 (0.0)	2 (.03)	2
Total	445	242 (54.4)	155 (34.8)	33 (7.4)	15 (3.4)	48

Table 3 Number of disorder and procedure concepts added to the subset for each iterative downwards traversal step

Downwards traversal step	Number of disorder and procedure concepts
0	404
1	5570
2	22606
3	44206
4	62212
5	73818
6	79705
7	82031
8	82806
9	83064
10	83122
11	83125

need to isolate the concepts related to the APACHE IV diagnostic categories and their subordinates and attributes by downwards traversal. To this end, SNOMED CT's core data structure was imported into a Microsoft Access Database to isolate the subset using SQL queries. The results of the map-

ping procedure (i.e. the mapped SNOMED CT concepts) were used as target concepts. From each target concept, downwards traversal and downwards traversal from the attributes was performed (► Fig. 2) [17]. All identified links and nodes were added to the subset during iterative steps. ► Table 3 shows the number of disorder and procedure concepts and the number of attributes added to these core concepts in the subset during each iterative downwards traversal step. If the graph traversal depth would be limited to, for example, two steps the 404 source concepts resulted in identification of 22,606 disorder and procedure concepts, but when the graph traversal depth would not be limited this would result in 83,125 disorder and procedure concepts in the subset to be used in the interface terminology. At the end, the Microsoft Access database included three SNOMED CT tables, i.e. concepts table, descriptions table and relationships table containing the SNOMED CT subset.

3.4 Extending the Subset with Local Concepts, Relationships, and Descriptions

In this phase of the development, the isolated SNOMED CT subset was extended

with local concepts, relationships, and descriptions to represent the APACHE IV diagnostic categories which were not covered (by pre-coordinated concepts) in SNOMED CT. The extensions were based on the categories described in Section 2.2.4. For all concepts representing a diagnostic category from the APACHE IV classification, the APACHE IV term with the related APACHE IV code was added as the preferred term. For diagnostic categories that were mapped to a post-coordinated concept, the post-coordinated match was added as a pre-coordinated concept using the SNOMED CT guides [19, 20, 37]. Also for the partial matches, a new pre-coordinated concept was added to the SNOMED CT subset. This step was necessary to enable the aggregation of the detailed clinical information to the appropriate APACHE IV categories required for calculation of APACHE IV mortality risks. Finally, the hierarchical relations between the APACHE IV categories were added to the subset to represent the APACHE IV classification tree in the subset. For instance, we added the class "APACHE IV Medical disorder" as a subordinate of the SNOMED CT concept "Clinical finding" and then added "Is-A" relations from all APACHE IV classes (i.e. system disorders,

Table 4 Examples of extension types, i.e. new content to the SNOMED CT subset

Extension types	Number	Examples		
		APACHE IV diagnostic category with the code	Relationship	Linked Concept(s) in the subset
New description	599	Repair of abdominal aortic aneurysm (228)	<i>Preferred term to</i>	405525004 repair of aneurysm of abdominal aorta
New leaf concept	168	Repair of ventricular aneurysm (227)	<i>Is-A</i> <i>Procedure site – Direct</i>	75087007 repair of aneurysm 21814001 cardiac ventricular structure
New leaf concept with multiple parents	113	Abdomen/face trauma (192)	<i>Is-A</i> <i>Is-A</i>	128069005 injury of abdomen 125593007 injury of face
New node concept	47	Carbon monoxide, arsenic or cyanide poisoning (146)	<i>Is-A</i>	75478009 Poisoning
Leaf concepts to the newly added node concept		17383000 Toxic effect of carbon monoxide 64189001 Arsine poisoning 66207005 Toxic effect of cyanide	<i>Is-A</i> <i>Is-A</i> <i>Is-A</i>	Carbon monoxide, arsenic or cyanide poisoning (146) Carbon monoxide, arsenic or cyanide poisoning (146) Carbon monoxide, arsenic or cyanide poisoning (146)

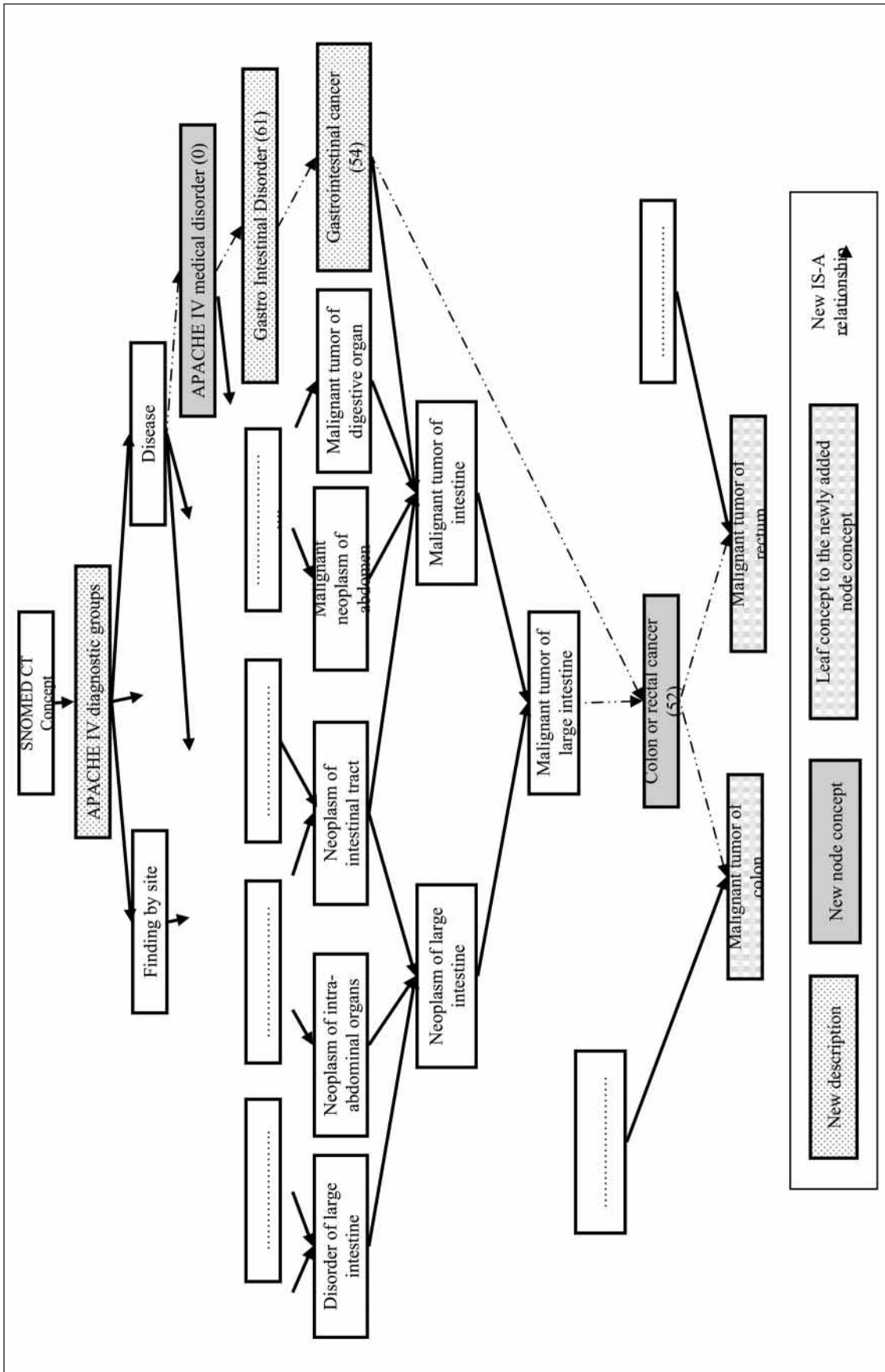


Fig. 4 Subset-extension examples in SNOMED CT

transplant, and trauma categories) to this class. This step was necessary to enable creation of aggregated patient data based on APACHE IV categories.

The type and the amount of extensions to the extracted subset are presented in ►Table 4. ►Figure 4 provides some examples of the different extension types in the SNOMED CT hierarchy. The extensions were all added to the extracted SNOMED CT core data tables. The subset was extended with 325 concepts, 1243 relationships for these concepts, and 597 descriptions.

3.5 Constraining the Subset

It is hard to decide to which level the depth of the graph should be restricted. Use of SNOMED CT in daily care practice requires high granularity but if the depth of the graph is not restricted and only the non-human concepts in SNOMED CT were excluded from the subset this would result in 233,782 concepts, 660,008 descriptions and 665,394 relationships in the database. This can hardly be named a SNOMED CT subset. The use of the SNOMED CT subset in real practice should reveal to which level of depth the SNOMED CT graph should be restricted and whether the SNOMED CT subset needs further constraining. Therefore, this phase of the interface terminology development will be repeated after initial evaluation in an intensive care setting.

3.6 Terminology Server

The interface terminology was deployed using the local terminology server DICE (Diagnoses for Intensive Care Evaluation). DICE consists of a SOAP-based Java terminology service together with a client for knowledge browsing and post-coordination [38]. The DICE client can be used to add controlled compositional descriptions to clinical records. The implementation of DICE offers physicians three ways to search for the appropriate reason for admission: a) a short list containing the most frequently occurring reasons for admission in an ICU, b) the APACHE IV classification

entry, and c) entry of (a part of) its preferred or synonymous term. The system returns all terms matching the given free-text query. The user can then select anyone of the returned terms, after which, a list of subordinate concepts is shown, if these exist.

Once a concept is selected, DICE enables post-coordination to provide concepts with more detailed information. For example, Sepsis can be further qualified by finding site, causative agent and the underlying diagnosis, but such further specification is not mandatory. In case a concept can not be found in the interface terminology it is also possible to enter the diagnostic information in free text. Finally, users can also provide comments on each entry.

4. Discussion

The use of all of SNOMED CT as an interface terminology in a specific clinical setting is the subject of discussion in many studies [8, 12, 13]. It has been argued that using SNOMED CT for a particular specialist application requires the creation of an interface terminology based on a domain-specific subset [15]. Nevertheless, very few studies focused on the process of designing such an interface terminology on SNOMED as a whole. In this study we contribute to this issue by enumerating and combining the different phases in the processes of identifying an interface terminology on SNOMED CT. As far as we know, this study is the only work that provides an extensive overview of the total process and can be used as a reference by other researchers who also intend to develop a domain-specific interface terminology on SNOMED CT.

Due to the lack of a Dutch SNOMED CT translation using historical Dutch data as a starting point for the subset is infeasible. Furthermore as in this case study one of the important use cases is to calculate mortality risks we used the APACHE IV classification as a starting point to identify the intensive-care-specific SNOMED CT subset. The APACHE IV classification is the most extensive and widely used classification for the intensive care.

Use of the interface terminology by its end users should reveal whether the

SNOMED CT subset based on the APACHE IV classification is sufficient for collection of diagnostic information for multiple use, including daily care processes. It should also shed light on the optimal size of the interface terminology, i.e., whether a full extension is needed and useful, or whether a limited extension better enables the collection of diagnostic information. Eventually, other resources might be required to optimize our interface terminology content [22]. The free text entries and comments provided by the users in DICE, for instance, might provide a common ground for this purpose.

To identify the equivalent concepts for the APACHE IV categories in SNOMED CT, a mapping was realized. Previous studies on mapping aggregate terminologies to SNOMED CT have shown that the mapping can be influenced by the structure and the content characteristics of both systems [4, 6, 25]. In the APACHE IV classification, for instance, some categories are provided with classification rules, e.g. “Respiratory Arrest” is provided with the additional rule “without cardiac arrest”. While these kinds of rules are used in aggregate terminologies to make clear what should and what should not belong to a class, they are not included in terminological systems such as SNOMED CT, which are used to document clinical data [21]. Consequently, these rules are also not accounted for in the mapping. Furthermore, the diagnostic categories including the word “other” (e.g. “other cardiovascular disorder” or “other respiratory disorder”) are residual categories for conditions that cannot be allocated to more specific categories. These categories formed the largest part of the “partial match” group. Again, while these kinds of categories are used in aggregate terminologies to classify data, they are not included in formally defined reference terminological systems such as SNOMED CT [21]. These diagnostic categories were all matched to a superordinate concept in SNOMED CT and were included as a pre-coordinated concept in the extended subset.

An intensivist was consulted to interpret ambiguous APACHE IV categories. As this interpretation may be subjective, a duplicate interpretation by two independent do-

main experts is advised, with a conflict resolution process similar to that used for the mapping.

Post-coordination has been proposed and demonstrated as an approach to improve the content coverage (12). Also in this study, the use of post-coordination resulted in better matching scores as 34.9 % of all complete matches and 31.4 % of all partial matches were realized through post-coordination. The use of post-coordination is thus an important factor to bear in mind when considering SNOMED CT for data collection.

The results of the mapping procedure were used to create the relevant SNOMED CT subset. One of the advantages of the subsetting is to isolate the users from the complexity of SNOMED CT and to extend the subset with user-required content which was not included in SNOMED CT [12]. Depending on the restriction of the depth of the graph traversal the numbers of disorder and procedure concepts range from 5570 (i.e., 2% of all active SNOMED CT concepts) to 83,125 (29% of all active SNOMED CT concepts). In addition, when fully expanding the subset, more than 150,000 concepts can be added as attribute values to further specify the disorder and procedure concepts, resulting in a subset of 233,782 concepts (82% of all active SNOMED CT concepts). A previous study showed that a subset of about 2700 concepts from SNOMED CT was sufficient to cover 96% of clinical notes for patients admitted to the intensive care over a period of five years [16]. Even under the condition that the study was performed in another setting with a different basic assumption, it indicates that it is likely that restricting the level of the graph is necessary. However, the domain of intensive care is rather complex and broad, involving a diversity of clinical problems. A simple appendicitis, for instance, may lead to severe sepsis with a wide range of possible underlying microbiological agents. So, for the interface terminology to facilitate sharing and aggregating this kind of data for different purposes, it should enable a detailed registration of widely diverging clinical problems and their attributes. An evaluation of the interface terminology by its end users should reveal whether shrinking is necessary and, if so, what parts could be reduced.

The advantage of creating an interface terminology is among others the possibility to include concepts, relationships, and terms which are not present in the reference terminology, but which are required by the user. Therefore, as diagnostic categories such as “head trauma with chest trauma” occur frequently in the ICU and are part of the APACHE IV classification, these should be included as pre-coordinated concepts in an intensive-care-specific interface terminology. Yet, because SNOMED CT does not support the creation of collections (i.e., defining A “with” B), it is hard to represent these composite diagnostic categories correctly [39]. Instead, in SNOMED CT, collections are usually defined using the logical “And”. For instance, the concept “195878008|Pneumonia and influenza” is modeled as a taxonomic child of “6142004|Influenza” AND of “75570004|Viral pneumonia”, referring to situations in which a patient has both influenza and viral pneumonia. Although not strictly logically correct, in line with the SNOMED CT modeling, within our study we also represented the composite diagnostic categories such as “head trauma with chest trauma” as a taxonomic child of two or more other concepts. As argued before, further research is needed to gain more insight into the consequences this representation will have in practice, mainly for the purposes of querying patient data [39].

A major issue in using the (extended) subset in an interface terminology in practice is its maintenance. A new version of SNOMED CT is published every six months and may introduce (large) changes in the terminology content [40–42]. Also the SNOMED CT rules regarding for instance post-coordination might change [42]. These changes can affect the local extensions and consequently the consistency of the captured data. Therefore, in order to retain its utility, the interface terminology needs to be updated on a regular basis and an update mechanism, as for instance described in the SNOMED CT Reference Set Specification, should be in place to organize this process [40, 43, 44].

The real test of our interface terminology will come when it is used in practice in a computer-based patient record. Although a lot of studies have focused on the content

coverage of medical terminologies, only a few examined how clinicians interact with the terminological system during data entry [45, 46]. In the next phase of our project, we will evaluate the interface terminology in an intensive care setting in an electronic health record. The evaluation will not only focus on the terminology content, but also on the user interface of the terminology client [3]. Especially the effect that the size and comprehensiveness of the content have, and ways in which users can be adequately supported to handle these need further study.

5. Conclusion

To our knowledge, we are the first to report on the process of developing a domain-specific interface terminology on SNOMED CT. The application of the proposed approach to the domain of intensive care resulted in an interface terminology which can be used to facilitate sharing and aggregation of data for different purposes. Future work should reveal whether the method is also applicable for other domains.

References

1. de Keizer NF, Abu-Hanna A, Zwetsloot-Schonk JH. Understanding terminological systems. I: Terminology and typology. *Methods Inf Med* 2000; 39 (1): 16–21.
2. German E, Leibowitz A, Shahar Y. An architecture for linking medical decision-support applications to clinical databases and its evaluation. *J Biomed Inform* 2009; 42 (2): 203–218.
3. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. A Model for Evaluating Interface Terminologies. *J Am Med Inform Assoc* 2008; 15 (1): 65–76.
4. Brown SH, Husser CS, Wahner-Roedler D, Bailey S, Nugent L, Porter K, Bauer BA, Elkin PL. Using SNOMED CT as a reference terminology to cross map two highly pre-coordinated classification systems. *Medinfo* 2007; 12 (Pt 1): 636–639.
5. World Health Organization. International classification of diseases, manual of the international statistical classification of diseases, injuries and causes of death. 9th revision. 1977.
6. Imel M, Campbell JR. Mapping from a Clinical Terminology to a Classification, AHIMA's 75th Anniversary National Convention and Exhibit Proceedings, 2003.
7. Young JD, Goldfrad C, Rowan K. Development and testing of a hierarchical method to code the reason

- for admission to intensive care units: the ICNARC Coding Method. Intensive Care National Audit & Research Centre. *Br J Anaesth* 2001; 87 (4): 543–548.
8. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. *J Am Med Inform Assoc* 2006; 13 (3): 277–288.
 9. Spackman KA, Campbell KE, Cote RA. SNOMED RT: a reference terminology for health care. *Proc AMIA Annu Fall Symp* 1997. pp 640–644.
 10. Richesson RL, Andrews JE, Krischer JP. Use of SNOMED CT to Represent Clinical Research Data: A Semantic Characterization of Data Items on Case Report Forms in Vasculitis Research. *J Am Med Inform Assoc* 2006; 13 (5): 536–546.
 11. Elevitch FR. SNOMED CT: electronic health record enhances anesthesia patient safety. *AANA J* 2005; 73 (5): 361–366.
 12. Brown SH, Elkin PL, Bauer BA, Wahner-Roedler D, Husser CS, Temesgen Z, Hardenbrook SP, Fielstein EM, Rosenbloom ST. SNOMED CT: utility for a general medical evaluation template. *AMIA Annu Symp Proc* 2006. pp 101–105.
 13. Campbell JR. Semantic features of an enterprise interface terminology for SNOMED RT. *Medinfo* 2001; 10 (Pt 1): 82–85.
 14. Asbeh N, Peleg M, Schertz M. Creating Consistent Diagnoses List for Developmental Disorders Using UMLS, Next generation information technologies and systems. Lecture notes in computer science. Berlin/Heidelberg: Springer; 2006 (4032). pp 333–336.
 15. Lopez-Osornio A, Luna D, Gambarte ML, Gomez A, Reynoso G, de Quiros FG. Creation of a local interface terminology to SNOMED CT. *Medinfo* 2007; 12 (Pt 1): 765–769.
 16. Patrick J, Wang Y, Budd P, Rector A, Brandt S, Rogers BJ, Herkes R, Ryan A, Vazirnezhad B. Developing SNOMED CT Subsets from Clinical Notes for Intensive Care Service. 7th Annual HINZ Conference and Exhibitions, Rotorua, New Zealand, 2008.
 17. Seidenberg J, Rector A. Web ontology segmentation: Analysis, classification and use. In: 5th International World Wide Web Conference 2006. Available at: <http://www.co-ode.org/resources/papers/seidenberg-www2006.pdf>. Last accessed: September 2009.
 18. The International Health Terminology Standards Development Organisation. Available at: <http://www.ihtsdo.org/our-standards/snomed-ct/>. Last accessed: September 2009.
 19. SNOMED Clinical Terms® Technical Implementation Guide. The International Health Terminology Standards Development Organisation. January 2008.
 20. SNOMED Clinical Terms® Technical Reference Guide. The International Health Terminology Standards Development Organisation. January 2008.
 21. SNOMED Clinical Terms Editorial Guidelines, Content Inclusion Principles and Process, The International Health Terminology Standards Development Organisation. May 2008.
 22. Castro AG, Rocca-Serra P, Stevens R, Taylor C, Nashar K, Ragan MA, Sansone SA. The use of concept maps during knowledge elicitation in ontology development processes – the nutrigenomics use case. *BMC Bioinformatics* 2006; 7: 267.
 23. Cimino JJ. From data to knowledge through concept-oriented terminologies: experience with the Medical Entities Dictionary. *J Am Med Inform Assoc* 2000; 7 (3): 288–297.
 24. Ryan A, Patrick J, Herkes R. Introduction of enhancement technologies into the intensive care service. Royal Prince Alfred Hospital, Sydney. *HIM J* 2008; 37 (1): 40–45.
 25. Vikstrom A, Skaner Y, Strender LE, Nilsson GH. Mapping the categories of the Swedish primary health care version of ICD-10 to SNOMED CT concepts: rule development and intercoder reliability in a mapping trial. *BMC Med Inform Decis Mak* 2007; 7: 9.
 26. Min Z, Baofen D, Weeber M, van Ginneken AM. Mapping OpenSDE domain models to SNOMED CT. Applied to the domain of cardiovascular disease. *Methods Inf Med* 2006; 45 (1): 4–9.
 27. Wang Y, Patrick J, Miller G, O'Halloran J. Linguistic mapping of Terminologies to SNOMED CT. First European Conference on SNOMED CT organised by the Network of Excellence Semantic Mining. 2006. Available at: <http://www.hiwww.org/smcs2006/proceedings/6WangSMCS2006final.pdf>. Last accessed: September 2009.
 28. Wilson PS. What mapping and modeling means to the HIM professional. *Perspect Health Inf Manag* 2007; 4: 2.
 29. Rosenbloom ST, Brown SH, Froehling D, Bauer BA, Wahner-Roedler DL, Gregg WM, Elkin PL. Using SNOMED CT to represent two interface terminologies. *J Am Med Inform Assoc* 2009; 16 (1): 81–88.
 30. Wang Y, Patrick J, Miller G, O'Halloran J. A computational linguistics motivated mapping of ICPC-2 PLUS to SNOMED CT. *BMC Med Inform Decis Mak* 2008; 8 (Suppl 1): S5.
 31. Noy NF, Musen MA. Specifying ontology views by traversal. In: International Semantic Web Conference 2004. pp 713–725. Available at: <http://www.stanford.edu/~natalya/papers/traversalViews-SW.pdf>. Last accessed: September 2009.
 32. Wasserman H, Wang J. An applied evaluation of SNOMED CT as a clinical vocabulary for the computerized diagnosis and problem list. *AMIA Annu Symp Proc* 2003. pp 699–703.
 33. Chute CG, Elkin PL, Sherertz DD, Tuttle MS. Considerata for a clinical terminology server. *Proc AMIA Symp* 1999. pp 42–46.
 34. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006; 34 (5): 1297–1310.
 35. DesHarnais SI, Forthman MT, Homa-Lowry JM, Wooster LD. Risk-adjusted quality outcome measures: indexes for benchmarking rates of mortality, complications, and readmissions. *Qual Manag Health Care* 1997; 5 (2): 80–87.
 36. Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 2004; 363 (9415): 1147–1154.
 37. SNOMED Clinical Terms® User Guide. The International Health Terminology Standards Development Organisation. January 2008.
 38. Cornet R, Prins AK. An architecture for standardized terminology services by wrapping and integration of existing applications. *AMIA Annu Symp Proc* 2003. pp 180–184.
 39. Cornet R, Schulz S. Relationship Groups in SNOMED CT. *Stud Health Technol Inform* 2009; 150: 223–227.
 40. Mapoles J, Smith CS, Brian Levy JC. Strategies for Updating Terminology Mappings and Subsets using SNOMED CT®. Proceedings of the 3rd international conference on Knowledge Representation in Medicine (KR-MED): Representing and sharing knowledge using SNOMED. 2008.
 41. Spackman KA. Rates of change in a large clinical terminology: three years experience with SNOMED Clinical Terms. *AMIA Annu Symp Proc* 2005. pp 714–718.
 42. Wade G, Rosenbloom ST. The impact of SNOMED CT revisions on a mapped interface terminology: Terminology development and implementation issues. *J Biomed Inform* 2009.
 43. SNOMED Clinical Terms® Reference Set Specification. International Healthcare Terminology Standards Development Organisation. January 2008.
 44. Bakhshi-Raiez F, Cornet R, de Keizer NF. Development and application of a framework for maintenance of medical terminological systems. *J Am Med Inform Assoc* 2008; 15 (5): 687–700.
 45. Cimino JJ, Patel VL, Kushniruk AW. Studying the human-computer-terminology interface. *J Am Med Inform Assoc* 2001; 8 (2): 163–173.
 46. Kushniruk A, Patel V, Cimino JJ, Barrows RA. Cognitive evaluation of the user interface and vocabulary of an outpatient information system. *Proc AMIA Annu Fall Symp* 1996. pp 22–26.