
**Coherent SNOMED CT
Implementation Facilitating
Re-use of Data**

Coherent SNOMED CT Implementation Facilitating Re-use of Data

PhD Thesis by

Anne Randorff Højen

*Medical Informatics Group,
Department of Health Science and Technology,
Aalborg University, Denmark*


River Publishers
Aalborg

ISBN 978-87-93102-83-5 (e-book)

Published, sold and distributed by:

River Publishers
Niels Jernes Vej 10
9220 Aalborg Ø
Denmark

Tel.: +45369953197
www.riverpublishers.com

Copyright for this work belongs to the author, River Publishers have the sole right to distribute this work commercially.

All rights reserved © 2014 Anne Randorff Højen.

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without prior written permission from the Publisher.

Acknowledgements

First and foremost, I would like to express my gratitude to all the people who made it possible for me to work on this very interesting project. To Trifork A/S who contributed financially and with initial discussions.

I am very grateful to my main supervisor, Stig Kjær Andersen, for his positive attitude and ability to guide me in the right direction and to discuss how best to disseminate the results of my research. I am also very grateful to my other supervisor, Pia Britt Elberg, who has helped me immensely throughout the whole period. She was always available to discuss difficult problems and guiding me in the writing of papers.

I would like to thank Kirstine Rosenbeck for our close collaboration and for her help in answering study-related questions and propose possible solutions. She is a great colleague and supported me in conducting a valuable study. I would also like to thank Erik Sundvall, who contributed with development of the visualisation tool, and the people from the North Jutland Region. I enjoyed our collaboration and interesting discussions.

My thanks go to all my great colleagues at Aalborg University, who were always there to help me when I needed them.

Finally, I would like to thank my family, friends and especially my wonderful husband, for supporting me throughout the entire process and for keeping family life a joy even in peak periods.

List of publications

The main results of this PhD work is published in four papers enclosed as appendices of this thesis.

Paper A

Rasmussen, A. R., & Rosenbeck, K. (2011). SNOMED CT implementation: implications of choosing clinical findings or observable entities. *Studies in Health Technology and Informatics*, 169, 809-13.

Paper B

Højen, A. R., & Gøeg, K. R. (2011). SNOMED CT Implementation. Mapping guidelines facilitating reuse of data. *Methods of Information in Medicine*, 50(5), 472-478.

Paper C

Højen, A. R., Gøeg, K. R., Sundvall, E. (2013). Methods and applications for visualization of SNOMED CT concept sets. *Applied Clinical Informatics*. In review process.

Paper D

Højen, A. R., Elberg, P.B., Andersen, S.K.. (2013). SNOMED CT in Denmark. Why is it so hard? Submitted for the Medical Informatics Europe (MIE) 2014 conference.

Additional scientific contributions of the author related to the above-mentioned work follows:

Conference papers

Dueholm, S. S., Rasmussen, J. H., Poulsen, R. S., Boysen, M. A., Rasmussen, A. R., & Johansen, M. D. (2011). Anatomy note-taking software supporting different learning modalities. *ICERI2011 Proceedings*, p. 6571-6576.

Rosenbeck, K. H., Rasmussen, A. R., Elberg, P. B., & Andersen, S. K. (2010). Balancing centralised and decentralised EHR approaches to manage standardisation. *Studies in health technology and informatics*, 160 (Pt 1), 151.

Rosenbeck, K. H., & Rasmussen, A. R. (2010). Does mapping to SNOMED CT improve precision of subjective clinical evaluations?. In Scandinavian Conference on Health Informatics. Tapir Akademisk Forlag 2010, p. 39-43

Pape-Haugaard, L., Rasmussen, A. R., Rosenbeck, K. H., & Andersen, S. K. (2010). Pitfalls when integrating terminology systems and EHRs. In Scandinavian Conference on Health Informatics. Tapir Akademisk Forlag 2010, p. 34-38.

Rasmussen, A. R., & Markussen, D. (2009, August). Methods for Testing SNOMED CT in a Defined Clinical Domain. In Studies in Health Technology and Informatics, Vol. 150, 2009, p. 248.

Abstracts

Højen, A. R., Gøeg, K. R., & Elberg, P. B. (2012) Applying visualization to support coherent terminology implementation. In Scandinavian Conference on Health Informatics, SHI, Linköping University Electronic Press, 2012. p. 49 (Linköping Electronic Conference Proceedings; No. 70).

Jensen, A M. D., Jensen, M. M., Korsager, A. S., Rasmussen, A. R., Elberg, P. B. Using the SNOMED CT-model for standardization of clinical information in labour documentation. (2010). In Scandinavian Conference on Health Informatics, SHI2010 Proceedings: p. 62.

Pape-Haugaard, L., Rasmussen, A. R., Elberg, P. B. , & Andersen, S. K. (2010). Conflicts between terminology and EHR information models as obstacles to semantic interoperability : a scientific review.
In: Studies in Health Technology and Informatics, Vol. 160, 2010, p. 1374.

Abstract

The objective of this research is to develop methods and tools to enable clinicians to register data of high quality that are useful for supporting clinical practice and at the same time serve as basis for management, reimbursement, quality assessment and research. Reuse of data within and across Clinical Information Systems (CIS) can be realized by use of the international standardized clinical terminology SNOMED CT. The prerequisite for reuse of data from different patient contacts is that the terms and concepts entered by clinicians are well-defined, structured and can be processed automatically by computers. SNOMED CT can support automatic terminological reasoning but it requires that the CISs are developed and configured to enter and store SNOMED CT concepts and expressions. The size and complexity of SNOMED CT challenge consistent use of concepts across organizational borders, hampering comparability and hence reuse of data. Thus, methods that support consistent concept selection is important for SNOMED CT-implementation projects, to utilize the structure of SNOMED CT to ensure comparability of clinical data. This thesis presents the theoretical foundation for comparability of data along with four studies addressing SNOMED CT implementation and reusability of clinical data.

The first study presents an analysis of how concepts from two different SNOMED CT hierarchies can be used to represent clinical expressions in two data entry templates. This study shows that, besides from consistent concept selection, it is important to be aware of the definition and hierarchical composition of the concepts to select concepts with the best comparability and retrieval-properties.

To support consistent and semantically coherent concept selection the second study presents the development of a set of SNOMED CT mapping guidelines. These guidelines are based on an iterative mapping process conducted on clinical content from 14 different EHR-templates from five different Danish and Swedish EHR-systems. Each mapping is assessed against overall quality criteria enabling specification and refinement of the guidelines.

The third study documents the development and evaluation of a web-based visualization tool, SNOVIEW, which graphically visualize sets of SNOMED CT concepts, their hierarchical relationships and a set of terminological features. The fourth study documents the results of exploring reasons for lacking adaption of SNOMED CT in Denmark through group work and focus groups interviews with representatives from three central groups of stakeholders related to SNOMED CT implementation. The focus group discussions were focused on the participants' experiences and expectations related to terminology management.

SNOMED CT implementation that facilitates comparability and reuse of data requires knowledge about the intended use of data and understanding of the logical model of SNOMED CT. Further research should demonstrate the value of SNOMED CT in actual clinical settings, and demonstrate how to incorporate the logical representation of SNOMED CT concepts in existing CISs to enable automatic retrieval and terminological reasoning of both pre- and post-coordinated concepts.

Abstract in Danish

Målet med denne forskning er at udvikle metoder og værktøjer, der kan gøre det muligt for klinikerne at registrere data af høj kvalitet på en måde så data kan bruges til at informere andre klinikere og samtidig kan danne grundlag for ledelsesinformation, afregning, kvalitetssikring og forskning. Genbrug af kliniske data i og på tværs af kliniske informationssystemer kan realiseres ved hjælp af en internationalt standardiseret sundhedsterminologi som SNOMED CT. Forudsætningen for genbrug af data fra forskellige patientkontakter er at de termer og begreber klinikerne anvender, er veldefinerede, strukturerede, og kan behandles automatisk af computere. Dette muliggøres igennem SNOMED CT, men kræver, at de kliniske IT-systemer udvikles og konfigureres til at kunne repræsentere de kliniske data vha. terminologien. Pga. SNOMED CT's størrelse og kompleksitet er det imidlertid vanskeligt at sikre konsistens i valg af begreber på tværs af organisatoriske enheder, hvilket er en afgørende hindring for at sikre sammenlignelighed af kliniske data og dermed genbrug af data. Derfor er det vigtigt med metoder til at støtte SNOMED CT-implementering og udnytte strukturen af SNOMED CT således sammenlignelighed af kliniske data sikres. Afhandlingen præsenterer den teoretiske baggrund for sammenlignelighed af data samt de fire studier der er gennemført.

I det første studie analyseres og sammenlignes begreber fra to hierarkier i forhold til at kunne repræsentere det kliniske indhold fra to kliniske inddateringskemaer. Dette studie viser, at det udover konsistent brug af SNOMED CT, forudsætter viden om begrebernes definition og hierarkiske sammenhæng, at kunne udvælge de begreber fra der sikrer bedst mulige forudsætninger for at kunne genanvende og sammenligne data.

Til at støtte konsistent og semantisk sammenhængende udvælgelse af begreber fra SNOMED CT er der i det andet studie udviklet et sæt retningslinjer for udvælgelse af SNOMED CT-begreber til brug for system-konfiguratører. Retningslinjerne er udviklet på baggrund af en analyse af indholdet i 14 forskellige EPJ-skabeloner hentet fra fem danske og svenske EPJ-systemer som i en iterativ proces blev mappet til SNOMED CT. Hver mapning er vurderet forhold til et overordnet kvalitetskriterium, som har ført til specificering af retningslinjerne. I det tredje studie er der udviklet et visualiseringsværktøj, SNOVIEW, som grafisk præsenterer sæt af SNOMED CT-begreber, deres hierarkiske relationer og desuden et sæt af terminologiske funktioner. En kvalitativ evaluering af SNOVIEWs anvendelsesmuligheder i det fjerde studie indikerer at visning af en hierarkisk sammenhæng imellem sæt af SNOMED CT begreber har potentiale til at understøtte fx intelligent søgning i kliniske informationssystemer og uddannelse i samt håndtering af terminologi.

Implementering af SNOMED CT kræver viden om datas potentielle anvendelse og om SNOMED CT's struktur og logiske opbygning, hvis SNOMED CT skal kunne understøtte sammenlignelighed af data på tværs af systemer. Fremtidig forskning bør demonstrere hvilken værdi implementering af SNOMED CT har i praksis, herunder hvordan man udnytter den logiske repræsentation af begreber fra SNOMED CT i eksisterende kliniske informationssystemer, således automatisk udtræk af både præ- og postkoordinerede begreber muliggøres.

Table of content

Abstract	10
Abstract in Danish	11
Table of content	13
Introduction: Implementation of Clinical Terminologies - Potential and challenges	15
<i>The role of clinical terminologies</i>	15
<i>Terminological challenges in configuring Clinical Information Systems</i>	20
<i>Research Objective</i>	24
Method	25
<i>Method for analyzing theory and practice of semantic comparability using SNOMED CT</i>	26
<i>Method for development of tools supporting semantic comparability when selecting SCT concepts</i>	32
Results (part 1)	35
Theoretical foundation for semantic comparability using SNOMED CT	37
<i>Logical foundation of SNOMED CT</i>	37
<i>Hierarchical comparison</i>	38
<i>Meaning-based comparison</i>	42
Empirical study of semantic comparability using SNOMED CT	51
<i>Implications for people configuring CISs</i>	51
<i>Implications of technological incompatibility</i>	55
<i>Implications of local and central Governance</i>	56
Summary of findings of the theoretical and empirical analysis	58
Results (part 2)	61
Mapping guidelines	63
<i>Quality criteria for concept selection</i>	63
<i>Material and method</i>	64
<i>Hierarchy selection</i>	66
<i>Concept selection</i>	67
Visualizing sets of concepts	69
<i>Tool for visualizing sets of concepts</i>	69
<i>Potential of visualization of sets of concepts</i>	71
Summary of development of tools supporting semantic comparability when selecting SCT concepts	75

Discussion	77
<i>Terminology binding</i>	77
<i>Observable entities vs. clinical findings</i>	79
<i>Post-coordination aspects</i>	81
<i>The difficulty of SNOMED CT evaluations</i>	83
Conclusion	84
<i>Towards consistent and coherent SNOMED CT adoption</i>	84
References	86
Paper A	91
Paper D	93
Paper B	95
Paper C	97

Introduction: Implementation of Clinical Terminologies - Potential and challenges

The role of clinical terminologies

Clinical terminologies are highlighted in the scientific literature as a key factor for improving communication of clinical data and increase availability of relevant information for the various stakeholders within the health sector. Following sections presents potential and challenges related to implementation of clinical terminologies.

Reuse of clinical data

Clinical terminologies have potential to support development and configuration of Clinical Information Systems (CISs) that enable semantic interoperability (Lopez, Blobel 2009) and support efficient and effective data entry and retrieval. Various definitions of Semantic interoperability in health exist. (Dolin, Alschuler 2011, Stroetmann et al. 2009, European Commission 2008, Elkin et al. 2010, Cimino 2011, Cimino 2007, Cimino 1998, Aguilar 2005) Examples are:

“semantic interoperability means ensuring that the precise meaning of exchanged information is understandable by any other system or application not initially developed for this purpose.”

(European Commission 2008)

“Semantic interoperability, which defines the ability for information shared by systems to be understood at the level of formally defined domain concepts so that the information is computer processable by the receiving systems”

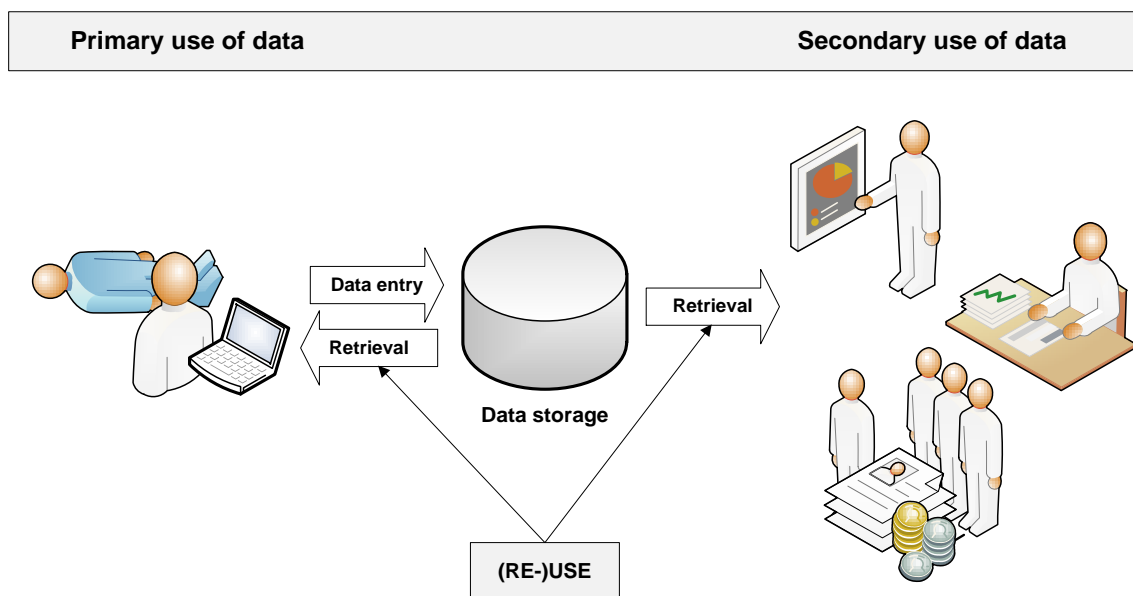
(Aguilar 2005)

The first definition focuses on the meaning-preserving aspect and the second definition focuses on understanding at the level of formally defined domain concepts. Hence, a prerequisite in achieving Semantic interoperability is standardized concept systems (like SNOMED CT) because they are the mean to representing clinical meaning unambiguously.

CISs are configured to transform the clinicians' documentation needs and requirements for functionality into templates which best support the clinical practice. In addition to clinicians' requests, the national and regional authorities, administration and management also want to reuse the clinical data for secondary purposes, such as

overview of clinical activities, expenditures within a specific clinical department, quality measures, waiting time statistics, reimbursement etc. Clinical data captured with the primary purpose to support decisions on patient care and treatment must therefore also serve as basis for secondary purposes. (Cimino 2007, Elkin et al. 2010, Barton et al. 2011) Hence, system configuration enabling reuse of data requires bridging the flexibility necessary for documenting the fine-grained clinical information with the need for coarse-grained information necessary for analytics. Figure 1 illustrates how data entry is done in a clinical care setting and retrieval can be done for either primary or secondary purposes.

Figure 1 Overall illustration of data reuse. Data entry is done in a specific clinical situation in a specific CIS. Whenever data is being retrieved, either for primary or secondary purposes, we define it as reuse of data.



In (Cimino 2007) it is explained how clinical terminology used to capture a specific clinical use case can be reused for various purposes, such as computerized clinical guidelines, alert systems, automatic retrieval etc. In this study, the definition of data reuse is when information captured in a specific clinical situation for clinical purposes is retrieved from its original storage and applied for either primary or secondary purposes.

Both in a semantic interoperability context and in bridging fine-grained and coarse-grained information, clinical terminologies are essential. The next section introduces the terminology of focus in this project, SNOMED CT.

SNOMED CT: Great potential, limited implementation experience

IHTSDO¹ distributes and develops the clinical terminology SNOMED CT which is regarded as one of the major clinical terminologies in the world. (Cimino 1998) proposed a set of desiderata's, which specified requirements for coding systems to be used for clinical documentation. (Cimino 1998) These requirements have been used to guide the development of SNOMED CT and include:

1. Vocabulary content
2. Concept orientation
3. Concept permanence
4. Non-semantic concept Identifiers
5. Poly-hierarchy
6. Formal definitions
7. Rejection of "not elsewhere classified" terms
8. Multiple granularities
9. Multiple consistent views
10. Context representation
11. Graceful evolution
12. Recognized redundancy

The main objective of SNOMED CT is to contribute to the improvement of the clinical documentation by underpinning development of CISs that support meaning-based data entry and retrieval. (IHTSDO) Currently, SNOMED CT is being adopted (Wade, Rosenbloom 2008) as one of the most comprehensive clinical terminologies (Cornet 2009). SNOMED CT concepts represent clinical meanings defined by formal description logic rules allowing terminological reasoning with the purpose to support common interpretation, sharing and reusability of clinical data. (IHTSDO)

Through various studies SNOMED CT has proven to meet clinical documentation needs with respect do content coverage (So, Park 2011, Wade, Rosenbloom 2008, Wasserman, Wang 2003, Rosenbloom et al. 2009). Leiberman and Ricciardi (2003) demonstrated how the use of SNOMED CT can simplify development and maintenance of clinical querying without semantically reducing the query results. Additionally (Bodenreider 2008) describe the potential of biomedical ontologies to support linkage to enable decision support, data aggregation and data selection.

Despite these potential benefits and proven coverage of SNOMED CT across a range of clinical domains, actual implementation and use of SNOMED CT in clinical settings is limited (Lee et al. 2013, Cornet, de Keizer 2008, Lee et al. 2012). A literature review on the use of SNOMED CT shows that a large number of scientific studies

¹ The International Health Terminology Standards Development Organization (www.ihtsdo.org)

address SNOMED CT. Lee et al. categorize 488 papers into four groups and show that despite increased attention on SNOMED CT implementation the majority of studies focus on theoretical and predevelopment/design which is the first step towards adoption. Currently, limited documentation about actual SNOMED CT implementations exist (Lee et al. 2013). The study concludes that more effort is needed to bring SNOMED CT into routine clinical use and to demonstrate utility and applicability of the poly-hierarchical terminology.

Mono-hierarchies vs. poly-hierarchies

In (Rector 1999) A. Rector provides the following definition of clinical terminologies and terminological reasoning:

“Clinical terminology concerns the meaning, expression, and use of concepts in statements in the medical record or other clinical information system.”

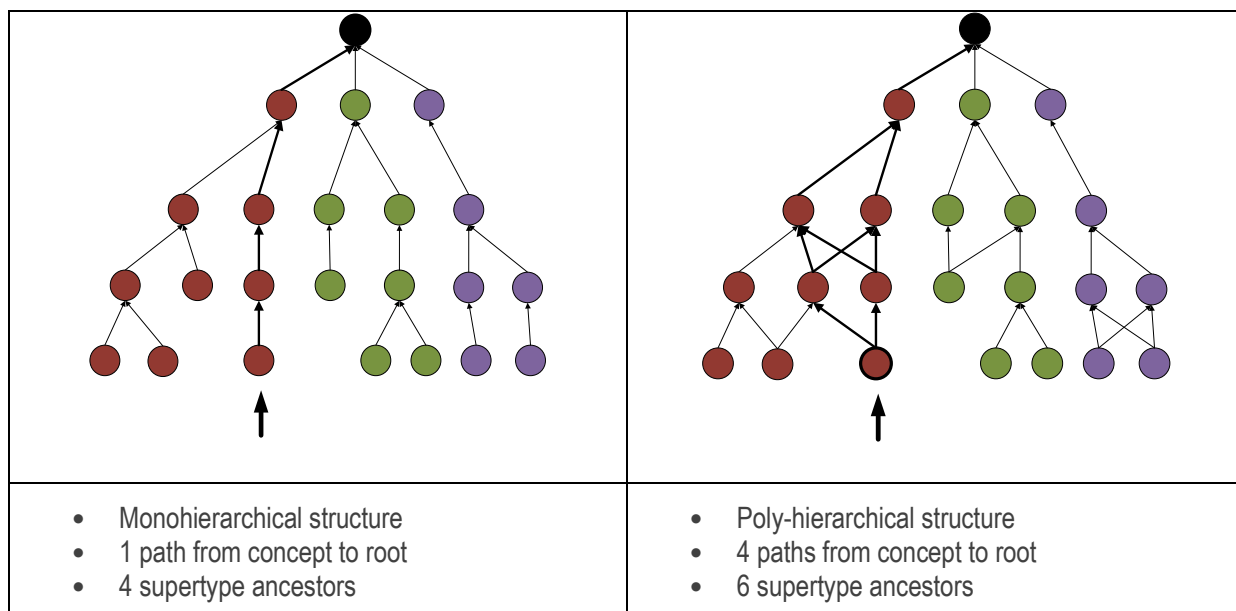
“Terminological reasoning is that reasoning which can be performed on the basis of the classification, relations and comparison of isolated concepts from a medical record or information system.”

Considering the nature of the underlying formal logic and the designation of concepts to represent unique clinical meanings, many types of clinical terminologies can be described, as nomenclatures, thesauruses, taxonomies, etc.(de Keizer, Abu-Hanna & Zwetsloot-Schonk 2000) Ingenerf and Giere (1998) divided medical terminologies into two categories based on purpose:

- **Statistical classification systems**, such as ICD-10 ICPC and NCSP, are mainly developed to support secondary purposes, such as aggregation and statistics.
- **Concept systems**, such as SNOMED CT, are developed to represent all clinically relevant information and have greater flexibility and granular representation of clinical information. (Ingenerf, Giere 1998, Rector 1999) Also, concept systems can represent defining characteristics for clinical information using relationships to other concepts.(Ingenerf, Giere 1998)

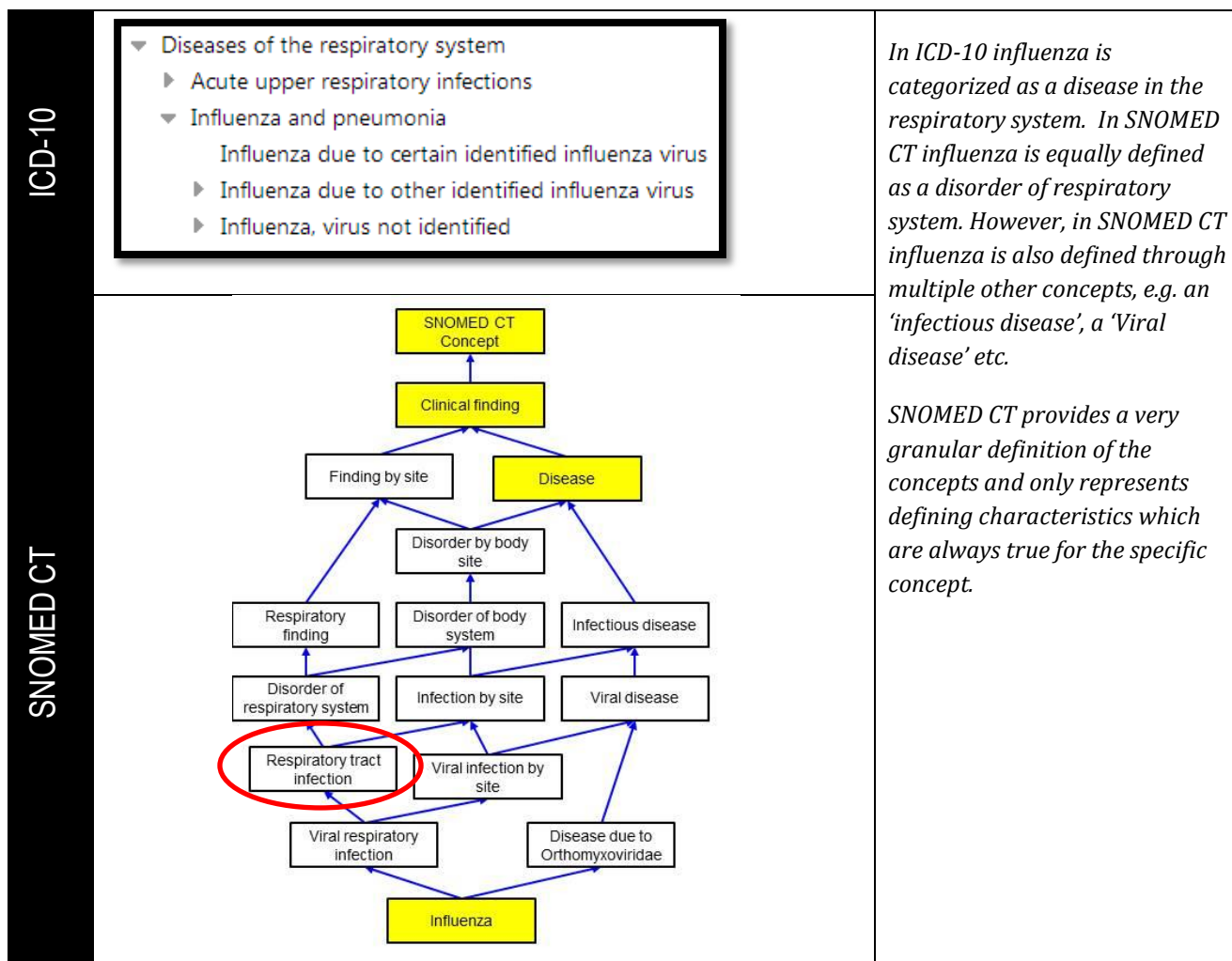
A main difference between statistical classifications and concept systems is that classifications are mono-hierarchical whereas concept systems are polyhierarchal. This means that concepts can have more than one supertype or parent concepts. This makes the subtype hierarchies in SNOMED CT direct acyclic graphs (DAGS). Mono-hierarchical classification systems only allow for one supertype parent per concept, which results in only one path from a concept to the root concept. In contrast, the complexity of traversing DAGs is much more complex. Figure 2 illustrates the difference between mono- and poly-hierarchies and shows that poly-hierarchies allow for a larger number of supertype ancestors and therefore more complex graphs.

Figure 2 Comparing monohierarchies and poly-hierarchies. Poly-hierarchies will lead to an increased graph complexity due to the multiple paths between to concepts and because of an increased amount of supertypes.



Technologically, the increased complexity of the poly-hierarchical structure, can be challenging, however clinically, the poly-hierarchical structure supports accurate definition of the meaning of a concept. For example, in ICD-10, the disease 'influenza' is categorized as a 'Disease of the respiratory system'. This definition is correct, however it is not sufficient, as 'influenza' can also be defined as a viral disease. The poly-hierarchical structure of SNOMED CT allows for this more fully defined representation, as illustrated in Figure 3.

Figure 3 ICD-10 vs. SNOMED CT comparison.



For ICD-10, with the specific purpose being statistical classification, the monohierarchical structure serves its purpose. However, the poly-hierarchical structure of SNOMED CT is necessary for supporting meaning-based entry and retrieval for clinical purposes.

Terminological challenges in configuring Clinical Information Systems

Configuration of CISs is central for terminology implementation, because this includes decisions about what clinical content should be included in the CISs and how the content should be represented. Hence, their decisions are foundation for how data can be retrieved and reused when the systems come in routine use.

According to (Rosenbeck et al. 2010) clinical content can be defined as the clinical knowledge built into CISs. Clinical content is expressed as domain specific terms, rules and structures, with the overall purposes being:

- To define the interface terminology and relation between interface terminology and a standardized clinical terminology, such as SNOMED CT.
- To define the constraints in the input to EHR systems; e.g., to define a diastolic blood pressure as a number within a certain range.
- To structure the GUI of EHR systems, in order to support a certain clinical workflow

(Rosenbeck et al. 2010)

When implementing a clinical terminology in a CIS, the people responsible for configuration must be able to apply these decisions in a way which utilizes the terminology potential to achieve the vision for clinical data use. A. Rector states that the main objective of clinical terminologies is implementation and use.

“SNOMED CT itself is only a part of the solution to addressing the requirements for effective electronic clinical records. A terminology on its own “does” nothing unless it is implemented as part of an application and used.”

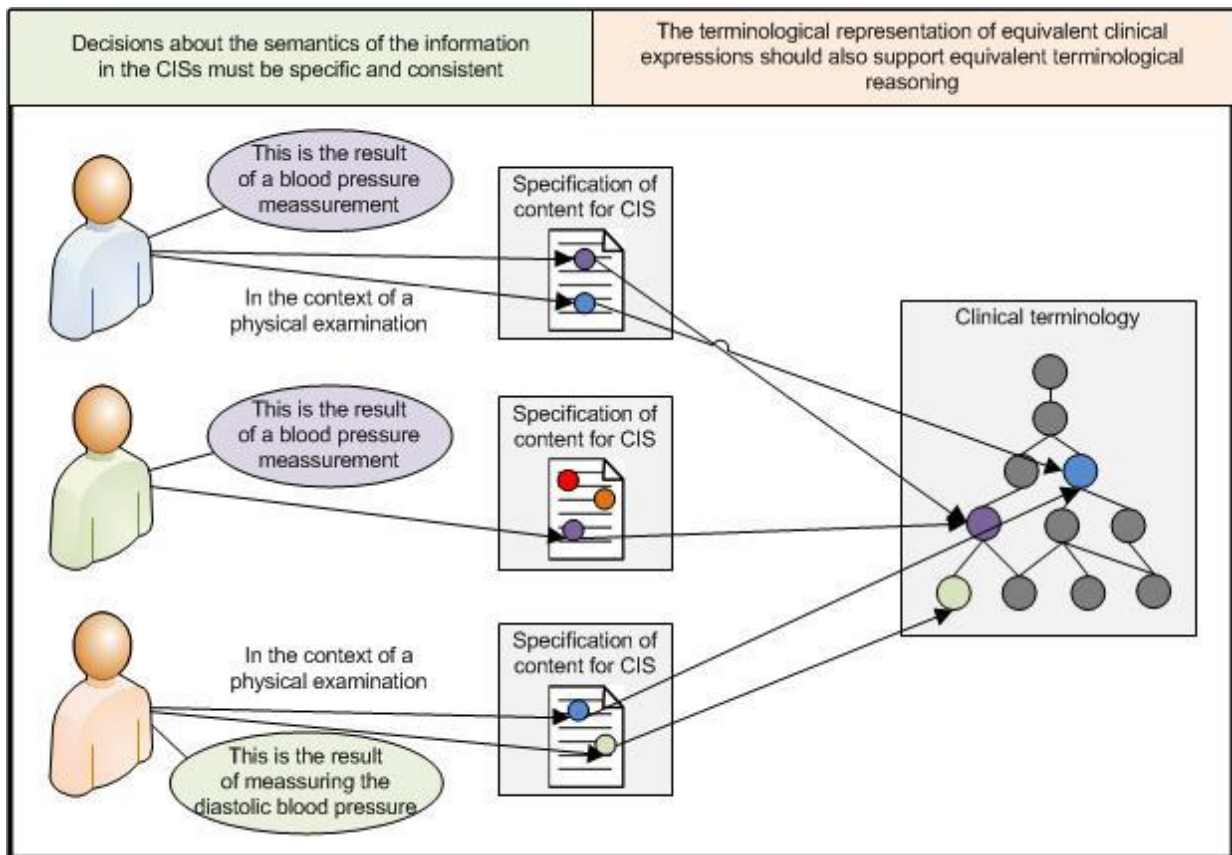
(IHTSDO 2012-07-31b)

One concept system does not ensure data reuse

Applying concepts from the same clinical terminology is not adequate to fulfill requirements for reusability of data. Reuse of clinical data requires for data to be comparable in order to be able to terminologically reason for common and diverse characteristics. Firstly, comparability of data requires specific and consistent assessment about the semantics of the clinical expressions to be documented as illustrated in

Figure 4. Secondly, the terminological representation for that specific clinical meaning should also be consistent and support equivalent terminological reasoning. Hence, supporting comparability in SNOMED CT enabled applications demands for consistent and well balanced SNOMED CT implementations.

Figure 4 Comparability of clinical data requires common representation and clinical reasoning of the applied concepts.

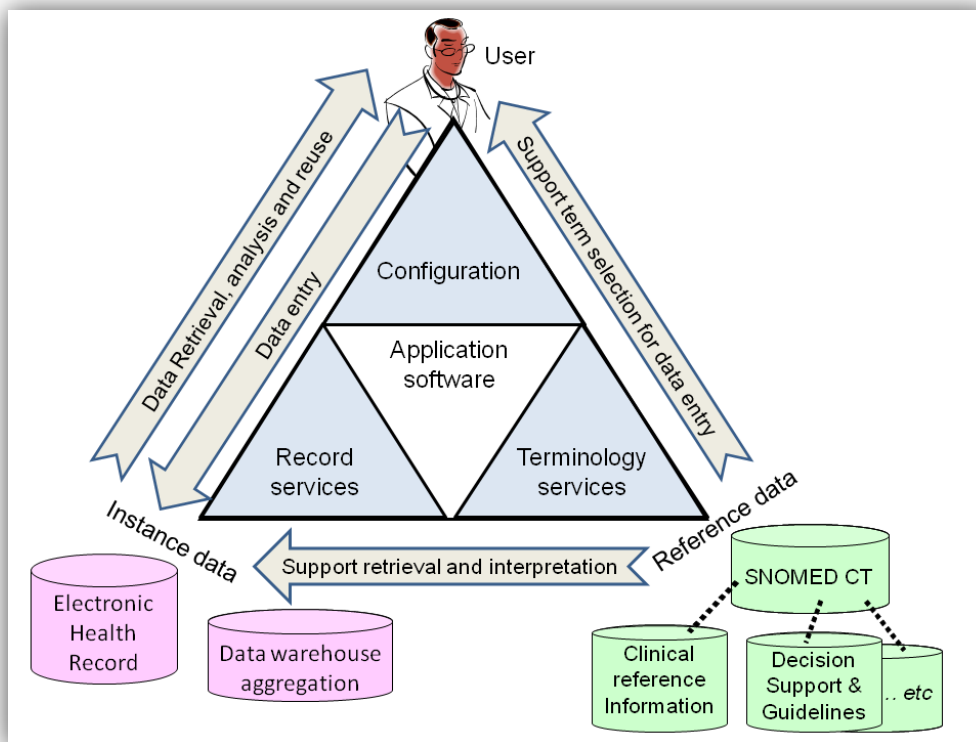


SNOMED CT has the potential to support comparability, however, state of art in SNOMED CT implementations indicate that application of medical terminologies to meet the anticipated benefits is not straight forward. The number of scientific papers on actual SNOMED CT implementations or experiences with utilizing SNOMED CT for specific retrieval purposes are limited (Cornet, de Keizer 2008, Lee et al. 2012, Lee et al. 2013).

The importance of mapping consistency

IHTSDO defines three central parts of developing SNOMED CT enabled software, namely configuration, record services and terminology services, as illustrated in Figure 5, derived from (IHTSDO 2012-07-31b).

Figure 5 Relationship between users application software and SNOMED CT. January 2013 International Release, Copyright 2002 2013 International Health Terminology Standards Development Organisation (www.ihtsdo.org)



terminology services enable access to Reference data, i.e. the terminological foundation of the application, in this case SNOMED CT and additional information, such as clinical guidelines, decision support, quality measures etc. Through reference sets, IHTSDO provides a formalized way of representing the sets of SNOMED CT components (e.g. concepts, descriptions, relations) to be included in a record. However, when selecting what concepts to be included in a specific set (to be used for data entry) it is important to be aware of the described issues on ambiguity and inconsistency, so that each set support data retrieval in order to meet the vision of reusability.

Various studies address the importance of mapping consistency and unambiguity when applying SNOMED CT as a standard reference for the content within the applied CISs. (Andrews et al. 2008a, Patrick et al. 2008, Elkin et al. 2010) show that even when using SNOMED CT as a common clinical terminology system, redundant representations of identical clinical information occur. Redundant use of SNOMED CT concepts can lead to a limited applicability, as retrieval, aggregation and comparison of clinical data across systems cannot be done in an accurate and trustworthy manner. Redundancy occurs due to the compositional structure of SNOMED CT which allows for post-coordination where two or more concepts are combined to form a specific clinical meaning,

With post-coordination there is a risk of composing expressions where the clinical meaning is already present as pre-coordinated expression in SNOMED CT. (Cornet 2009) Without services that support comparing for expression equivalence it is not possible to process and overcome redundancy and inconsistency. Alternatively, strategies can be applied supporting unambiguous use of concepts when implementing SNOMED CT to improve comparability of information. These challenges motivate the focus of this PhD project.

Research Objective

Poly-hierarchical clinical terminologies such as SNOMED CT, are useful for clinical documentation, as they support a representation of the detailed clinical information in a way that is computer processable. On one hand such terminologies address the needs and expectations of use of clinical information for user oriented and meaning-based retrieval. On the other hand, the comprehensiveness and the complex structure of such terminologies include issues related to consistency and ambiguity, as pitfalls to comparability.

Achieving data comparability within and across systems becomes challenging due to the practical implementation setting. One issue is coding variability among the persons who are responsible for selecting concepts to be used (Andrews et al. 2008a, Patrick et al. 2008), and another is the heterogeneity of the applied CIS, challenging data comparability. CIS are built up using different information models, the model of meaning and the model of use, and have different ways of storing and managing the data structures and terminologies. (Rector, Iannone 2012, Pape-Haugaard et al. , Rector, Qamar & Marley 2009). This project aims to explore pragmatic approaches to support consistent use of terminology as foundation for terminological harmonization of clinical data in CISs.

In the process of configuring CISs and striving to meet the user needs and expectations, deviations from theory will be inevitable. An objective of this study is to contribute to current research on SNOMED CT implementation by exploring how the theoretical potentials of SNOMED CT can be utilized in actual implementation settings to make reuse of clinical data possible within and across CISs.

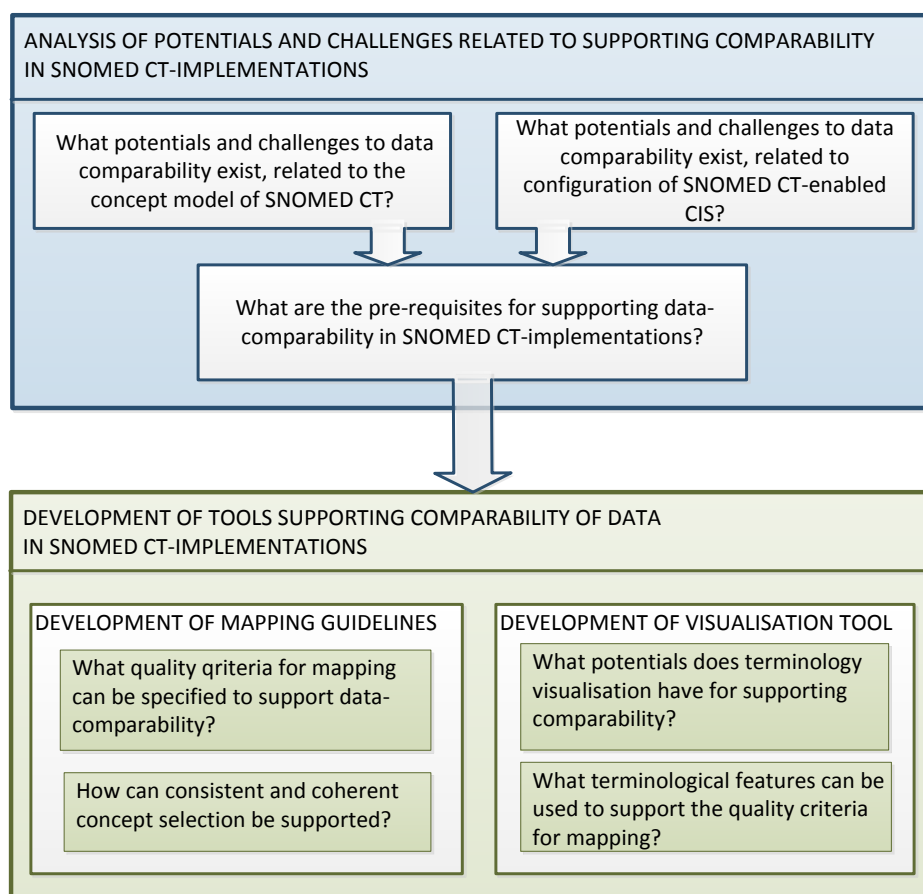
The research questions addressed in this study are:

1. *What are the prerequisites for supporting comparability in SNOMED CT implementations?*
And
2. *How can the complexity of SNOMED CT be managed, in order to support data comparability when selecting concepts for data entry?*

Method

The approach taken in this study to answer the research question can be described through the research framework presented by (Hevner et al. 2004) combining behavioral science and design science paradigms. This framework combines analysis of a specific environment, as people, organizations, and technologies with analysis of the knowledge base, being the theories, models, instructions and methodologies. The purpose of the analysis is to explore needs, prerequisites and state of the art knowledge on the subject area, to set up requirements for so called “artifacts” that can be used to meet the user needs and add knowledge to the existing knowledge base. Figure 6 illustrate the show the method applied to answer these research questions. The method is divided into two parts including an analysis of the potential and challenges related to SNOMED CT implementation and development of tools for supporting data comparability in SNOMED CT implementations.

Figure 6. Overall method for answering the research questions of this study.



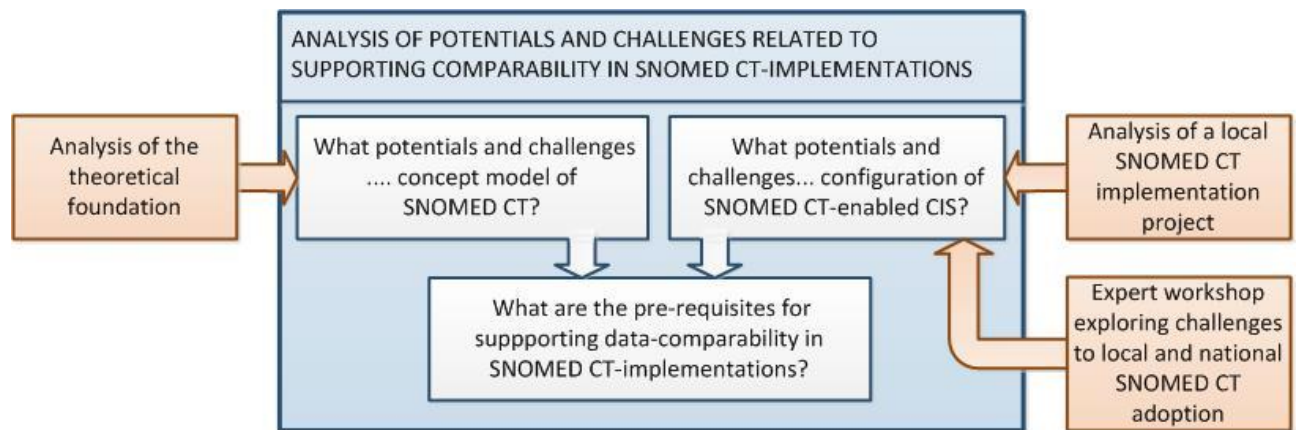
The analysis was conducted to answer the first part of the research question by analyzing the potentials and challenges within SNOMED CT as a clinical terminology along with exploring implications of applying SNOMED CT in an actual EHR implementation setting, in order to meet the theoretical potential. Hence, the analysis explores the environment and the knowledge base, according to the design science framework (Hevner et al. 2004)

The results of the analysis formed the basis for developing a set of detailed mapping guidelines and a visualization tool aimed at aiding the SNOMED CT implementation process. The decision to develop these two types of tools was made on the basis of knowledge obtained through the analysis about what challenges exist in SNOMED CT implementation projects.

Method for analyzing theory and practice of semantic comparability using SNOMED CT

A study of the theoretical foundation was conducted to analyze the potential and challenges to comparability related to the SNOMED CT concept model. Figure 8 illustrates the studies conducted to address the research questions.

Figure 7 Research analysis. Three studies were conducted to answer the research questions on potential and challenges related to data comparability



As described in the introduction, little is currently documented about actual implementation and use of SNOMED CT and Lee et al. state that *“Most implementations are not published in the scientific literature, requiring a look beyond the scientific literature to gain insights into SNOMED CT implementations”*(Lee et al. 2013). The analysis conducted in this research go beyond the scientific literature and explore SNOMED CT implementation through an actual implementation setting involving the configuration of an EHR system using SNOMED CT to represent the clinical content. Also, the analysis included an expert workshop, where stakeholders involved in SNOMED CT

implementation, including terminology experts, vendors of CISs and health information managers were gathered to discuss the potential and challenges related to local as well as national SNOMED CT adoption.

The following sections present how the analysis was conducted through both theoretical and empirical studies. The methods applied to develop tools to support SNOMED CT implementation are presented. A detailed description of the methods is found in the enclosed papers and reports.

Method for theoretical analysis

The theoretical analysis documented in the thesis presents the results of an analysis of the logical model of SNOMED CT, based on the technical specifications provided by the IHTSDO. However, both theoretical and scientific foundation of SNOMED CT has served as the underlying knowledgebase throughout this project. The aim first was to gain knowledge on the structure and composition of SNOMED CT. The second aim was to identify the parts of SNOMED CT that could impact use of data when the overall objective was to ensure the best possible prerequisites for reuse of the data. The last aim was to gain knowledge about the state of the art in the scope of SNOMED CT.

This theoretical analysis was conducted by reviewing existing scientific literature on SNOMED CT combined with an in-depth reading of the existing SNOMED CT documentation provided by IHTSDO. Reviewing existing SNOMED CT literature was an iterative process, and was conducted throughout the project period. The scientific literature has been point of departure, and the topics addressed in the literature have been further reviewed in the SNOMED CT documentation to fully understand the application and the logic behind that specific topic. For example, when specific topics, such as 'primitive concepts' or 'concept equivalence' has been addressed in a paper the SNOMED CT documentation has been examined to understand what the actual definition, purpose and logic behind the specific topic. This approach has led to an increased understanding of the specific topics being addressed in the literature and also an increased knowledge on the topics of concern related to SNOMED CT implementation. Additionally, examining SNOMED CT documentation led to insight on related topics which then served as input for new literature searches. A powerful method for identifying high quality literature from different sources is snowballing (Greenhalgh, Peacock 2005), which was used when searching for papers addressing related topics of interest.

Method for empirical analysis of SNOMED CT implementation in practice

This study was conducted based on implementation of SNOMED CT in an EHR system in the Northern Region NJ in Denmark. The terminology was implemented concurrently while configuring the EHR system. The aim was to assess prerequisites for enabling use of SNOMED CT that support quality, consistency and reusability of the clinical content within the EHR system.

Background

The vision in Denmark is to apply to common reference architecture and national standards for improving the quality and efficiency of clinical practice, to enhance the coherency between CISs and to facilitate communication across the actors within the health sector. (Statens Serums Institut 21st June 2013) Nationally, SNOMED CT has been translated into Danish, with the aim to use it for structured documentation in EHR systems. However, there exists no detailed strategy for how to bring SNOMED CT into routine use in Denmark and the Danish regions have been reluctant to initiate use of SNOMED CT. Despite the great effort to translate and distribute SNOMED CT, initiatives and demands from related stakeholders are not appearing and the clinical IT projects are limited to local implementations. The link between the national vision and regional terminology issues related to development of clinical content to be implemented in the EHR systems seems to be lacking.

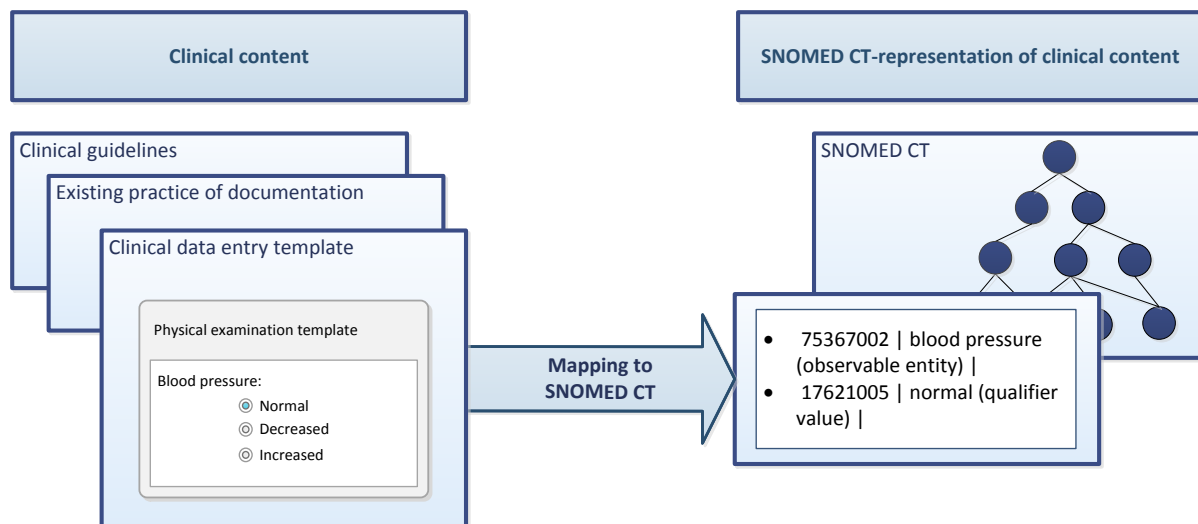
Whether or not SNOMED CT is being adapted within the Danish regions there is a risk of redundant efforts. Either because five (or more) regional terminologies are being invented which later must be adapted to SNOMED CT, or because SNOMED CT can be implemented very differently and these must be harmonized afterwards.

Obtaining consensus on how SNOMED CT should be adapted in the regional EHR systems, is therefore seen as an important national task. However, it is difficult to establish national guidelines for the use of SNOMED CT with no experience in the field. Therefore, the aim of this study is twofold. First, to understand the specific terminology related problems in the setting of Region Northern Jutland (Region NJ) and to contribute to the knowledge about the methods and guidelines relevant to manage SNOMED CT implementations from both national and international perspectives.

Using SNOMED CT for defining clinical content in Northern Jutland Region

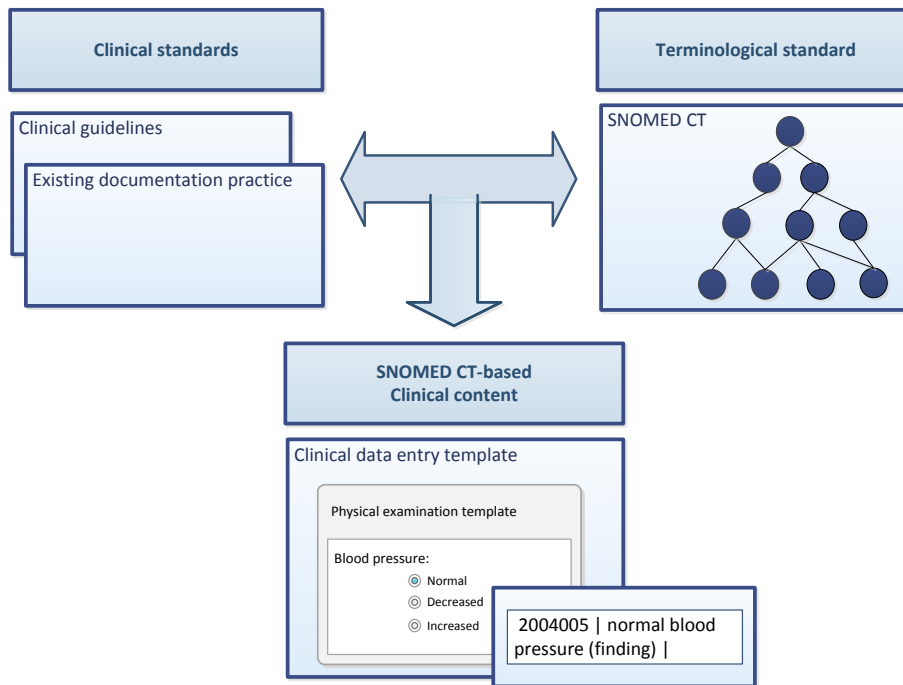
The typical approach for assessing SNOMED CT's ability to support clinical documentation is by applying SNOMED CT to existing interface terms while aligning with the existing composition of the content in the clinical templates, see Figure 8.

Figure 8 Typical approach for mapping interface terminologies to SNOMED CT. The mapping reflects the specific terms and composition of the clinical content source.



As it cannot be expected that the content of each sub hierarchy of SNOMED CT is directly correlated with the content of all existing formalizations, there is a need for mutual adjustment in the process of developing the clinical content. For example, some clinical expressions in existing formalizations may need to be composed using post-coordination to represent the specific detail, or the logical relations between the content in existing formalizations may differ to that of SNOMED CT. The applied approach taken in this study, as illustrated in Figure 8, should to a greater extent allow for SNOMED CT to be a source of inspiration for how the formalization can be made. This means that both the clinical standards and informatics standards, including SNOMED CT, provide a common basis for defining a highly structured and implementation oriented formalization within a clinical area for example, by supporting meaning-based retrieval.

Figure 9 Strategy for developing SNOMED CT based clinical content in this study. The strategy relies on mutual adjustment of the clinical content structure and the choice of SNOMED CT concepts.



The local EHR configuration team from Region NJ posed the clinical content specifications to be configured, see Table 1 for an overview of the included material. The structuring level indicates whether the template was a low structured (LS) template or a high structured template (HS).

Table 1. Overview of EHR templates included in study.

EHR template	Number of clinical expressions	Structuring level
Physical examination	25	LS
Other organ systems	9	LS
Nursing status	15	LS
Social conditions	13	LS
Chronic Obstructive Lung Disease (COPD)	109	HS

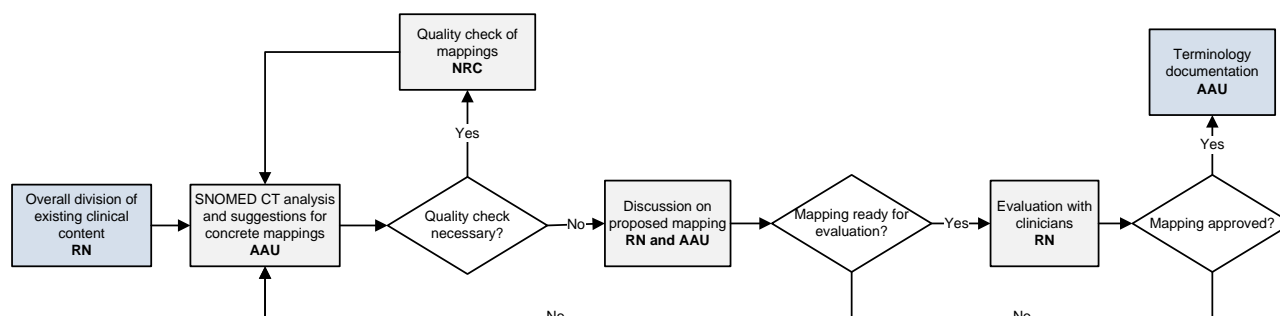
Most templates in the Region NJ EHR system were low structured templates, as they were clinical notes. For these templates a structured narrative approach was chosen. Structured narratives combine the familiarity, ease of use and freedom of expression of the narrative with the ability to browse data based on the gross structure represented by sections, fields and paragraphs. (Johnson, Bakken & Dine 2010) The COPD template was a high structured EHR template and included the use of lists, checkboxes, and constrained input fields. In the HS template the mapping involved much more granular concepts, whereas the LS template included more coarse grained concepts, covering for every possible narrative result.

Figure 10 Illustration of HS and LS templates.

<p>Objektiv undersøgelse</p> <p>Almentilstand</p> <p>Ernæringstilstand</p> <p>Objektivt psykisk</p> <p>Vitale værdier</p> <p>Hoved og hals</p> <p>Stethoscopia cordis</p> <p>Stethoscopia pulmonis</p> <p>Mammæ</p> <p>Abdomen</p>	<p>Eksacerbationer de sidste 6 måneder</p> <p>Antal</p> <p>Ekspektorat</p> <p>Ekspektorat sendt til D+R <input type="radio"/> Ja <input type="radio"/> Nej</p> <p>Kronisk ekspektoration <input type="radio"/> Ja <input type="radio"/> Nej</p> <p>BMI</p> <p>Vægt <input type="text"/> kg Højde <input type="text"/> cm BMI <input type="text"/> NIP</p> <p>Rygeadfærd</p> <p><input type="radio"/> Aldrig røget <input type="radio"/> Tidligere ryger <input type="radio"/> Ryger NIP</p> <p>Røget antal år <input type="text"/></p> <p>Cigaretter <input type="text"/> stk Cigarer <input type="text"/> stk 1 cigaret = 1g 1 cigar = 4g</p> <p>Cerutter <input type="text"/> stk Pibestop <input type="text"/> stk 1 cerut = 3g 1 pibestop = 3g</p> <p>Beregnet gram tobak dagligt</p> <p><input type="radio"/> 1-10g</p> <p><input type="radio"/> 11-20g</p> <p><input type="radio"/> 21-30g</p> <p><input type="radio"/> Mere end 30g</p> <p>Antal pakkeår <input type="text"/></p>
<p>Example of Low structured template (from the Physical examination template)</p>	<p>Example of High structured template (From the COPD template)</p>

Figure 11 illustrates the applied workflow of the SNOMED CT implementation project. The AAU researchers (Anne Randorff Højen and Kirstine Rosenbeck) did the actual concept mapping and then discussed potential mapping questions with representatives from the National Release Center (NRS), and the RN. In the first iteration of this study, the RN was intended to complete the mapping concurrent with AAU. However, due to limited competencies in effective SNOMED CT mapping, along with the labor intensiveness of this process and other important work tasks for the RN team, we chose the strategy illustrated instead. The RN did the mapping evaluation with the clinicians through workshops.

Figure 11 Workflow of the implementation project.



Expert workshop exploring challenges to SNOMED CT adoption

In the project in Region NJ the conditions for SNOMED CT implementation along with the related challenges were explored in a local EHR configuration setting. To underpin the knowledge gained through this study and to explore challenges to SNOMED CT adoption, including experiences and views from various stakeholders and

organizations, an expert workshop was conducted. The objective was to explore what challenges adoption of SNOMED CT face in a Danish perspective.

Representatives from three central groups of stakeholders related to SNOMED CT implementation, from different Danish positions were gathered to discuss experiences and expectations related to terminology management and specifically SNOMED CT, with the aim to explore reasons for lacking adaption of SNOMED CT in Denmark.

The method applied for this study was based on a mapping exercise conducted in predefined groups followed by a focus group interview aimed at discussing experiences related to terminology management and expectations to SNOMED CT adoption in Denmark. Also, the aim with this workshop was to evaluate the tools developed as part of this project, presented in the following. This study is documented in the enclosed paper "SNOMED CT adoption in Denmark. – Why is it so hard?", see Appendix "Paper D

Method for development of tools supporting semantic comparability when selecting SCT concepts

From the outcome of the analysis we were able to characterize the prerequisites for supporting SNOMED CT implementation and used these as the requirements for the tool development. A set of mapping guidelines and a tool for visualization of sets of SNOMED CT concepts was developed, to support SNOMED CT implementation facilitating data reusability.

Mapping guidelines

The mapping guidelines were developed with the objective to provide a clear methodology for SNOMED CT mapping to enhance applicability of SNOMED CT despite incompleteness and redundancy. The mapping guidelines were applied in an in-depth analysis of 14 different EHR templates retrieved from five Danish and Swedish EHR systems. Each mapping was assessed against defined quality criteria and the specified mapping guidelines. The mapping guidelines have been documented in the enclosed paper (Randorff Højen, Rosenbeck Goeg 2012), and the specific material and method of development can be found on page 63 and page 95.

Visualization tool

The visualization tool was developed with the objective to support the use of the developed mapping guidelines by providing an overview of sets of concepts and their hierarchical relationships in SNOMED CT. This study aimed to explore ways of illustrating common pathways and ancestors of particular sets of concepts, to support consistent use of SNOMED CT in EHR system implementation processes. The developed prototype is an interactive web-based reimplementation of the terminology visualization tool TermViz. It was developed as an

open source prototype and contains terminological features that are of relevance when exploring and comparing sets of concepts in SNOMED CT. This includes interactively rearranging graphs, fetching more concept nodes, illustrating least common parents and shared pathways in merged graphs etc. Further description of the applied method, technologies and evaluation results are presented in the enclosed paper, on page 69 and appendix "Paper C".

Results (part 1)

Analysis of theory and practice of semantic comparability using SNOMED CT

In this research a theoretical analysis has been conducted to explore how the logical model of SNOMED CT supports concept comparison. Moreover, an empirical analysis has been conducted exploring the potential and challenges to data comparability related to the configuration of SNOMED CT-enabled CISs, based on findings from a local SNOMED CT implementation project and an expert workshop.

Theoretical foundation for semantic comparability using SNOMED CT

This chapter explains how the logical foundation of SNOMED CT and hence, how the core components of SNOMED CT can serve as the terminological foundation for supporting semantic comparability in CIS.

The core structure of the SNOMED CT concept model is the concepts, the relationships and the descriptions. Each concept in SNOMED CT represents a clinical idea with an assigned unique concept identifier (conceptID). These concepts are also called pre-coordinated expressions. The conceptID is a numeric machine readable identifier, and it provides the possibility to unambiguously identify distinct concepts from each other.

The concept model of SNOMED CT specifies the logical definition of each SNOMED CT concept and is core to understanding how the compositional structure of SNOMED CT can support meaning-based retrieval. The subtype hierarchies and the concept definition using attribute relationships enable different ways of comparing concepts.

Logical foundation of SNOMED CT

A clear and consistent logical commitment is important to support correct and unambiguous interpretation of concepts and their meanings in CIS (Schulz, Cornet & Spackman 2011, Schulz, Cornet 2009). This is also why great efforts has been put into developing and refining the reference model of SNOMED CT according to the standard formalism provided in Description Logic (DL) (Spackman et al. 2002, Spackman, Campbell & CÃ 1997). DL is a formalism within the field of Knowledge Representation (KR) that uses logic-based semantics to consistently define concepts within a specific domain, and use these concepts to specify properties of objects and instances within that given domain. Moreover, an important feature of DL is to allow reasoning, which means that the implicit meaning of concepts can be inferred from the defined meanings of related concepts. (Baader, Lutz & Suntisrivaraporn 2006, Baader, Brandt & Lutz 2005).

The formal logic of SNOMED CT is represented through the DL framework \mathcal{EL}^+ , where concepts are defined by conjunctions of other concepts, subsumption (\sqsubseteq) as well as role-value pairs, which are represented as existential restrictions (\exists) (Schulz, Cornet & Spackman 2011).

Hierarchical comparison

SNOMED CT hierarchies

SNOMED CT contains more than 311,000 active concepts organized into hierarchies. The top level hierarchies define the types of concepts available to describe clinical information e.g. clinical findings, observable entities, procedures and body structures. (IHTSDO 2012-07-31b)

A main feature in DL that enable the logical property of the SNOMED CT hierarchies is subsumption, which means that all instances of a particular concept is also instances of the more general concepts (Baader, Nutt 2003). Exemplified by the illustration below (Figure 12) this means that all instances of the concept “Disorder of eye” is also instances of [Visual system disorder].

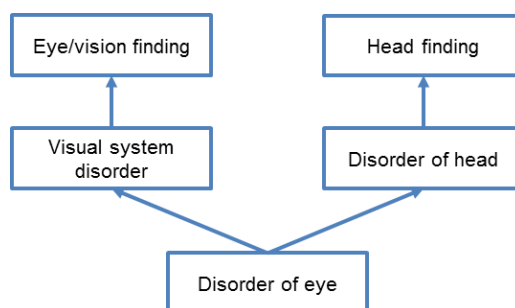


Figure 12 Subsumption hierarchy. [Disorder of eye] is subsumed by both [Visual system disorder] and [Eye vision finding], and this means that [Disorder of eye] inherits the properties of these two parent concepts.

In DL this is denoted using the subsumption operator (\sqsubseteq), and hence:

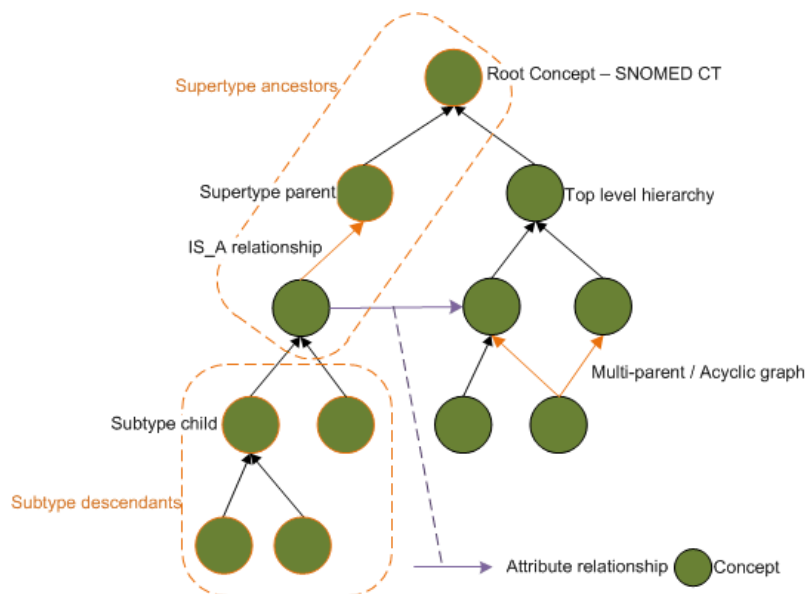
$$Disorder\ of\ eye \sqsubseteq Visual\ system\ disorder \sqsubseteq Eye\ vision\ finding$$

IHTSDO represent this type of subtype relationship using the [is a] relationship type. The [is a] relationships are always directed from a granular to a more general level. Every concept, except from the root concept has at least one supertype parent. To logically represent that a concept is subsumed by more than one parent concept, the logical operator “and” is used \sqcap :

$$Disorder\ of\ eye \sqsubseteq Visual\ system\ disorder \sqcap Disorder\ of\ head$$

Figure 14 gives an overview of the structure SNOMED CT hierarchies and the terms used to explain the core structure of SNOMED CT hierarchies.

Figure 13 Overview of the structure of SNOMED CT hierarchies and the terms used.



Methods for hierarchical comparisons

To explain how hierarchical comparison can be developed on a more advanced level than simple 1:1 ID-matching the logic behind subsumption testing is used. Subsumption testing uses information about the supertype ancestors of a concept to determine whether a concept is subsumed by another concept. The supertype ancestors can be determined by either traversing the graphs using the relationships table or by computing a transitive closure view of each concept. The latter is recommended by IHTSDO, but information about both can be found in the SNOMED CT documentation, see (IHTSDO 2012-07-31b).

Subsumption testing means testing one of the two equal hypotheses:

SUBSUMPTION HYPOTHESIS:

- The candidate concept is one of the subtype descendants of the predicate concept.

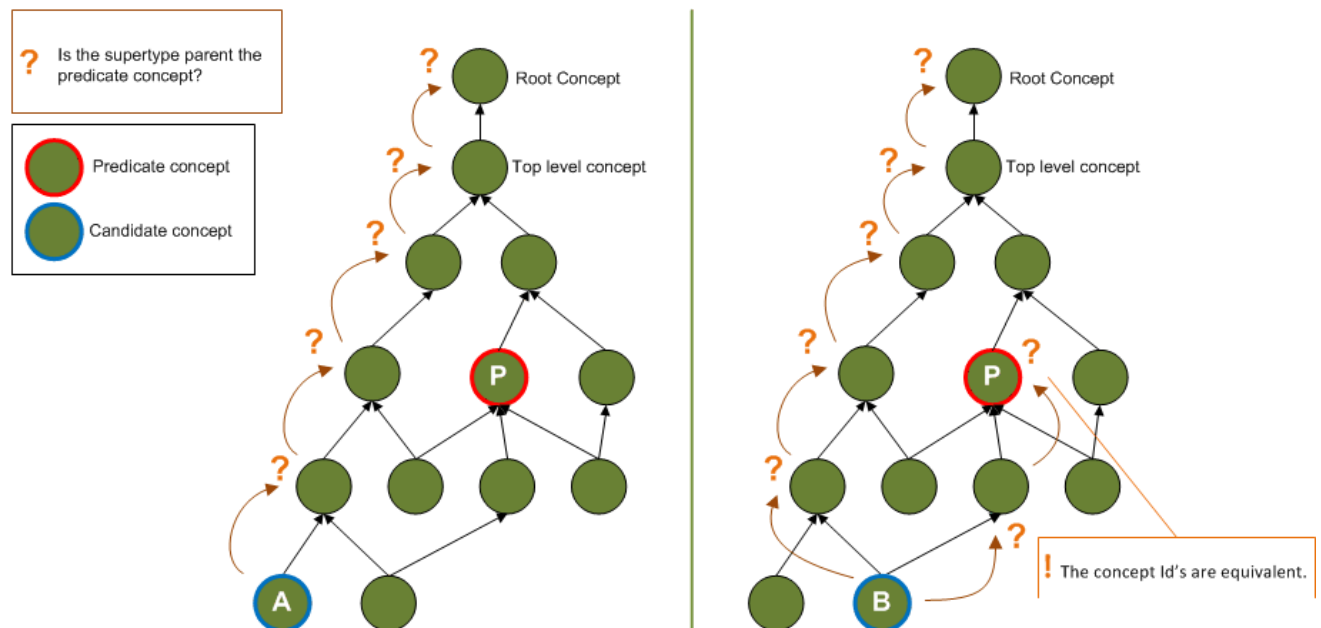
OR

- The predicate concept is one of the supertype ancestors of the candidate concept.

Figure 14 illustrates the basic idea with subsumption testing. In this example the test checks if the candidate concepts A and B are subtypes of the predicate concept P. The test checks if the conceptID's of all supertype ancestors are equivalent with the concept ID of the predicate concept. The result of the test to the left will reject the hypothesis, because the candidate concept is not one of the subtype descendants of the candidate concept.

The example to the right will accept the hypothesis, as the candidate concept B is subsumed by the predicate concept.

Figure 14 Illustration of the subsumption test using graph traversing. In the example to the left the result of testing the 'subsumption hypothesis' would be to reject the hypothesis and the example to the right would be to accept the hypothesis.



IHTSDO provides detailed guidance on how to perform and optimize the algorithms used for graph traversing. However, the basic steps of a subsumption testing of two concepts include:

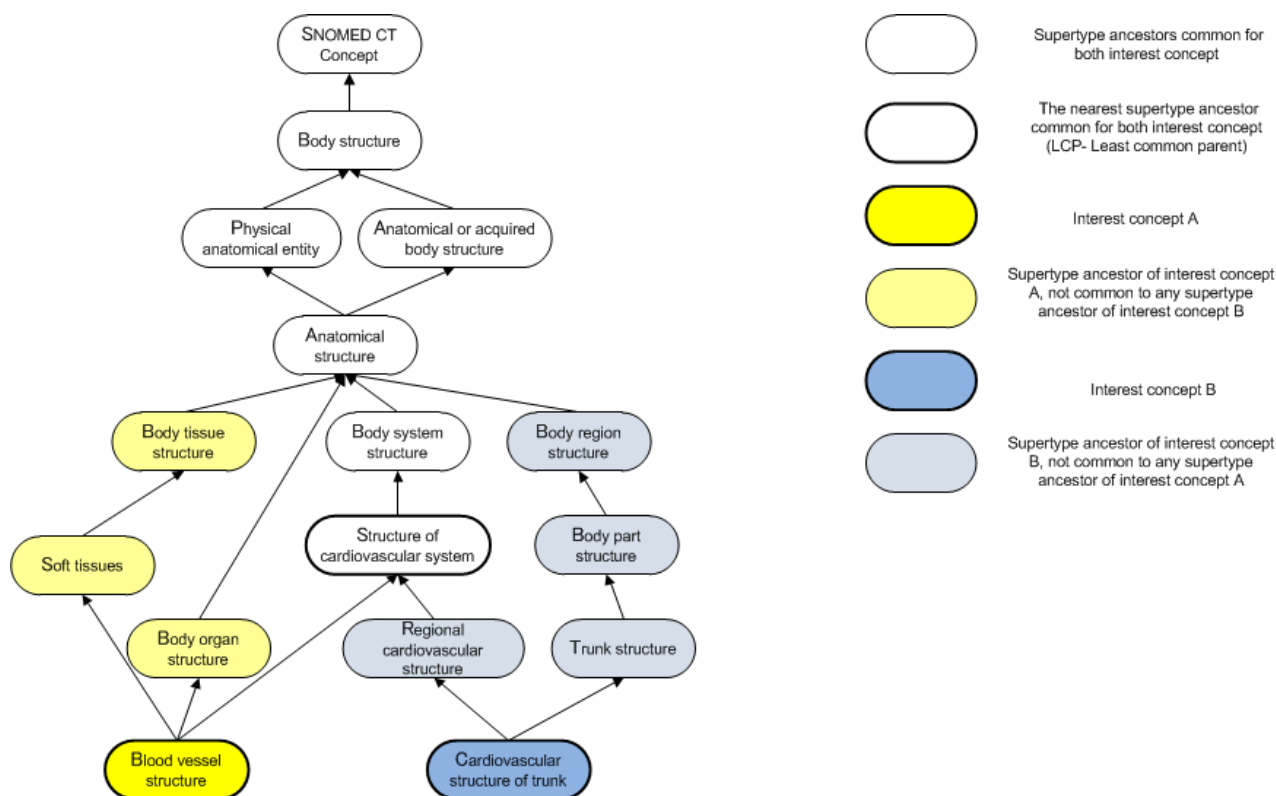
- 1) Checking if the concepts are active. If one, or both concepts are not active, the historical relationships [SAME AS] and [REPLACED BY] provide information of what concepts can be used for replacing the concept of interest.
- 2) Checking if the concept ID's of the two concepts are the same.
- 3) Traversing the subtype hierarchy.
 - a. The candidate concept is one of the subtype descendants of the predicate concept, or
The predicate concept is one of the supertype ancestors of the candidate concept.

Comparing concepts by the hierarchical structure of SNOMED CT requires services such as subsumption testing. However, comparing two or more concepts will require increased functionality, to determine common and uncommon supertypes of the concerned concepts.

Analyzing common supertypes of two or more candidate concepts can be used to determine the level for which it is possible to compare the concepts. Figure 16 shows how two concepts can be compared using the hierarchical structure of SNOMED CT. Considering the two concepts:

- Blood vessel structure (id: 59820001)
- Cardiovascular structure of trunk (id: 281233007)

Figure 15 Example of how the subtype hierarchy can be used to compare concepts.



Comparing the two concepts by their unique concept id's will obviously show that the two concepts are different. However, determining the common ancestors of both concepts will reveal that the two concepts have common hierarchical characteristics, see Table 2.

When querying data by using the hierarchical structure of SNOMED CT it is possible to use the common supertypes of a set of concepts to determine the level of granularity for which it is possible to compare the concepts. At a database level concept comparison is done by comparison of the concept IDs of the supertype ancestors of each of the concepts of interest. Table 2 include concept IDs of the supertype ancestors of the two concepts | Blood vessel structure | and | Cardiovascular structure of trunk |, and by comparing these Id's and counting the distance from the concept to each supertype concept it is possible to determine the common concepts and the nearest common concept, in this example | structure of cardiovascular system | .

Table 2. Supertype ancestors of the two interest concepts, and distance from interest concept and the actual supertype. The concept | structure of cardiovascular system | is common supertype for both interest Concepts, and has the smallest aggregated dist, whereas this concept is the Least Common Parent (LCP) for the two concepts.

Interest Concepts	Supertypes													
Blood vessel structure			87784001	113343008	113257007		91689009		85756007	91723000	442083009	91722005	123037004	138875005
Dist			1	1	1		2		2	2	3	3	4	5
Card. struc. Trunk.	22943007	281827007			113257007	38866009	91689009	52530000		91723000	442083009	91722005	123037004	138875005
Dist	1	1			2	2	3	3		4	5	5	6	7
Description	trunk structure	regional cardiovascular structure	soft tissues	body organ structure	structure of cardiovascular system	body part structure	body system structure	body region structure	body tissue structure	anatomical structure	anatomical or acquired body structure	physical anatomical entity	body structure	SNOMED CT Concept

The nearest common supertypes, also described as the Least Common Parents (LCPs), can be thought of as the lowest denominator for a set of concepts. The LCP is therefore valuable, as it represents the entry-point for a subtype query that retrieves all concepts of interest as a result.

Meaning-based comparison

SNOMED CT supports meaning-based comparison of data, because it enables a formal representation of the semantic meaning of a concept based on how a concept is defined through relations to other concepts. As for the SNOMED CT hierarchies the applied DL framework supports: A logical representation of concept definitions, enabling consistent representation of the defining characteristics of a concept and hence automatic reasoning. [#]

Concept definition

Besides the unique concept ID, a concept is represented through a set of defining characteristics. Each defining characteristic is represented through relations to other SNOMED CT concepts, as illustrated in Figure 16. The |is a| relationships are the core defining attribute used for all sub-hierarchies in SNOMED CT, but there are also other attribute relationships used to define concepts.

In DL the existential quantifier \exists is used to represent more complex relationships and to represent the existence of a specific role for a given concept. Hence, the formal expression for the existential quantifier is $\exists r.C$, where r is a defined role and C is a concept. This can be exemplified through the expression:

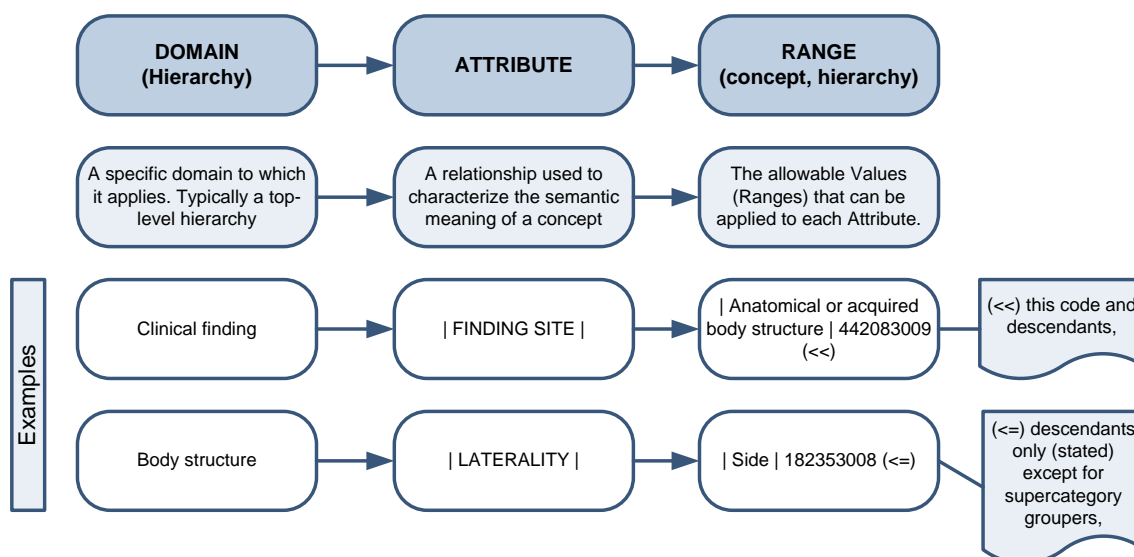
$$Cyst \sqcap \exists \textit{finding_site}.Breast$$

This expression is instantiated by all instances of *Cyst* and is related through the relationship | *finding_site* | some instances of *Breast*. Moreover, such existential expression can be used to represent the equivalence between specific concepts and expressions using the equivalence operator (\equiv), and hence representing specific defining properties of a given concept, for example that the meaning of the concept | *Cyst of breast* | is equivalent to the above expression:

$$Cyst \textit{ of breast} \equiv Cyst \sqcap \exists \textit{finding_site}.Breast$$

In SNOMED CT these existential quantifications are called attribute relationships, i.e. relationships that relate concepts across hierarchies and for each attribute relation a predefined domain and range is specified. This means that the concepts in SNOMED CT solely provide a unique clinical meaning, and at the same time it can serve as a part of another concept definition.

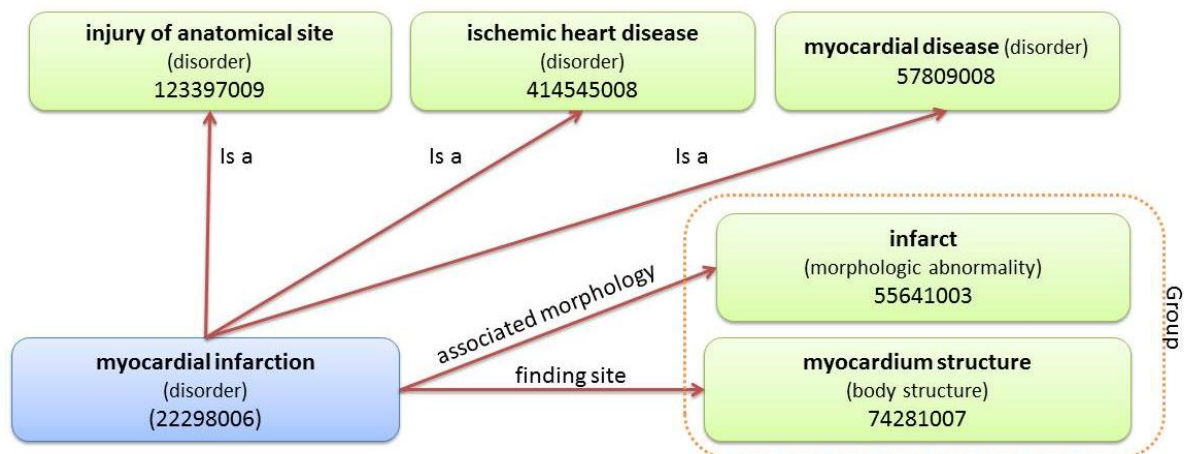
Figure 16 Attribute relationships in SNOMED CT. The Domain denotes the source of the relationship, the attribute is the SNOMED CT representation of roles, i.e. denoting the specific type of relationship. The Range represent the ‘range’ of values that a specific attribute can be applied to given a specific domain.



For example, Figure 17 illustrates the defining characteristics of the concept |myocardial infarction (disorder) | with the unique conceptID “22298006”. The defining characteristics can be used to interpret the meaning of the

specific concept, i.e. that a myocardial infarction is both a type of [injury of anatomical site], [an ischemic heart disease], [a myocardial disease] and it has a specified group relationship to the [associated morphology] [infarct] and the [finding site] [myocardium structure]. Hence, the defining characteristics represent what is definitely clinically true about this concept.

Figure 17 Defining characteristics of myocardial infarction.



The unique conceptID “22298006” and the associated description of [myocardial infarction] does not solely allow for any automatic interpretation (reasoning) of the concept and its defined at inherited meaning. This requires application of the logical model of SNOMED CT, and hence interpretation of the defining characteristics, providing a granular view of the meaning of a concept.

Relationship groups (role groups)

Relationship groups (marked “Group” in figure 18) were first introduced by (Spackman et al. 2002) with the development of SNOMED CT, emerging from SNOMED RT and Clinical Terms V3. Relationship groups (role groups, **rg**) were introduced to represent attribute-value pairs, i.e. making it explicit when two or more attribute relationships must be interpreted together to obtain the correct concept interpretation. These groups are important for concepts that involve more than one site, more than one method or more than one morphology.

An example of role groups in SNOMED CT is the procedure |cholecystectomy and exploration of bile duct| that is defined using two role groups:

- cholecystectomy and exploration of bile duct ≡
- ∃ **rg** (∃ *method.excision-action* ∩ ∃ *procedure site-Direct.gallbladder structure*)
- ∃ **rg** (∃ *method.exploration-action* ∩ ∃ *procedure site-Direct.bile duct structure*)

In this example the role groups are used to define that instances of this concept are excision of some instance of gallbladder structure and exploration of some bile duct structure. It is important to interpret these relationships

pair-wise, as the reverse reading of these relationships would include excision of bile duct, which would be a misinterpretation of this concept. Hence, these groups are important for concept comparison two concepts can include common attribute relationships and still be diverse because the relationships belong to different groups.

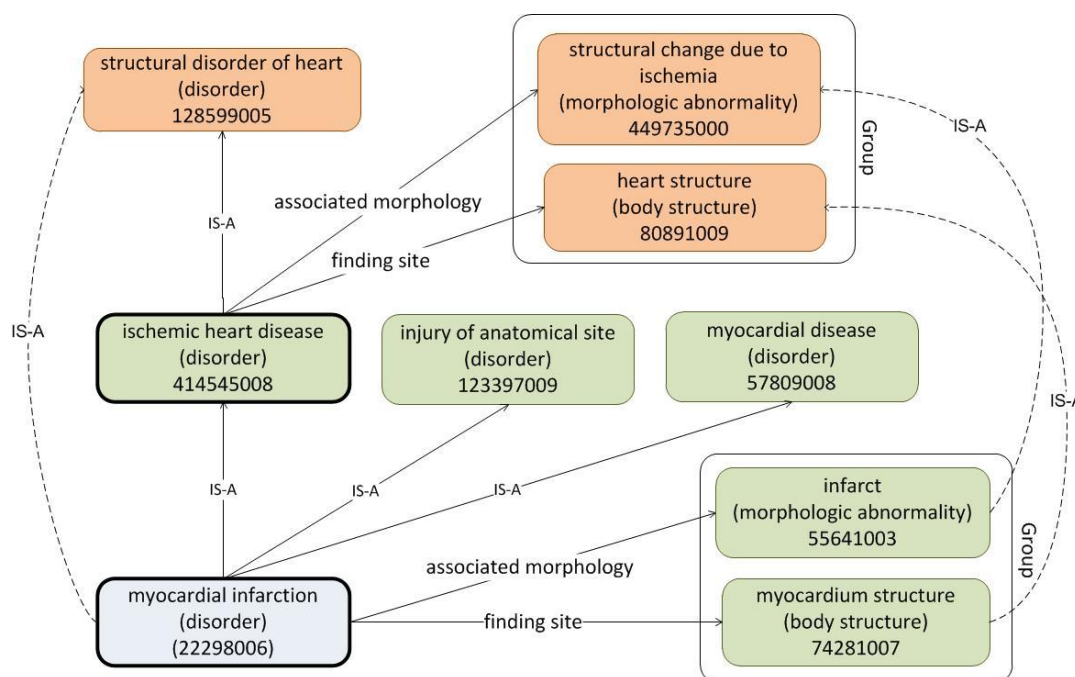
Comparison of concept definitions

Figure 18 illustrates the defining characteristics of the two concepts:

- 1) Ischemic heart disease (defining characteristics: orange)
- 2) Myocardial infarction (defining characteristics: green and orange(inherited))

As illustrated in Figure 18, |myocardial infarction| is a subtype of |ischemic heart disease|. The defining characteristics of Ischemic heart disease are the relationship group of associated morphology |structural change due to ischemia| and a finding site in the |heart structure|.

Figure 18 An overview of the inherited definition of the concept ‘myocardial infarction’. The orange concepts are inherited defining characteristics which are attached to ‘myocardial infarction’ through its relationship to the concept ‘ischemic heart disease’



The concept |myocardial infarction| inherits the defining characteristics |ischemic heart disease|, i.e. *Myocardial infarction* \sqsubseteq *Ischemic heart disease* \sqsubseteq *Structural disorder of heart* \sqsubseteq ... \sqsubseteq *Clinical finding*, and from a retrieval perspective the concept |myocardial infarction| will be a result of queries asking for concepts that are:

- an |ischemic heart disease|

- an |injury of anatomical site|
- a |myocardial disease|
- a disease with an associated morphology |infarct| and a finding site |myocardium structure|

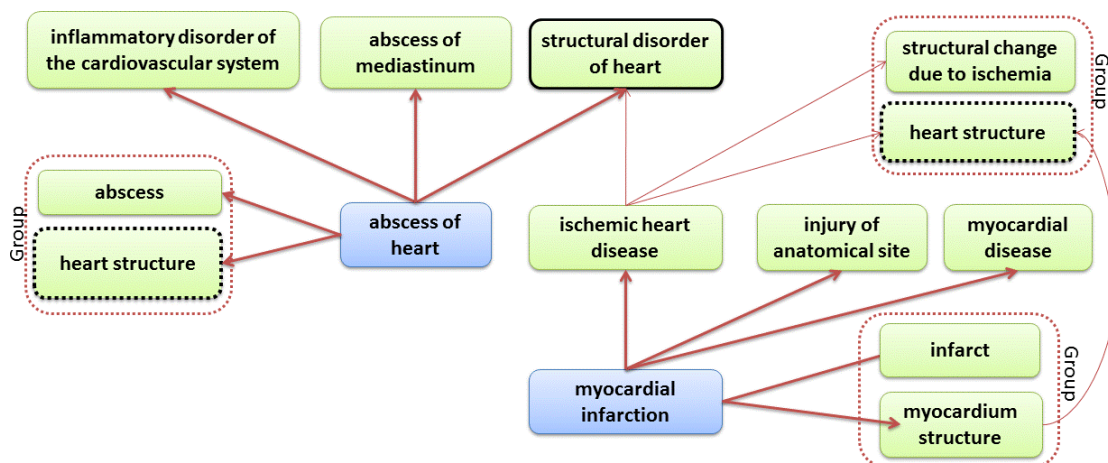
And because of its inherited meaning, the concept myocardial |infarction| will also be a result of queries asking for concepts that are:

- a |structural change due to ischemia|, (as *infarct* \sqsubseteq *structural change due to ischemia*)
- a disease with an **associated morphology** |infarct| and a **finding site** |heart structure| (as *infarct* \sqsubseteq *structural change due to ischemia* and *myocardium structure* \sqsubseteq *heart structure*)

Taking into consideration the definitions of each of the supertypes of |myocardial infarction| will increase the number perspectives where the concept can be identified as a result of a specific query. Additionally, the defining characteristics can be used to compare both pre- and post-coordinated concepts and enable selective meaning-based data retrieval.

In the common view of concepts |abscess of heart| and |myocardial infarction| shown in Figure 19 it is illustrated that the concept |structural disorder of heart| is a common concept, as it serves as a supertype for both concepts. The inherited defining characteristic of |myocardial infarction| is valuable when one wants to compare concepts to concepts that are not sibling concepts. Another comparable view is the finding site |heart structure|. Retrieving disorders that have the finding site |heart structure| will result in retrieving both |abscess of heart| and |myocardial infarction|.

Figure 19. Common view of the defining characteristics of the concepts |abscess of heart| and |myocardial infarction|.



These exemplifications illustrate how the defining relationships in SNOMED CT can be used to develop advanced query features. Moreover, it emphasize the importance, and necessity, of applying the logical model of SNOMED CT in the implementation of SNOMED CT, as this is essential to utilize the concept definitions for retrieval

purposes. The various defining relationships can serve as entry point for a query predicate, and enable that two or more concepts can be compared on their common defining characteristics.

Primitive concepts

As described above the logical model of SNOMED CT is designed to support consistent representation of concepts and their definitions, based on the formal and lightweight DL framework \mathcal{EL}^+ .

However, the potential to rely on the defining characteristics of concepts can be challenged if all the relationships required to define the meaning of a concept is not specified, as is the case for primitive SNOMED CT concepts.

Concepts in SNOMED CT are regarded as either primitive or fully defined. The defining characteristics of a concept are represented as one or more attribute relationships (and/or relationship groups) that are necessarily true for all instances of that concept (IHTSDO 2012-07-31b). For example, the concepts |myocardial infarction| and |myocardial infarction with complication| is defined using the same attribute relationships:

Myocardial infarction:	Myocardial infarction with complication
myocardial infarction \equiv myocardial disease \sqcap ischemic heart disease \sqcap necrosis of anatomical site \sqcap \exists rg (\exists associated_morphology .infarct \sqcap \exists \exists finding_site . myocardium structure)	myocardial infarction with complication \sqsubseteq myocardial infarction \sqcap \exists rg (\exists associated_morphology .infarct \sqcap \exists finding_site . myocardium structure)
Fully defined (\equiv)	Primitive (\sqsubseteq)

The concept |myocardial infarction| is fully defined as it is defined through relationships that are sufficient to define and distinguish it relatively to its immediate supertypes. In contrast, it is not possible to distinguish |myocardial infarction with complication| from its supertype |myocardial infarction|, and therefore this concept is primitive. In DL the fully defined concepts are denoted using the equivalence operator (\equiv), and primitive concepts are denoted using the subsumption operator (\sqsubseteq).

A primitive concept can be primitive for two reasons:

- 1) SNOMED CT does not contain any concept that can be used to sufficiently define the concept and distinguish it from its parent and sibling concepts.
- 2) Relationships necessary to fully define a concept needs yet to be specified. Defining attributes in SNOMED CT were first assigned to those hierarchies where retrieval of clinical data is most useful and relevant: procedure, finding, and situation with explicit context clinical findings. (IHTSDO 2012-07-31a)

Regardless of the reason for lacking concept definition, the use of primitive concepts has implications for data retrieval potential (Randorff Rasmussen, Rosenbeck 2011). Therefore it is important to be aware of how primitive concepts are used in SNOMED CT implementation projects. In the enclosed paper (Randorff Rasmussen, Rosenbeck 2011), the primitive concepts are used as a measure to analyse the retrieval potential of a concept. See Appendix "Paper A".

Post-coordination

Besides from representing the defining characteristics of a concept using the logical relationships it is also possible to combine two or more concepts to build new meanings. This process is called post-coordination and the attribute relationships are used to form the post-coordinated expressions. Post-coordination has implications for the adaptability of SNOMED CT as it has implications for the size of the terminology and the scope coverage. Because of post-coordination it is possible to limit the amount of pre-coordinated concepts without reducing the content coverage of SNOMED CT within specific clinical domains. This is because new meanings can be composed to fit a specific clinical expression without being limited by the amount of pre-coordinated concepts. This underpins the flexibility of SNOMED CT and allowing for users to define the level of granularity needed for a specific purpose, and the defined rules for post-coordination support only logical correct compositions to be made. (IHTSDO 2012-07-31b, Cornet 2009)

Besides from these benefits post-coordination can also support structured data entry and facilitate data retrieval. Post-coordination can for example support consistent data entry by allowing expressions to be constructed identically each time that information is documented. Compared to searching through hundreds of similar terms to find a precise match, post-coordination can support an effective way of supporting consistency between sets of similar variants for different concepts. An example of this is Figure 20 illustrating how post-coordination is used to specify laterality of a specific group of clinical findings. The right side of the figure shows how the post-coordinated expression can be stored in a normalized form supporting equivalent characteristics to be queried.

Figure 20 Using post-coordination for specifying laterality of specified clinical findings.

<div style="border: 1px solid gray; padding: 5px;"> <div style="background-color: #f0f0f0; padding: 2px; margin-bottom: 5px;">Select finding...</div> <ul style="list-style-type: none"> fracture of tibia <li style="background-color: #e0e0e0;">fracture of foot fracture of patella fracture of acetabulum fracture of ankle fracture of femur fracture of fibula <div style="margin-left: 20px;"> Specify laterality: <input checked="" type="radio"/> Left <input type="radio"/> Right </div> </div>	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <th style="background-color: #f0f0f0;">Close to user view</th> </tr> <tr> <td>15574005 fracture of foot : 272741003 laterality = 7771000 left </td> </tr> <tr> <th style="background-color: #f0f0f0;">Normal form</th> </tr> <tr> <td>64572001 disease : { 116676008 associated morphology = 72704001 fracture , 363698007 finding site = (84167007 bone structure of foot : 272741003 laterality = 7771000 left) }</td> </tr> </table>	Close to user view	15574005 fracture of foot : 272741003 laterality = 7771000 left	Normal form	64572001 disease : { 116676008 associated morphology = 72704001 fracture , 363698007 finding site = (84167007 bone structure of foot : 272741003 laterality = 7771000 left) }
Close to user view					
15574005 fracture of foot : 272741003 laterality = 7771000 left					
Normal form					
64572001 disease : { 116676008 associated morphology = 72704001 fracture , 363698007 finding site = (84167007 bone structure of foot : 272741003 laterality = 7771000 left) }					

Challenges in using post-coordination also exist and must be considered along with the above mentioned potentials of post-coordination when implementing SNOMED CT in CISs.

Post-coordination can influence the human readability of clinical terms, and induce that well-known and simple clinical terms are not used for data entry. For example, the diagnosis appendicitis can be composed as a disease with associated morphology | inflammation | and the finding site | appendix structure |. And the procedure [endoscopy] can be composed as a procedure with method | inspection – action | and using device | endoscope |. Hence, post-coordination for direct data entry is a balance act that must take into consideration human readability as well as the complexity of entering composite expressions compared to entering a single term.

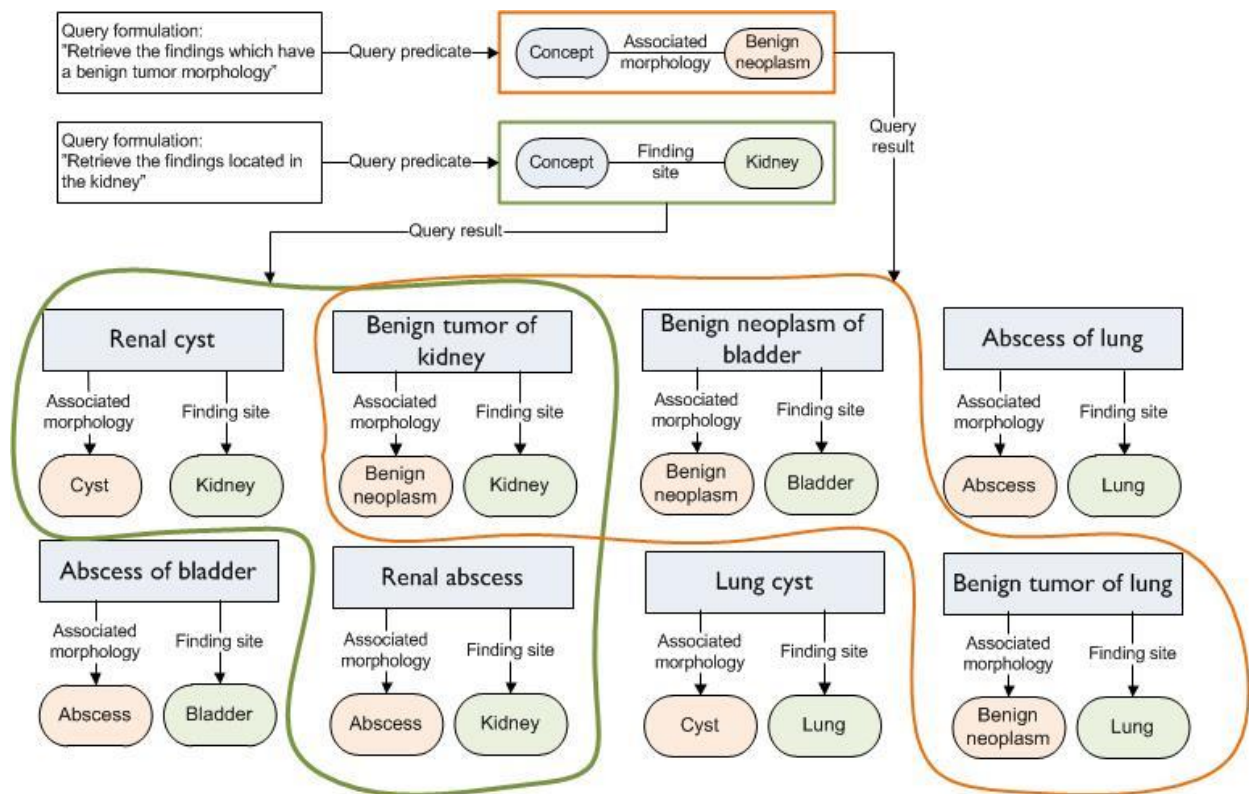
Post-coordination also has implications for data storage and retrieval, because of the increased complexity of storing and querying expressions of a variable and potentially very long length compared to single numeric identifiers. Related to retrieval it is a challenge that several post-coordinated expressions may mean the same thing and may also mean the same as a pre-coordinated concept. Hence, it requires specific services that support retrieval and comparison of both pre- and post-coordinated expressions to utilize post-coordination and hence allow for post-coordination in a data entry situation.

Method for retrieving and comparing pre- and post-coordinated expressions

In the paper *“Selective retrieval of pre- and post-coordinated expressions”* Dolin et. al. introduced a method for which pre- and post-coordinated expressions can be retrieved and compared to a query predicate, enabling retrieval of concepts based on their defining characteristics. (Dolin, Spackman & Markwell 2002) To underpin comparison of query predicates in pre- and post-coordinated expressions each expression is transformed into a common form. IHTSDO describes how concepts can be transformed into normal forms aimed at expression comparison (IHTSDO 2012-07-31b).

Figure 21 illustrates the idea behind meaning-based retrieval. The query predicate is transformed into a form which is directly comparable with that for the expressions. As shown in Figure 21 a query predicate may specify that any expression with the | finding site | defined as | Kidney | should be identified as a result of the query. These expressions are encircled with green. Additionally, a query can be specified to retrieve all expressions with an associated morphology | benign neoplasm |. These expressions are encircled with orange.

Figure 21 Illustration of selective retrieval



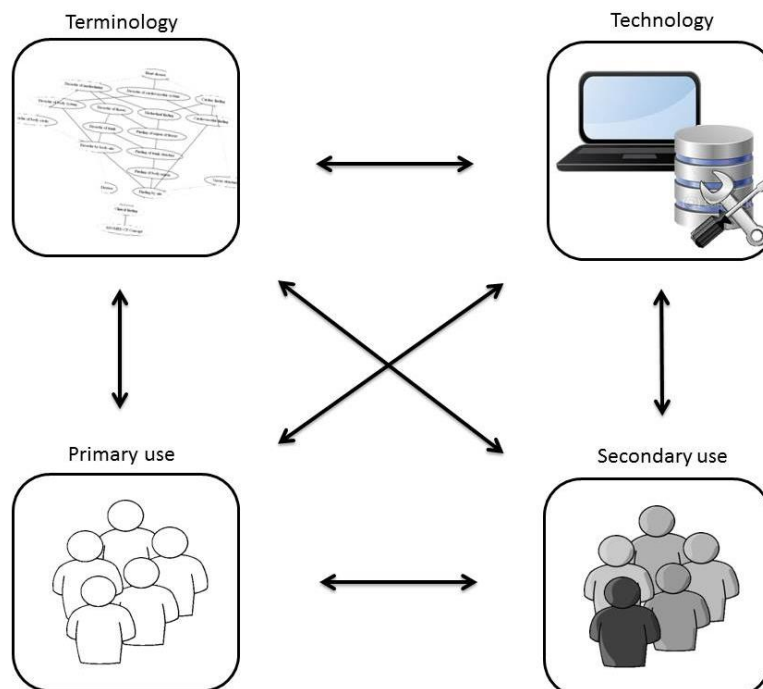
The theoretical analysis of how SNOMED CT support comparability of data shows that at least three levels of comparison can be supported: Comparing concept IDs, hierarchical comparison and meaning-based comparison. The findings are summarized in "Summary of findings of the theoretical and empirical analysis" page 58.

Empirical study of semantic comparability using SNOMED CT

Implications for people configuring CISs

SNOMED CT implementation is not solely about representing clinical knowledge through appropriate concepts. One thing is to map an interface term to a specific SNOMED CT concept, but the concept representation should also be compatible with the information model of the applied technology and support the intended use of data. Beside the need for deep insight to the terminology, SNOMED CT implementation is therefore matters of balancing the needs from different stakeholders involved as well as balancing the technological and terminological framework, see Figure 22. The people making decisions about system configuration therefore play a central role in enabling comparability when implementing SNOMED CT.

Figure 22 EHR-configuration must consider both the terminology, the technology along with primary and secondary needs when configuring EHRs. Therefore, SNOMED CT implementation is not solely a question about selecting concepts to fit a specific interface term.



Analyzing SNOMED CT implementation has shown two important prerequisites for use and implementation. These include:

- Knowledge and mapping skills to support consistent use of SNOMED CT
- Information management processes and systems to support coherence across borders

SNOMED CT competencies and mapping strategies

Working with the configuration team in Region NJ clarified the need for SNOMED CT competencies and mapping strategies in order to facilitate consistent concept selection, and thereby supporting data comparability.

Free text search is a common starting point for approaching concept selection in SNOMED CT. However, a free text search that is done without constraining the search result to fit the overall context of a clinical expression will often lead to a large number of search results. Therefore, it is important to have knowledge on the structure of SNOMED CT and knowledge on how to assess whether a concept represents the meaning of the clinical expression.

When mapping interface terms to SNOMED CT the search will often be conducted with the interface term being used directly as the search string. However, the interface term used in a specific clinical context, e.g. as a label to a data entry field could potentially not be accurate with respect to the meaning that it represents. The interface term is therefore only part of the criteria to consider when conducting a SNOMED CT search. It is more important to go beyond the term and focus on the meaning which the term represents. This became evident from the expert workshop where the mapping-exercise showed that knowledge on documentation practices, on SNOMED CT structure and of the overall strategy for how the resulting clinical information system was to be used is essential for selection concepts. Similar expressions may carry different meaning and the resulting mapping will induce inconsistency in use of terms in the system if mapping of terms becomes a “free text search” in a large hierarchy.

In Denmark, where the translation process has not included development of Danish synonyms the search is even more challenging, as only one term is used to designate the concept from the relationships to other concepts. This situation increases the need for the people configuring systems to know how to interpret the meaning of a concept based on its defining characteristics and its place in the subtype hierarchy.

Studies show that even by the use of SNOMED CT, as a common clinical terminology system, redundant representations of identical clinical information occur (Spackman 2001, Andrews, Richesson & Krischer 2007). This is a significant barrier to consistent data representation and thereby to reusability of data. When selecting concepts from SNOMED CT it is not enough to simply find a term that matches the concerned interface term. From this study we experienced that variability occurred among the people configuring systems due to:

- Different interpretations of the same concepts resulted in selection of different concepts to represent the same clinical meaning.
- A specific clinical meaning cannot be found as a concept in SNOMED CT, either because of inadequate search or because of missing pre-coordinated concepts.
 - Using post-coordination to represent a clinical meaning can induce different compositions of pre-coordinated expression to form the same clinical meaning.

To overcome the above challenges, knowledge on SNOMED CT will not solely assure consistent concept selection. To avoid ambiguous representation of the same meaning it is important to have a clear mapping strategy which supports the process of selecting concepts consistently when selecting pre-coordinated concepts and also for post-coordination.

Content management

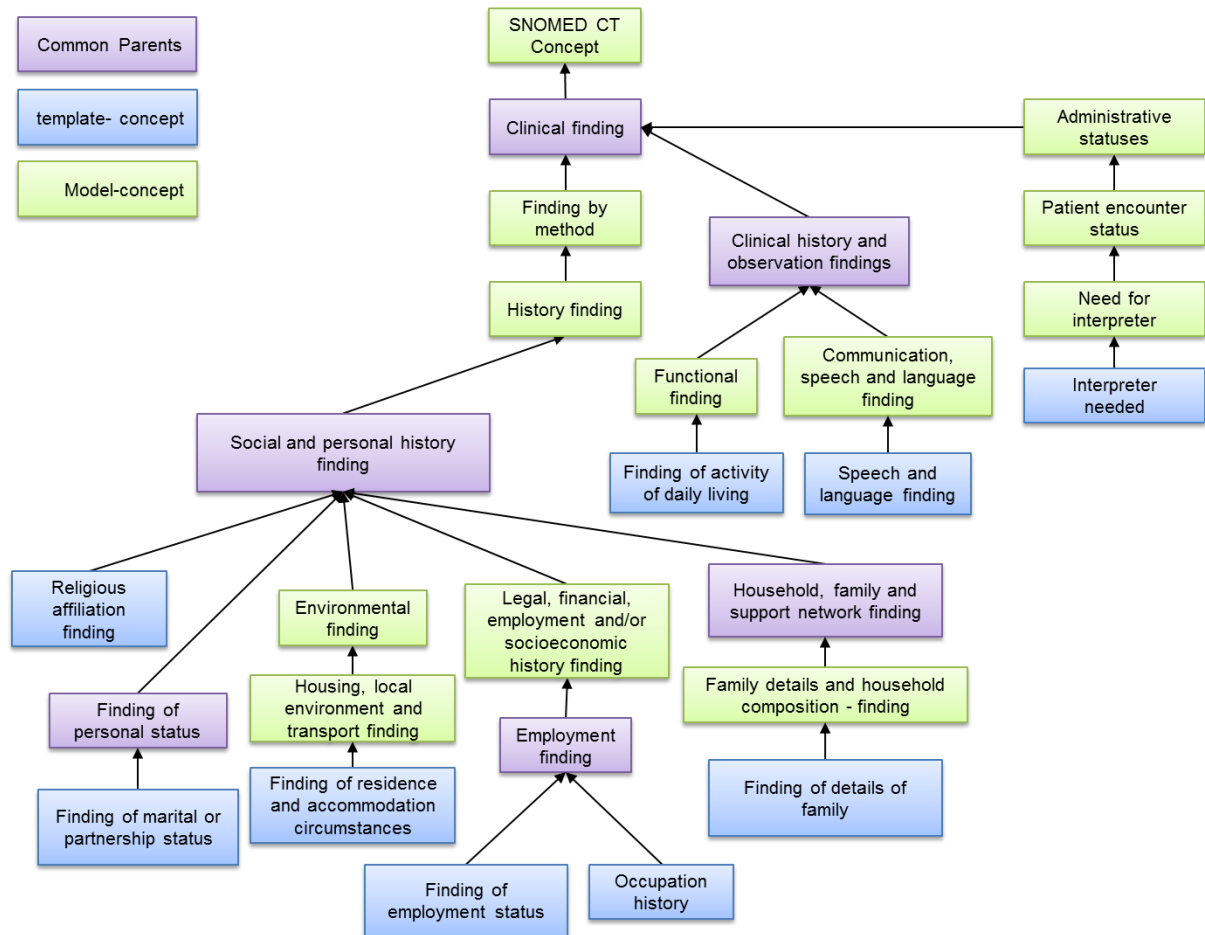
Consistent concept selection means that the same clinical meaning is always represented in the same way. This consistency ensures data retrieval and comparison on a single concept level and that data can be retrieved and compared based on a specific conceptID or ID's.

Supporting comparability based on the structure provided by SNOMED CT, such as the subtype relationships demands that the concepts also be coherent. Coherency means that concepts which are semantically related also are mapped to concepts which are related by their definition in SNOMED CT. Or, concepts that are fine granularity can be retrieved as a subtype of more coarse grained concepts.

We found that hierarchical coherency between the concepts is a pragmatic way of supporting comparability between sets of concepts. For example, using concepts from the same subtype hierarchy to represent content used in a specific clinical content, such as documenting the results of a physical examination. Hierarchical coherency is useful for supporting data comparability, especially, when the existing system does not support services that allows for querying SNOMED CT represented data by use of the defining characteristics of a concept.

In the SNOMED CT implementation project in Region NJ manual visualizations of the hierarchical coherence between the concepts of the mapping result were developed, see Figure 23.

Figure 23. Example of manual visualization of the mapping result developed in the study. This illustration shows the result of mapping to the clinical content of the nursing status template.

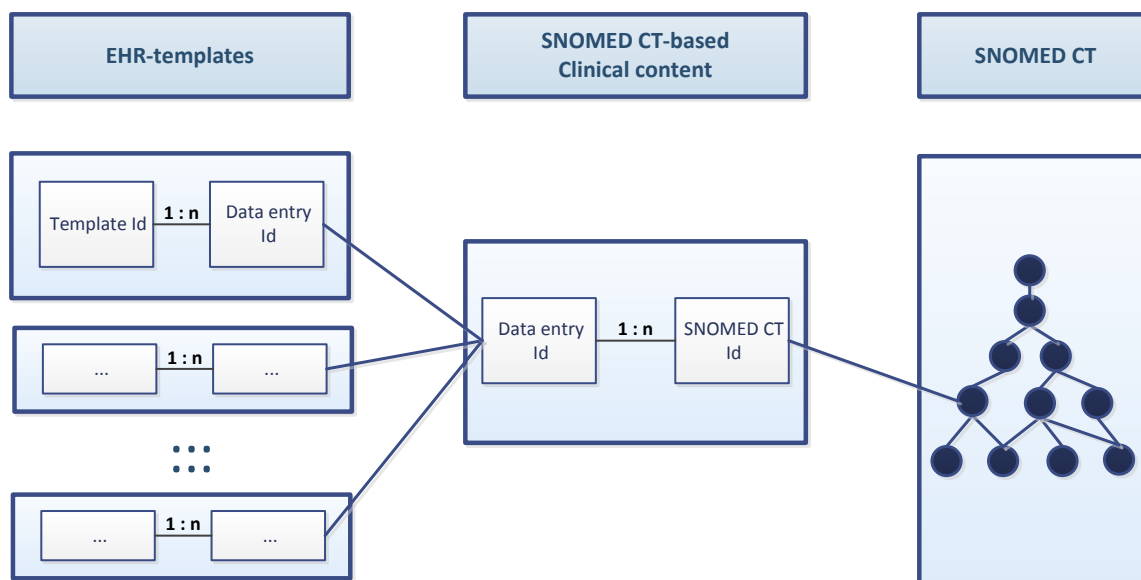


These visualizations were useful for discussing the mapping outcome and emphasizing the potential for reusability. The results of the study in Region NJ showed that SNOMED CT enabled content management has at least two purposes:

- To support consistency and coherency between the concepts
- To bridge the terminology with the record structure allowing for comparing content across system borders and data structures.

A minimum requirement for such a system is that it should at minimum relate each point of data entry to a SNOMED CT identifier, as illustrated in Figure 24.

Figure 24 Minimum requirements for SNOMED CT based content management.



For every point of data entry belonging to an EHR template a single, specified SNOMED CT concept or a specific set of concepts should be assigned to support consistent data entry within that data entry field. By linking the SNOMED CT concept ID's to the specific data entry points, it will be possible to query and use equivalent concepts across EHR templates. Content management that fulfills the above rather simple description can have great impact on the reusability of data across systems and templates. This supports an overview of what concepts are applied in the system and also specifies where the exact concepts are being used. It supports developing queries that use data from different EHR templates, to develop clinical overviews, aggregations etc. Content management also makes it is possible to compare where equivalent or similar expressions are being used (e.g. concepts with equivalent attribute relationships or subtypes of a certain concept).

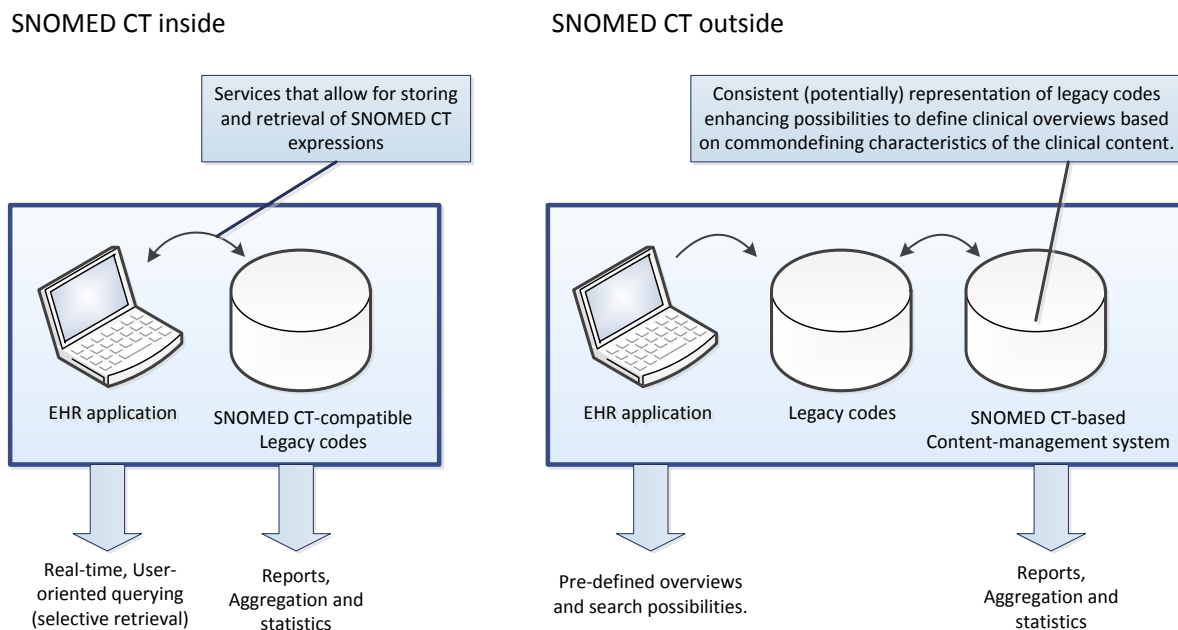
Implications of technological incompatibility

Services supporting terminological reasoning are a prerequisite for utilizing SNOMED CT and underpin advanced querying. Whether such services is incorporated in the CISs or as an interface (e.g. by using a content management system as described above) it is basically not pertinent for achieving a semantically deliberated representation of the clinical content in a CIS. However, the degree to which SNOMED CT can be utilized at the point of care and for secondary purposes depends on how SNOMED CT is implemented.

Implementing SNOMED CT at the interface (SNOMED CT outside) like was done in the Region NJ project, brings value by supporting semantic consistency between the EHR templates, and allows for querying data based on related SNOMED CT concepts (e.g. for creating overviews). However, incorporating SNOMED CT into the EHR

is preferable, as it will enhance the possibility of using SNOMED CT for advanced real time search functionalities and creating user oriented clinical overviews (SNOMED CT inside). The different strategies for SNOMED CT implementation, i.e. SNOMED CT inside vs. SNOMED CT outside is illustrated in Figure 25.

Figure 25. SNOMED CT can be used by either having "SNOMED CT inside" vs. "SNOMED CT outside". "SNOMED CT inside" enables the best potential for user oriented data entry and retrieval.



In Denmark, the applied CISs do not incorporate SNOMED CT inside and therefore it is relevant to explore methods for enabling SNOMED CT outside, as this for many cases will be the first step for SNOMED CT implementation.

Implications of local and central Governance

In both focus group discussions it became evident that the national infrastructure and governance has great impact on the course of directions on the regional strategy for use of terminology and influence the position and initiatives of the vendors.

The regional representatives described how most of the content within their system is developed locally in cooperation with clinicians and vendors, and state that the process of defining content and interface terms is a very extensive task demanding for overview and insight to both the technical framework and the clinical documentation practice. Generally, the CIS applied in Denmark today supports the use of ICD10, NCSP, ICF and other classification systems required for reporting. In addition, self-invented legacy codes are used to represent

the information which is not covered by the classifications or not required for reporting to national registers. One participant explained that the primary motivation for applying local terminology, compared to applying for example SNOMED CT, is that this approach quickly fulfills the basic and immediate documentation needs.

A summary of findings of the empirical analysis is presented in the following, along with a summary of the findings of the theoretical analysis.

Summary of findings of the theoretical and empirical analysis

Summary of findings on the theoretical foundation for semantic comparability using SNOMED CT

SNOMED CT supports concept comparison at different levels. The unique concept IDs, the subtype hierarchies and the logical definition of the semantics of each concept (and each expression) can be utilized to facilitate concept comparison.

<u>Id-comparison:</u>	Comparing the concept ID's of two or more concepts enable direct comparison of two concepts.
<u>Hierarchical comparison:</u>	Traversing the subtype hierarchies support hierarchical comparison of concepts, i.e. determination of the hierarchical relation between two or more concepts.
<u>Meaning-based comparison:</u>	The logical definition of a concept or post-coordinated expression enables comparison of the defining characteristics.

Id-comparison is done by comparing the concept-Ids of two or more concepts. When representing clinical data or interface terms by SNOMED CT concepts, this type of comparison can be used to determine whether instances of data or interface terms are represented by the same SNOMED CT concept.

Hierarchical comparison is done using the defined [is a]-relationships in SNOMED CT. These relationships can be used to determine the parent-children relationship of two or more concepts, and determine common supertypes of two or more concepts. The [is a]-relationship is the core relationship type in SNOMED CT, and all concepts in SNOMED CT has, at least, one [is-a]- relationship two a more general concept.

Meaning-based comparison is done by comparing the defining characteristics of two or more concepts. This means utilizing the poly-hierarchical structure of SNOMED CT and the defined relationships to compare concepts by their defining characteristics. Meaning-based comparison enables selective retrieval of concepts with common defined characteristics. However, this type of comparison requires methods enabling terminological reasoning of pre- and post-coordinated expressions and requires strategies for managing primitive concepts.

Summary of findings of Empirical study of semantic comparability using SNOMED CT

Two overall prerequisites for supporting comparability in SNOMED CT implementations have been derived from the empirical analysis.

- Consistent concept selection.
- Coherency between related concepts.

Both factors must be facilitated in order to support comparability based on the structure of SNOMED CT. Consistent concept selection means that the same clinical meaning is represented through the same SNOMED CT expression every time that clinical meaning occurs in the CISs, i.e. if the same clinical expression occurs in different clinical templates then it should be mapped to the same concept or expression. Coherency means that clinical expressions which are clinically or semantically related should be mapped to concepts which are related through direct or inherited relations in SNOMED CT. Consistency supports comparability on a single concept basis whereas coherency allows for comparing concepts based on their composition in the SNOMED CT hierarchies.

Beside these prerequisites for ensuring comparability of data, various factors were identified challenging in the SNOMED CT implementation process. The following challenges must be overcome in order to implement SNOMED CT and facilitate comparability of data in local implementations:

- The people responsible for system configuration and/or who are responsible for selecting concepts for data entry must have SNOMED CT competencies supporting correct interpretation of the meaning of a concept.
- Clear mapping strategies should be developed, supporting both consistent and coherent selection of concepts and to limit coding variability.
- Content management providing an overview of what concepts are applied and where in the CIS they are used is necessary for supporting data reuse and comparability across borders.
- Realistic expectations of the level of benefits SNOMED CT can induce given the actual technological capabilities are necessary, as SNOMED CT inside vs. outside has implications for how data can be retrieved and used within the CIS.
- Governance is an important factor for successful adoption of SNOMED CT nationally as well as locally. From a semantic interoperability viewpoint, comparing clinical data across organizations requires national strategies. As well, a local setting requires consistency and coherency to be able to reuse and compare information within a local CIS, so the same factors are equally important when wanting to reuse clinical data across organizations, i.e. across hospitals.

Results (part 2)

Development of tools for supporting semantic comparability when selecting SCT concepts

In this research two studies were conducted on basis of the results of the analysis and the knowledge obtained on prerequisites for facilitating data comparability. Hence, 'mapping guidelines' were developed along with a tool for visualizing sets of concepts in order to support hierarchical coherency and thereby comparability. Both tools were developed to support coherent and consistent selection of concepts which will ensure the data is comparable. The following sections summaries the developed mapping guidelines and the prototype for visualizing sets of SNOMED CT concepts.

Mapping guidelines

Following chapter is derived from the enclosed paper (Randorff Højten, Rosenbeck Goeg 2012), see appendix Paper B

Even though the research community continuously adds to the improvement of SNOMED CT, reported use of SNOMED CT for practical purposes is limited [19]. One possibility for practitioners could be to wait until the structure and content of SNOMED CT is complete. However, completeness is seldom cost effective, since much of the terminology would never be used in practice [20]. Therefore, the objective of this study is to provide a clear methodology for SNOMED CT mapping to enhance applicability of SNOMED CT despite incompleteness and redundancy. Redundancy can be reduced by applying a set of rules that clarifies which SNOMED CT term to choose among a set of candidate mappings. Earlier studies have shown that consistency can be improved through mapping guidelines [1, 21]. In this study guidelines are designed to make sense in a multi-organizational and multi-specialty context and to allow clinically meaningful querying for primary and secondary purposes.

Quality criteria for concept selection

The first quality criteria specifies that the | IS A | relationships are to be the only relationship type used to assess whether two concepts are terminologically related. For example, two concepts from two different subtype hierarchies can be terminologically related because they are defined by e.g. the same body structure. In theory these relationships can also be used to support comparability between mapped concepts. However, there are at least two reasons for restricting to the | IS A | relationships.

- 3) The | IS A | relationships are present for every concept in SNOMED CT, whether the concepts are fully defined or primitive.
- 4) EHR templates will often consist of clinical terms which can be thought of as being the same 'type', e.g. pick lists will typically allow for selection between a group of diagnoses, a group of body structures or substances etc. A list will seldom consist of a mix of these types. Therefore, it logically makes sense to use this relationship as basis for the mapping, as the | IS A | relations represent the sub-type relations.

The second quality criteria is about consistency across borders. This criteria is closely related to the first criteria, and to the content management aspect described in chapter "Content management". To facilitate data reusability across borders, it requires cohesion between data collected on every side of a border.

The tools developed in this study are designed for use in a multi-organizational and multi-specialty context and to allow clinically meaningful querying for primary and secondary purposes. Being aware of the importance of consistency in the selection of hierarchies and concepts facilitates the possibility of introducing fields for more granular information within the same clinical domain. Selecting concepts that are inherited by the coarse grained information will solely enlarge the sample space for a given query without being dependent on any query modification. Therefore, this approach supports the bridging between the coarse grained expressions applied in most of the general EHR templates and the fine grained expressions that are seen in the specialty specific templates.

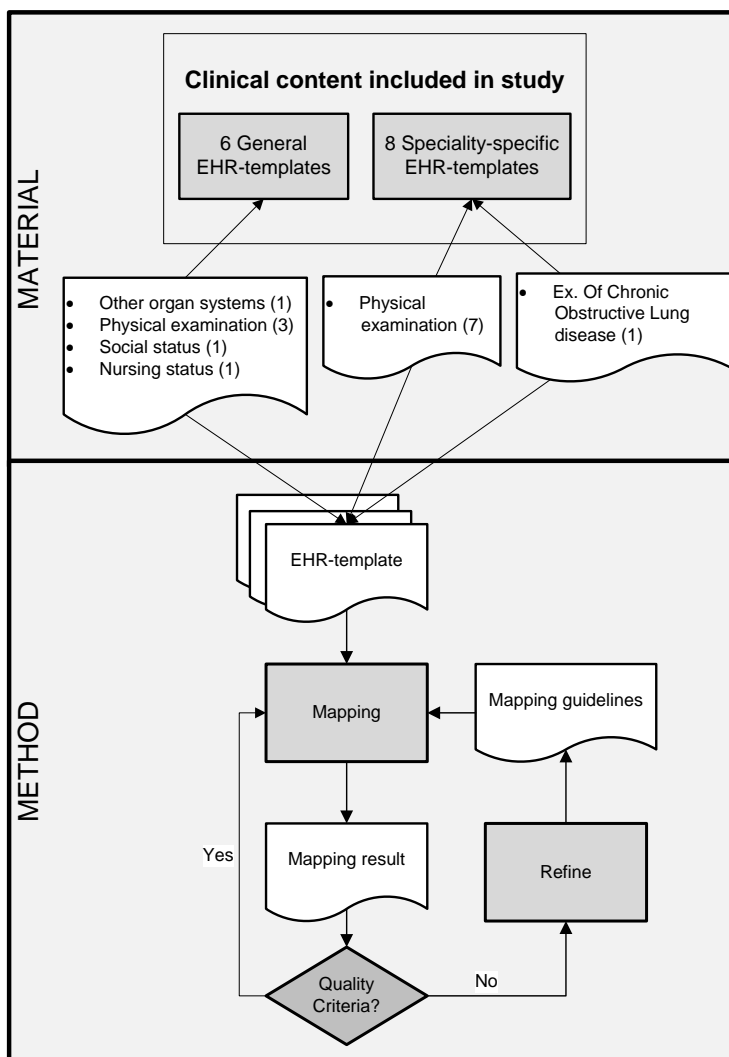
In summary, the quality criteria are:

- Meaningful relationships between SNOMED CT concepts must be ensured. Given the issues of SNOMED CT, meaningful relationships are | IS A | relationships, as these are always present.
- Consistent mapping should be ensured within and across organizations.

Material and method

In order to obtain a multi-organizational and multi-specialty focus, the guidelines presented are based on an in depth analysis of 14 different Electronic Health Record (EHR) templates from three Danish and two Swedish EHR system implementations. The EHR templates, included in this study, represent a variety of granularity, as both general and specialty specific information is represented. Thus, the guidelines can be adapted to different types of information and different levels of granularity. Also, the variety allows investigation of how mapping on different granularity levels can be kept consistent. Figure 26 illustrates the material used and mapping method developed and applied in this study.

Figure 26 Material and method applied for developing mapping guidelines facilitating reuse of data.



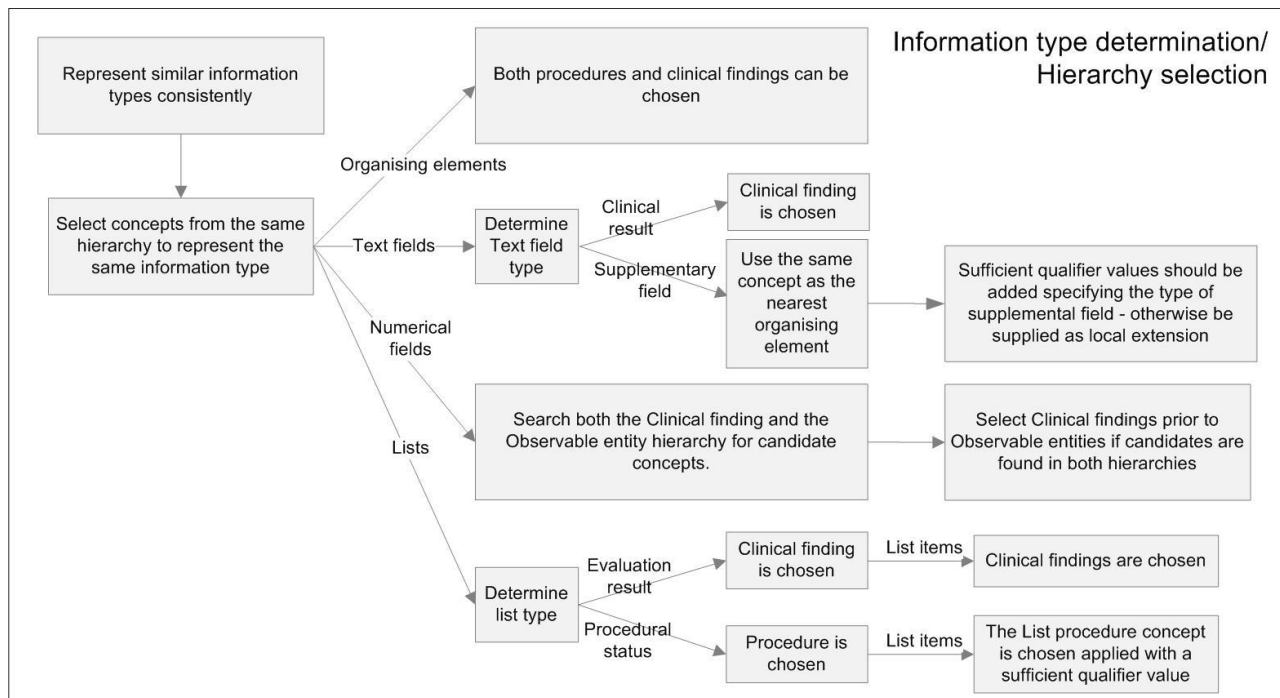
An iterative development process was conducted with the quality criteria set up as the overall mapping objective, see Figure 26. The guidelines were used to map the interface terms of EHR templates to SNOMED CT. The mapping result was assessed using the quality criteria. Problems were identified and possible solutions formulated as new guidelines replaced or supplemented the original guidelines. The refined guidelines were used to edit earlier mappings and continue with mapping the interface terms of more EHR templates. This continued until all 14 templates were mapped and fulfilled the quality criteria.

Hierarchy selection

As derived from the analysis, approaching a SNOMED CT search aimed at selecting a concept for data entry requires the competency to assess the implications of selecting a specific concept. In (Randorff Rasmussen, Rosenbeck 2011) we have documented that redundancy issues occur between SNOMED CT hierarchies, and we suggest that 'content coverage', 'granularity' and 'definition' might be useful parameters in determining which hierarchy reaps most benefits in terms of retrieval and reuse purposes. From the development of the mapping guidelines we also experienced that the information type of the clinical template can also support selection of the appropriate hierarchy. Even though each of the SNOMED CT top level hierarchies are aimed at representing different types of information, e.g. procedures, findings, body structures, determining what hierarchy a specific entry field or interface term in a clinical template should be represented by is not always straightforward. Clinical templates often consist of a mixture of information types and information levels, e.g. some information can be process oriented and some information represents evaluation results, some represent highly structured information and others are low structured information. The mapping guidelines specify how to assess the information type in order to select the appropriate hierarchy. In Figure 27, the resulting mapping guidelines for hierarchy selection based on type of information are illustrated.

Types of information: The highest level of information is 'organizing elements' and 'result elements'. The title of an EHR template, such as 'Physical examination' or 'Nursing status' is considered an 'organizing element'. However, an EHR template may consist of multiple section headings which are also organizing elements e.g. 'smoking behavior' or 'cardiovascular stress testing'. Result elements can be subdivided into text fields, numerical fields and lists. They specify result fields where the underlying SNOMED CT concept should represent the meaning of the interface term. Lists can refer to both drop down menus, check boxes and radio buttons.

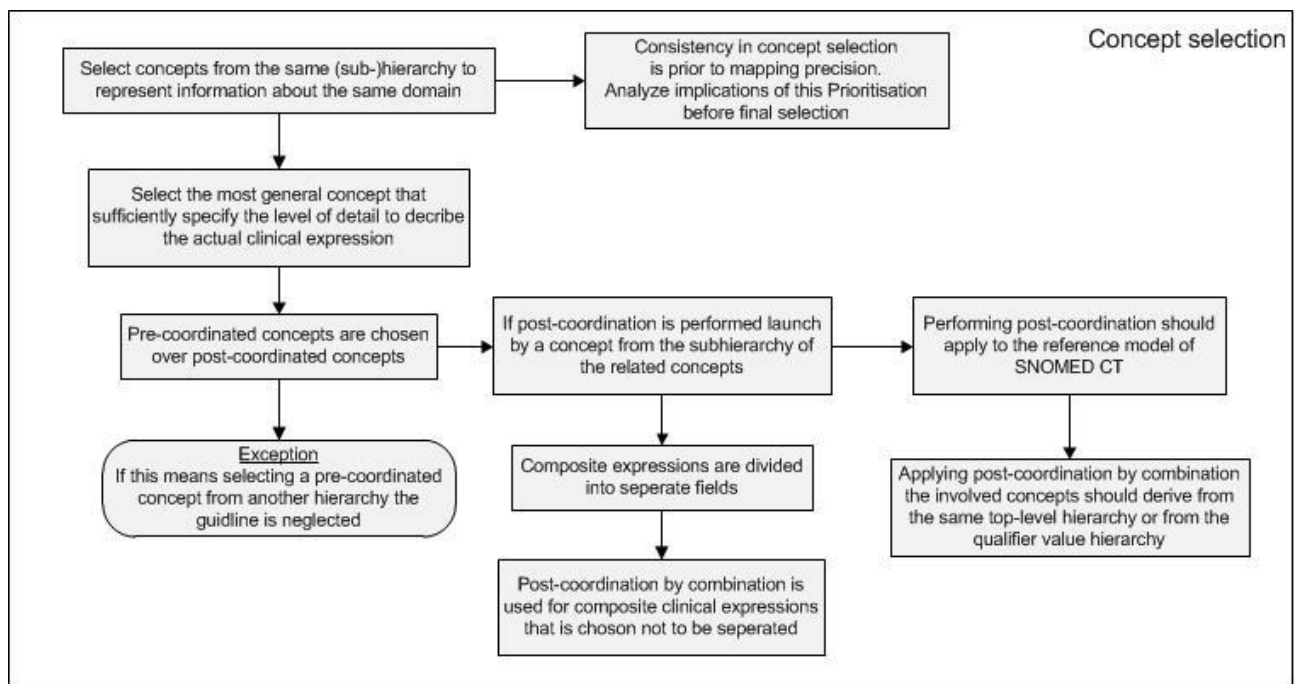
Figure 27 Illustration of the mapping guidelines related to the type of information of the clinical expression. Through the iterative approach, we selected concepts from as few hierarchies as possible to represent the interface terms.



Concept selection

After determining which hierarchy the specific clinical expression should be represented by, selecting the appropriate concept(s) can in some cases be a matter of balancing comparability vs. specificity, and in this project comparability is chosen prior to specificity. Figure 28 summarizes the guidelines specified for selecting SNOMED CT concepts and balancing use of pre- and post-coordination.

Figure 28 Mapping guidelines supporting concept selection and consistent use of post-coordination.



Generally, concept selection should be done in order to represent the exact meaning of the specific clinical expression. However, as also stated it is important to look beyond the specific clinical term and assess the actual meaning of the expression and assess the context of use and the related information.

One of the forces of SNOMED CT is its compositional structure that allows for post-coordination, which gives the possibility to build clinical meanings which are not represented as pre-coordinated expressions. In this case it also supports mapping consistency which enables data comparability.

Visualizing sets of concepts

Following chapter is derived from the enclosed paper, see appendix "Paper C"

Tool for visualizing sets of concepts

When using SNOMED CT browsers to search for specific clinical concepts, the browsers typically provide information about the defining characteristics and possibilities for qualification, refinement etc. In Rogers et. al (Rogers, Bodenreider 2008), a review of 17 existing SNOMED CT browsers was conducted. The study showed diversity in the implemented features, and requested increased research on what specific search features are important for specific use cases. One of the groups of features presented in the study is visualization and navigation, and the study show that most browsers support displaying search result in some sort of expandable tree view. However, it is not documented that any of the browsers apply visualization of more than a single concept, which could support consistent concept selection.

Through the empirical study of SNOMED CT implementation in Region NJ and the mapping guidelines development, described on page 37 and page 50, we got experience with browsing SNOMED CT when developing SNOMED CT based clinical content and with analyzing concepts to determine their retrieval potential. We found that visualization can provide an overview of concepts and their hierarchical relations, and show the coherency between the concepts used to represent the SNOMED CT enabled clinical content.

The following sections provide a summary of the terminological features and the visualization framework used in the prototype development, and also discusses the potential of visualizing sets of concepts.

Visualization framework

Developing a visualization tool for sets of SNOMED CT concepts requires a layout that supports visualization of graphs of very different sizes and complexity. With visualization it is important to consider the pitfalls that are entailed by the size and complexity of the graphs. (Rafiei 2005) describe challenges related to graph visualization "Visualizing "large" networks, however, can be quite challenging if not impossible. This is due to the limitations of the screen, the complexity of layout algorithms and the limitations of human visual perception."(Rafiei 2005)

The graphical tool used in e.g. the Cliniclue browser is based on a static layout. However this study explore an interactive approach to terminology vizualisation. The TermViz.js implementation is based on the versatile D3.js² (Bostock, Ogievetsky & Heer 2011) JavaScript framework that contains support for many types of graphs including force-based ones. This is possible because modern standards compliant web browsers now have the capacity to efficiently handle graphs drawn in Scalable Vector Graphics³ (SVG) backed by physics-based simulations implemented in JavaScript.

Terminological features

Three terminological features are defined as requirements for the visualization tool. Each of these features provide information useful for selecting concepts which meet the quality criteria set up in page 63.

- **Least common parents**

The LCPs are useful determining the level of which it is possible to compare two or more concepts. From a subtype hierarchy perspective the LCPs can be thought of as the least common denominators of the concerned concepts. LCPs are therefore indicative for the level of comparability of the clinical data based on the sub hierarchical view only.

- **Common parents and non common parents**

A graphical overview is useful when visualizing common structures of concepts such as common and non common parents. As many concepts have a very large number of supertype ancestors the graph will often be very large and with many nodes. Removing non common parents will provide a less complex graph and enhance the possibility to assess the common paths of a set of concepts.

- **Concept children**

Information about the children of a concept is useful for:

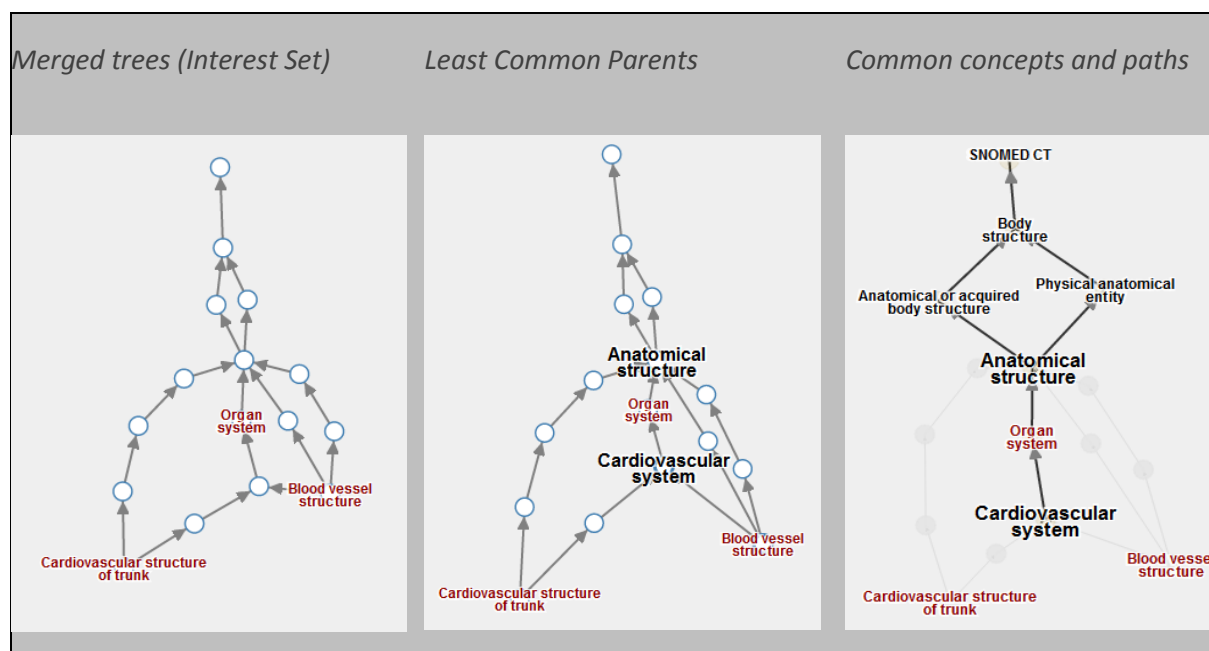
- Clarifying the semantic interpretation of a concept, by explaining information about which subtypes are related to the concept.
- If the specific concept represents the appropriate level of granularity or if a more refined concept should be used.

Figure 29 illustrates how the first merged graphs, the least common parents and the common concepts are visualized in the prototype.

² *Data-Driven Documents* . Available: <http://d3js.org/>.

³ *Scalable Vector Graphics (SVG)*. Available: <http://www.w3.org/Graphics/SVG/>.

Figure 29 illustrates the graphical view of an Interest Set consisting of three Interest Concepts (marked red in the graphs). Figure A shows the merged tree and the highlighted Interest Concepts. Figure B shows the Least Common Parents, and figure C shows all common concepts and links in the graph. Node labels are revealed when the mouse cursor hovers over them.



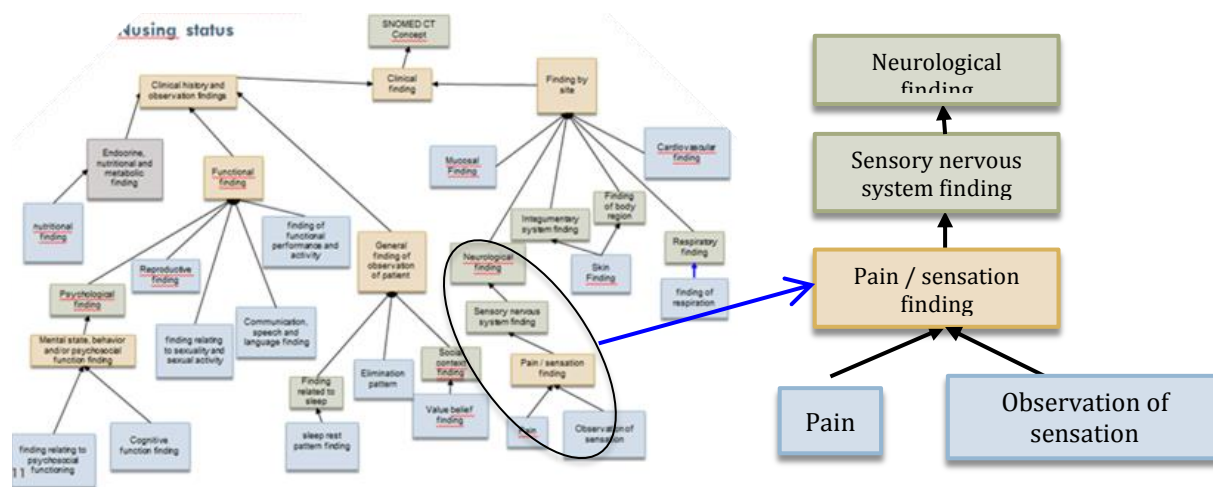
Potential of visualization of sets of concepts

As described in the enclosed papers (Randorff Højen, Rosenbeck Goeg 2012, Randorff Rasmussen, Rosenbeck 2011), we experienced that visualization supported overview of the sets of SNOMED CT concepts, useful for supporting consistent and coherent mapping. Also, an evaluation of the visualization tool, by 16 Danish terminology experts, showed various use cases for such tool. At least three direct potentials for such visualization is presented.

Overview of concepts and their relations

Visualizing concepts and their supertype ancestors provides a graphic overview of the concepts of interest and their hierarchical interrelationships. In contrast, a list of concepts represented by the concept descriptions is much less informative, as illustrated on the left side of Figure 30.

Figure 31 Illustration of how two concepts can be compared by the level of their 'least common parent'.



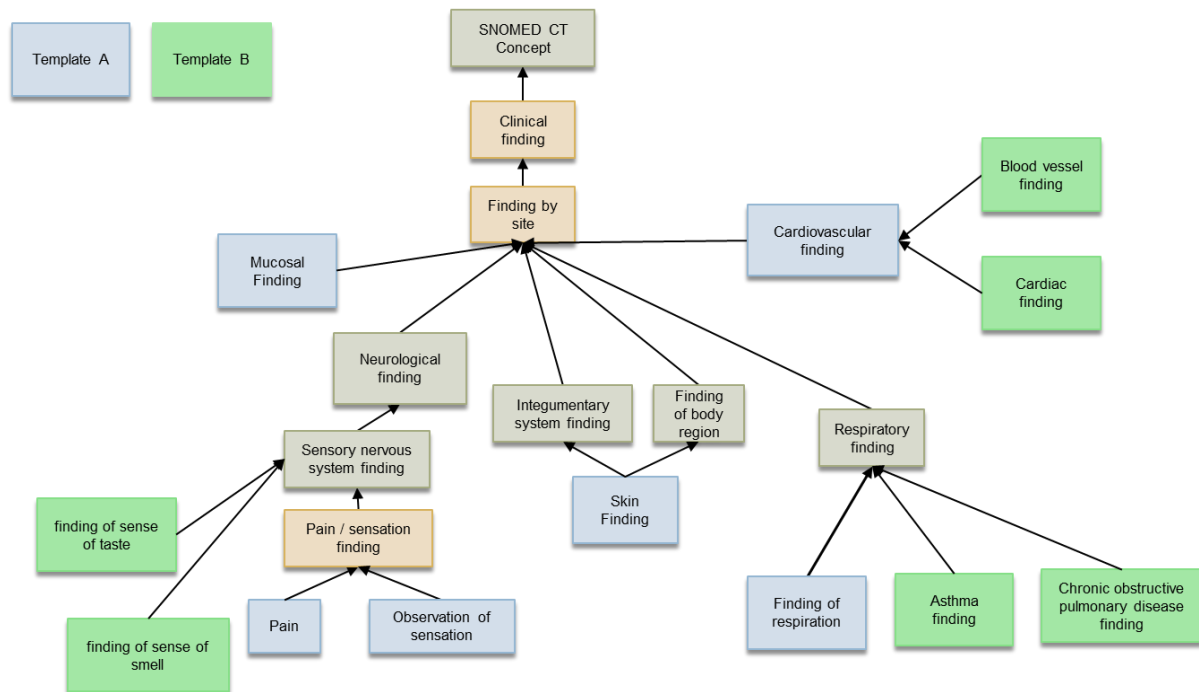
This way of visualizing common paths for different sets of SNOMED CT concepts can support definition of queries that retrieve concepts included in various clinical templates and provide a pragmatic way of utilizing the structure of SNOMED CT without being dependent on any retrieval functionalities.

Compare groups of concepts

To strengthen the link between clinical data entry templates and terminology further, using multiple simultaneous Interest Sets as input could be implemented as a feature. This would provide the possibility to compare concepts from different clinical templates with each other and would further make it possible to analyze on information shared across clinical templates. However, working with multiple Interest Sets will increase the graph size and complexity, and thus, future work should strive to improve visualization of large graphs, for example by applying clustering techniques (Herman, Melançon & Marshall 2000).

Figure 32 gives an example of how comparison of different concept sets could be implemented. The blue boxes illustrate concepts used in one template and the green boxes illustrate concepts from another template.

Figure 32 Illustration concepts from two different templates.



The graphical overview shows the coherence between the content in the two templates and supports reuse of data across these. For example, if we want to retrieve data recorded about a patients *Neurological findings* the illustration shows what concepts in the two templates are subtypes of this, namely *Pain*, *Observation of sensation*, *Finding of sense of taste*, *Finding of taste of smell*. We therefore see a potential use of applying visualization to support content management and query design.

Summary of development of tools supporting semantic comparability when selecting SCT concepts

Summary of findings related to mapping guidelines

Development of mapping guidelines has resulted in a methodology to obtain consistency and meaningful relationships in SNOMED CT mapping processes. The heterogeneous material applied in this study ensures that the mapping guidelines do not solely conform to a single EHR system or a single organization. The variety of the material, characterized by the different EHR systems and the different clinical specialties show that common guidelines can be developed to facilitate semantic unambiguousness, despite differences in IT solutions and clinical practices.

Even though, a common terminology system can support reusability of clinical information, it is necessary to be rigorous in the selection of the concepts from the terminology. Otherwise, redundancy or unsuitable hierarchy selection can challenge the possibility to request uniform information. The prerequisite of the quality criteria, is to apply SNOMED CT in accordance with the refined guidelines, considering the type of information and the balance between consistency and precision.

The overall mapping rule is to represent related information homogeneously by selecting concepts from the same sub-hierarchy. The guidelines provide a framework for achieving a consistent mapping procedure and thereby a well-defined foundation for data retrieval. This approach is practical for comparison of clinical information based on the hierarchical structure in SNOMED CT. Also, it will not limit the use of selective retrieval methods, as proposed by Dolin et al. [17] when sufficient definitions are added to SNOMED CT. Future work should strive at qualifying and validating the specified guidelines by applying them to other domains and EHR systems.

Summary of findings related to visualization tool

The visualisation tool visualise sets of concepts and their interrelations related to the hierarchical composition in SNOMED CT. The empirical study in Region NJ showed that visualizations were useful for discussing the mapping outcome and emphasizing the potential for reusability.

Visualization of complex graphs can be supported by specific features to support overview and interpretation of SNOMED CT concepts and sub-hierarchies. Four terminological features were implemented in the tool to support overview, comparison and interpretation of SNOMED CT concepts, i.e. Merged graphs, Least common parents, Common and non-common parents and concept children. These terminological features support concept selection according to the developed mapping guidelines and the specified quality criteria.

Evaluation of the tool show that visualisation of sets of SNOMED CT concepts has the potential to support overview of concepts and their relations, comparison of single concepts and comparison of sets of concepts. This potential can support different purposes, such as search and training, consistent selection of SNOMED CT concepts and content management.

The TermViz.js toolkit based on D3.js is released as an open source toolkit aimed at terminology visualization and can be used in different setups for different purposes. This makes it possible for others to further develop and integrate tools to enhance terminology exploration and usage. Further tuning and development of the force based layout is needed in order to improve the user experience when browsing SNOMED CT, especially when interactively adding more nodes to an already crowded graph. In future work, evaluation of the explored potentials of applying this tool will be an important priority.

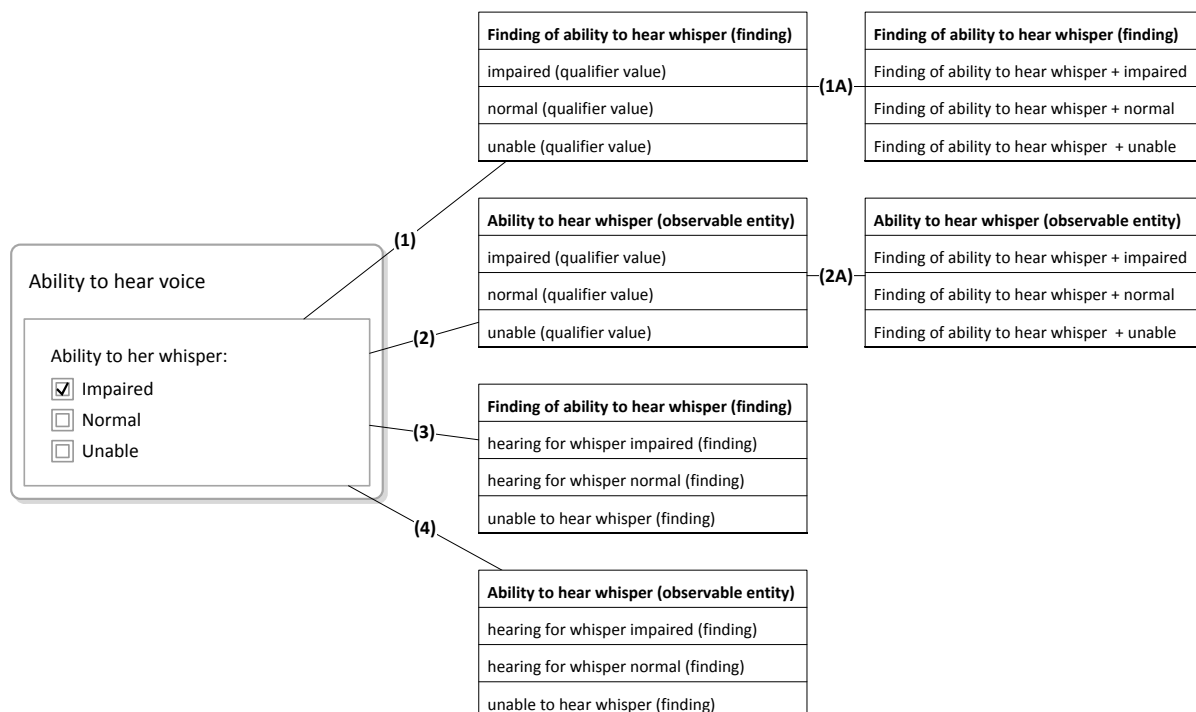
Discussion

Terminology binding

The importance and the necessity of binding terminology and information models in order to enable semantic interoperability is described in various studies (Qamar, Kola & Rector 2007, Sundvall et al. 2008, Markwell, Sato & Cheetham 2008). The mapping guidelines were developed using EHR-templates from 14 different EHR-systems, and developed to support consistent use of terminology across organizations and specialties. The mapping guidelines are not directed towards a specific information model; however the information type and user interface structures govern the choice of hierarchy. This approach is taken to enable semantic interoperability by integration of clinical data – not integration of clinical information systems.

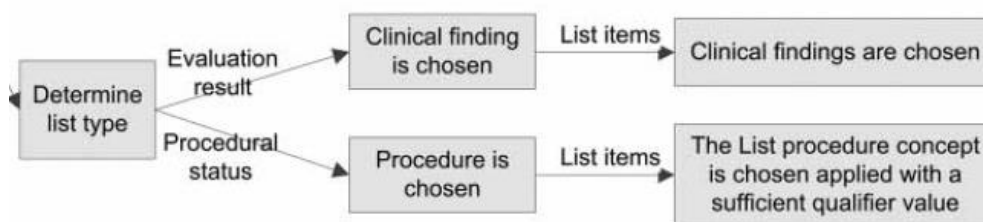
The guideline on how to represent list structures address the grey area between terminology and Information Model, which has been discussed in a previous study by Markwell et al. (Markwell, Sato & Cheetham 2008). For example, Figure 33 shows four different ways of representing a simple list, and hence represents the choices that must be taken for every list representation in a CIS.

Figure 33 Four examples of how equivalent clinical meaning may be stored in a CIS database. To ease interpretation this example uses the FSN instead of conceptID's.



The mapping guidelines recommend specific ways to represent the two types of check lists derived from the material of the study, see Figure 34. One list type represents evaluation results and the other list type represents a procedural status. In both cases data retrieval is supported because of consistent representation across the different EHR templates. Moreover, the semantic representation supports interpretation independently of how data is captured, and thus supports reusability.

Figure 34 Mapping guidelines on representation of lists.



The two list types defined in the mapping guidelines relates to the two views on how to represent and interpret information captured by check lists described in (Markwell, Sato & Cheetham 2008):

1. *Representing the information as captured.*
2. *Representing the information independently of the way in which it was captured.*

In Figure 33 the first two examples (1 and 2) shows how the SNOMED CT representation of the three check boxes represent the information as captured using qualifier values 'impaired', 'normal' and 'unable'. Hence, to interpret the actual meaning of these values requires knowledge on the context, and therefore these values must be linked to the label *Ability to hear whisper* which can be represented using either the observable entity "Ability to hear whisper" (1A in Figure 33) or the clinical finding "Finding of ability to hear whisper" (2A in Figure 33).

Example 3 and 4 in Figure 33 shows how the values of the check boxes can be represented independent of the way it was captured, as the semantics of the concepts chosen represent both the clinical meaning and the actual result value.

The recommendations given for the list items representing a specific status of a procedure, were that these items are represented by the post-coordinated expression of the specific procedural concept combined with a qualifier value descriptive for the status of the procedure. This recommendation conflicts with the logical foundation of SNOMED CT, as it can induce axis modification, meaning that the semantics of the concept is changed so that it cannot be regarded as an instance of its supertype concept, as for this example of a list representing statuses of a specific procedure:

Inhaler technique shown (Procedure)

- Inhaler technique shown (regime/therapy) + Not done (qualifier)
- Inhaler technique shown (regime/therapy) + Done (qualifier)

From an ontological perspective the instantiation of the expression |Inhaler technique shown (regime/therapy)| implies that the action “showing the inhalator technique” has actually occurred, whereas the qualifier |Not done| implies the opposite – namely that the action has not been performed. This complicates retrieval of actual occurrences of |Inhaler technique shown (regime/therapy)|, as this will require preclusion of occurrences where |Not done| is attached.

Ontological problems, as exemplified above, have been a matter of investigation in recent studies. As presented by (Schulz, Cornet 2009, Schulz, Cornet & Spackman 2011) SNOMED CT has unclear and diverse ontological commitments, which can lead to misinterpretations when implemented in CIS. I agree with the perspective that solutions are needed allowing that *“one concept C1 can be referred to in the definition of a concept C1 without stating that for each instance of C1 there must be an instance of C2.”*(Schulz, Cornet & Spackman 2011, Schulz, Cornet 2009)

Until such solutions are made, and the ontological commitment of SNOMED CT is changed to support such situations, it is necessary that implementation projects are consistent in their way of representing different information types using SNOMED CT, and explicitly state how to interpret each mapped expression. This is necessary to support concept comparison across implementations, but also to support future harmonization or altering according to changes in the ontological framework.

The developed mapping guidelines can serve as explicit reference, supporting interpretation of the mapped expressions, and therefore they have the potential to support future research and implementation projects addressing terminology binding. Future research should validate the adaptability of the mapping guidelines with respect to specific information models. Most likely, the guidelines must be refined to cover a broader scope of information types to ensure a common mapping strategy for each type. As consistency is important at the level of concept selection, a similar approach is important when addressing representation using specific information models.

Observable entities vs. clinical findings

In paper A, we explore what implications selection of observable entities vs. clinical findings have for data retrieval when representing the clinical content from two clinical templates through SNOMED CT concepts. This study questions the applicability of the observable entity hierarchy because the concepts within this hierarchy are poorly defined and currently lack quality and comprehensiveness. These shortcomings are the reason why the developed mapping guidelines recommend the use of clinical findings over observable entities, even though IHTSDO recommend the use of observable entities for “representing a question or procedure which can produce

an answer or a result.”⁴ Moreover, concepts from the clinical findings hierarchy are intended to represent the actual “result of a clinical observation, assessment or judgement”⁵. The attribute |interprets| is used to relate clinical findings and observable entities, for example the concept | finding of muscle tone (finding) | is related to | muscle tone (observable entity) |, as shown below:

Finding of muscle tone (finding) ≡
Muscle finding (finding) ⊓ ∃ finding_site.muscle structure (body structure)
 ⊓ ∃ **interprets**.Muscle tone (observable entity)

However, the clinical finding hierarchy includes many “abstract” clinical findings in addition to actual clinical results. The “abstract” clinical findings groups more specific concepts with common defining properties. As shown in Figure 35, the concept | finding of muscle tone | serves as a grouper for instances of muscle tone findings, and in a clinical documentation situation it is the descendants of these “abstract” clinical findings that will be the used to document the clinical result.

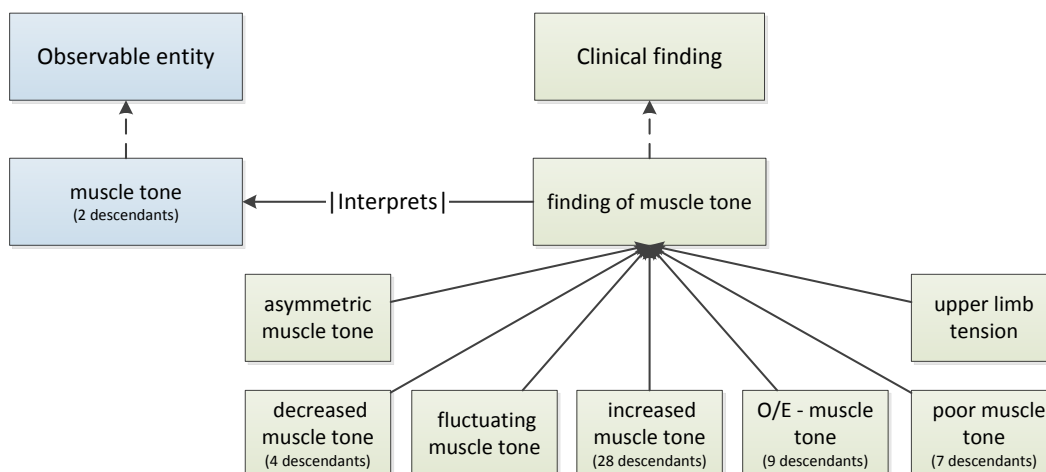


Figure 35 Illustration of the relationship between observable entities and clinical findings.

In our experience there is a distinct correlation between the “abstract” clinical findings and the observable entities, and our studies (paper A and B) show that the “abstract” clinical findings cover the needed information. Moreover, “abstract” clinical findings provide better retrieval properties than observable entities. Further studying the overlap between the observable entity hierarchy and the clinical findings hierarchy would be interesting. However,

⁴ www.snomed.org/eg?t=observables

⁵ www.snomed.org/eg?t=hier_clin_find

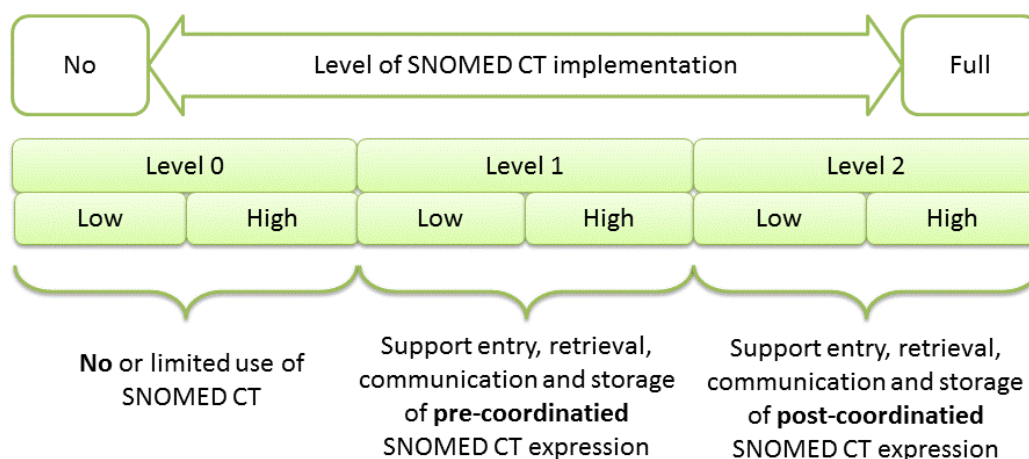
IHTSDO is currently working on changing the concept model for observable entities, and consequently, the value of such study is questionable in the current version of SNOMED CT.

Post-coordination aspects

In this research focus on supporting use of post-coordination is limited to the overall mapping rules. Post-coordination has been widely considered as one of the major strengths of SNOMED CT, as it induces flexibility and broadness of the terminology. (Cornet 2009, Rector, Iannone 2012)

For every application of SNOMED CT, and for every standard in general, there are different possibilities for implementation which are influenced by a range of factors such as the technical framework, scope of use, goals, implementation approach etc. IHTSDO distinguishes between Level 0, Level 1 and Level 2 implementations, as illustrated in Figure 36 (IHTSDO 2012-07-31b). The main difference between Level 1 and Level 2 implementation is the support of either pre- or post-coordinated expressions. In each case, the difference between the high and the low level is whether the implementation supports a fixed set of pre or post-coordinated expressions (low level), or if the implementation supports the full terminology (high level).

Figure 36 Implementation levels for SNOMED CT enabled applications.



Earlier studies describe how retrieval and comparison of both pre- and post-coordinated expressions are possible via the normal form representation of concept expressions (Dolin, Spackman & Markwell 2002, Spackman et al. 2002). As described in the section about “.

Meaning-based comparison” on page 42 comparison of concept expressions can be achieved by terminological reasoning of the defining characteristics of expressions. The services required for such comparison will be equivalent for both pre- and post-coordinated expressions.

Currently, there are few documented studies, on how to integrate post-coordinated expressions into a CIS. The reason for this could be the noticeable increased complexity between the management of pre-coordinated expressions compared to post-coordinated expressions, with respect to data entry as well as storage and services.

IHTSDO specifies that ‘*SNOMED CT enabled applications must support the storage of SNOMED CT expressions in ways that represent relevant information within the record system.*’ (IHTSDO 2010) This research clarifies that the EHR systems in use in Denmark do not support SNOMED CT adequately, as they are limited to represent information based on a single identifier, and therefore only allow for representing SNOMED CT expressions in its simplest form, i.e. pre-coordinated expressions.

Lee et. al. (Lee et al. 2012) describes different challenges related to use of post-coordination:

Implementing post-coordination continues to be a challenge both from a graphical user interface design and clinical terminology point of view (e.g., creating clinically nonsensical concepts, concept duplication and inefficiency of concept composition).

Furthermore, in (Rector, Iannone 2012) post-coordination is assessed through a lexical analysis of terms compared with semantic analyses of definitions, on the expressions “acute” and “chronic”. This study states that serious pitfalls occur when applying post-coordination by use of these qualifiers, and showed that 25% contained irregularities that could cause errors in post-coordination.

Future research should therefore be aimed at SNOMED CT implementation using concept expressions, e.g. addressing implementation and use of expression repositories (IHTSDO 2012-07-31b). This topic will be a step further towards level two implementations and an important part of enabling selective retrieval based on attribute relationships rather than limiting SNOMED CT implementation to using the sub-hierarchical structure defined by SNOMED CT, as proposed in this study.

The visualization tool developed as part of this study is limited to visualizing the subtype hierarchy, as this aligns with the proposed mapping guidelines. Visualization has great potential for supporting concept selection and ensures coherent use of the terminology. In future research the visualization tool could be further developed to visualize common and non-common attribute relationships. Also, visualization of post-coordinated expression is a topic yet unexplored but relevant from an implementation perspective.

Another potential which should be explored is how to use visualization to support education and training.

The difficulty of SNOMED CT evaluations

As the main result of this PhD project, the mapping guidelines and the visualization tool has been developed, aimed at supporting concepts selection and hence support SNOMED CT implementation and reusability of data. However, the fact that these tools have been developed does not demonstrate the impact of them. Empirical evidence for the added value of mapping guidelines and visualization is beyond the scope of this PhD project, and will be objectives for further research.

For the mapping guidelines, interrater variability measures between coding experts, similar to the methods applied in (Andrews, Richesson & Krischer 2007, Patrick et al. 2008), could be conducted to test whether the mapping guidelines improves mapping consistency between coders, compared to a situation where the guidelines are not applied. Prior to such an evaluation it should be explored how to ensure a comparable level of SNOMED CT related training among coders.

However, improving consistency does not demonstrate the added value, in terms of improved clinical information for researchers, administrators and clinicians which ultimately is intended to improve patient care. Terminology evaluation requires tight integration with complex information systems. This, however, induces a fundamental evaluation problem, since it is difficult to distinguish the success or failure of the information system from the success or failure of the terminology. (Rogers 2006) Therefore, adding to the knowledge on evaluation methodology is an important future priority.

Conclusion

Towards consistent and coherent SNOMED CT adoption

Denmark has a well-established set of procedures for reporting to national registers using a defined set of statistical classifications (ICD-10, ICF, NCSP). These procedures underpin national statistics and epidemiological research. The reuse of data is currently supported for secondary purposes. At the same time, locally defined interface terminologies or other terminologies are used to support clinical documentation. The “collect once, use many” vision for reuse of data (Cimino 2007, Barton et al. 2011, Elkin et al. 2010) is therefore not realized, because the practice could be summed up by the phrase “Register twice, use separately”.

The heterogeneous use of clinical terminology in CIS complicates the reuse of data for creating user-oriented overviews for clinical purposes to support patient treatment, and for sharing and comparing data across systems and organizational borders. Hence, this heterogeneity inhibits the realization of semantic interoperability.

SNOMED CT supports the documentation and use of data. It enables flexible and user oriented data entry and enables reusing data for multiple purposes. Despite this potential documentation addressing actual use and implementation of SNOMED CT is limited. For example, demonstrations are missing of benefits of applying SNOMED CT. One pitfall addressed in the scientific literature, has been that application of SNOMED CT can induce ambiguous and inconsistent representation of equivalent clinical meaning (Andrews, Richesson & Krischer 2007, Spackman 2001, Andrews et al. 2008b, Patrick et al. 2008), and this is a core barrier for data comparability.

In theory, and ideally, the flexible structure of SNOMED CT along with the clear rules of description logic, can obviate controlled and constrained use of SNOMED CT. However, when theory meets practice, reality becomes different. Existing CISs typically do not support expression storage, i.e. entry and retrieval of both pre- and post-coordinated expressions, along with terminological reasoning is not possible within the CIS, supporting real time search and retrieval functionalities. Currently, SNOMED CT is typically being implemented by use of pre-coordinated Concept Id's, while services supporting terminological reasoning are not applied. Adding to this discussion we see a dilemma which challenges bridging theory and practice. On one hand the concept model of SNOMED CT allows compensation for missing concepts (either intentionally or because of terminological gaps), through post-coordination. But on the other hand, the applied CISs do not interoperate with or conform to SNOMED CT to an extent which allows for utilizing the compositional and poly-hierarchical structure.(Lee et al. 2012) Because of this, the full theoretical potential of SNOMED CT is currently not realized in practical implementations.

This PhD study contributes to the knowledge about how to configure SNOMED CT enabled application software, utilizing the structure of SNOMED CT to facilitate comparability of data. Despite the technological foundation of the CISs, the quality criteria, the mapping guidelines and the visualization tool can contribute to adoption of SNOMED CT and potentially improving data quality, in terms of increased consistency and semantically coherent representation of the clinical content. An increased focus must be on utilization of SNOMED CT in the process of concept selection, and we believe that utilization of SNOMED CT requires deliberation about how data can be used afterwards. Therefore, concept selection must not be done blindly or without reflection about its intended use.

The results of this study show that comparability of clinical data is not only supported by assuring consistent concept selection. To utilize the structure of SNOMED CT, coherency between the selected concepts is important to meet support comparability of data based on the semantic meaning of concepts. The quality criteria for mapping, specified in this study, are prerequisites for ensuring comparability in SNOMED CT implementations:

- **Consistent concept selection must be assured**, i.e. common clinical expressions should be represented by the same pre-coordinated expression. Alternatively, applying services that support expression comparison, consistency should be ensured so that equivalent clinical expressions are represented by the same logically expressed meaning.
- **Coherency between related clinical expressions must be ensured**, i.e. clinical expressions which are semantically related in a way, so that they should serve as result of a specific query, should be mapped to concepts or expressions which are hierarchically related.

Further studies to elaborate and refine these guidelines in order to address aspects which are not covered by the material included in this study are recommended. The developed visualization tool supports a set of terminological features that can be used to meet above stated prerequisites for comparable use of SNOMED CT. Limited to visualizing subhierarchies and pre-coordinated concepts, further research should refine the visualization framework and validate the potential use cases for visualization.

Finally, future research must investigate how to utilize the compositional structure of SNOMED CT, both with respect to adjusting existing CISs using SNOMED CT outside, but also with respect to develop applications which use SNOMED CT inside and also support Level Two implementations.

References

- Aguilar, A. 2005, "Semantic Interoperability in the context of eHealth", *Research Seminar, DERI Galway, December 15th*Citeseer, .
- Andrews, J.E., Patrick, T.B., Richesson, R.L., Brown, H. & Krischer, J.P. 2008a, "Comparing heterogeneous SNOMED CT coding of clinical research concepts by examining normalized expressions", *Journal of Biomedical Informatics*, vol. 41, no. 6, pp. 1062-1069.
- Andrews, J.E., Richesson, R.L. & Krischer, J. 2007, "Variation of SNOMED CT coding of clinical research concepts among coding experts", *Journal of the American Medical Informatics Association*, vol. 14, no. 4, pp. 497-506.
- Andrews, J.E., Patrick, T.B., Richesson, R.L., Brown, H. & Krischer, J.P. 2008b, "Comparing heterogeneous SNOMED CT coding of clinical research concepts by examining normalized expressions", *Journal of Biomedical Informatics*, vol. 41, no. 6, pp. 1062-1069.
- Baader, F. & Nutt, W. 2003, "Basic description logics.", *Description logic handbook*, pp. 43.
- Barton, C., Kalle, C., Van Dyke, P., Mon, D. & Richesson, R. 2011, "Demonstrating "Collect once, Use Many"–Assimilating Public Health Secondary Data Use Requirements into an Existing Domain Analysis Model", *AMIA Annual Symposium Proceedings*American Medical Informatics Association, , pp. 98.
- Bodenreider, O. 2008, "Biomedical ontologies in action: role in knowledge management, data integration and decision support", *Yearb Med Inform*, vol. 47, pp. 67-79.
- Bostock, M., Ogievetsky, V. & Heer, J. 2011, "D³ Data-Driven Documents", *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 12, pp. 2301-2309.
- Chiang, M.F., Hwang, J.C., Yu, A.C., Casper, D.S., Cimino, J.J. & Starren, J. 2006, "Reliability of SNOMED-CT coding by three physicians using two terminology browsers", *AMIA Annual Symposium Proceedings*American Medical Informatics Association, , pp. 131.
- Cimino, J.J. 2007, "Collect Once, Use Many: Enabling the Reuse of Clinical Data through Controlled Terminologies", *JOURNAL-AHIMA*, vol. 78, no. 2, pp. 24.
- Cimino, J.J. 1998, "Desiderata for controlled medical vocabularies in the twenty-first century", *Methods of information in medicine*, vol. 37, no. 4-5, pp. 394.
- Cimino, J. 2011, "High-quality, standard, controlled healthcare terminologies come of age", *Methods of information in medicine*, vol. 50, no. 2, pp. 101.
- Cornet, R. 2009, "Definitions and Qualifiers in SNOMED CT", *Methods Inf.Med.*, , pp. 178.

- Cornet, R. & de Keizer, N. 2008, "Forty years of SNOMED: a literature review", *BMC medical informatics and decision making*, vol. 8 Suppl 1, pp. S2.
- de Keizer, N.F., Abu-Hanna, A. & Zwetsloot-Schonk, J. 2000, "Understanding terminological systems I: Terminology and typology", *Methods of information in medicine*, vol. 39, no. 1, pp. 16-21.
- Dolin, R.H. & Alschuler, L. 2011, "Approaching semantic interoperability in Health Level Seven", *Journal of the American Medical Informatics Association*, vol. 18, no. 1, pp. 99-103.
- Dolin, R.H., Spackman, K.A. & Markwell, D. 2002, "Selective retrieval of pre- and post-coordinated SNOMED concepts", *Proceedings / AMIA ...Annual Symposium.AMIA Symposium*, , pp. 210-214.
- Elkin, P.L., Trusko, B.E., Koppel, R., Speroff, T., Mohrer, D., Sakji, S., Gurewitz, I., Tuttle, M. & Brown, S.H. 2010, "Secondary use of clinical data.", *Studies in health technology and informatics*, vol. 155, pp. 14.
- European Commission 2008, "Commission Recommendation on cross-border interoperability of electronic health record systems Official Journal of the European Union", *L*, vol. 190, pp. 37-43.
- Greenhalgh, T. & Peacock, R. 2005, "Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources", *BMJ: British Medical Journal*, vol. 331, no. 7524, pp. 1064.
- Herman, I., Melançon, G. & Marshall, M.S. 2000, "Graph visualization and navigation in information visualization: A survey", *Visualization and Computer Graphics, IEEE Transactions on*, vol. 6, no. 1, pp. 24-43.
- Hevner, A.R., Salvatore, T.M., Park, J. & Ram, S. 2004, "Design Science in information Systems Research", *MIS Quarterly*, vol. 28.
- IHTSDO 2012-07-31a, *SNOMED Clinical Terms. Editorial Guide*.
- IHTSDO 2012-07-31b, *SNOMED Clinical Terms. Technical Implementation Guide*.
- IHTSDO 2010, *SNOMED Clinical Terms. Technical Reference Guide*.
- IHTSDO , *International Health Terminology Standards Development Organisation* [Homepage of ihtsdo], [Online]. Available: <http://www.ihtsdo.org/> [2013, 09/12].
- Ingenerf, J. & Giere, W. 1998, "Concept-oriented Standardisation and Statistics-oriented Classification: Continuing the Classification versus Nomenclature Controversy", *Methods Inf.Med.*, , no. 37, pp. 527.
- Johnson, S.B., Bakken, S. & Dine, D. 2010, "An Electronic Health Record Based on Structured Narrative", *Journal of the American Medical Informatics Association : JAMIA*, , pp. 54.
- Lee, D., Cornet, R., Lau, F. & de Keizer, N. 2012, "A Survey of SNOMED CT Implementations", *Journal of Biomedical Informatics*, .

- Lee, D., de Keizer, N., Lau, F. & Cornet, R. 2013, "Literature review of SNOMED CT use", *Journal of the American Medical Informatics Association*, .
- Lopez, D.M. & Blobel, B.G. 2009, "A development framework for semantically interoperable health information systems", *International journal of medical informatics*, vol. 78, no. 2, pp. 83-103.
- Markwell, D., Sato, L. & Cheetham, E. 2008, "Representing clinical information using SNOMED Clinical Terms with different structural information models", *Proceedings of KR-MED 2008*, .
- Pape-Haugaard, L., Rasmussen, A.R., Andersen, S.K. & Elberg, P.E. "Conflicts between terminology and EHR information models as obstacles to semantic interoperability: a scientific review", .
- Patrick, T.B., Richesson, R., Andrews, J.E. & Folk, L.C. 2008, "SNOMED CT Coding Variation and Grouping for "other findings" in a Longitudinal Study on Urea Cycle Disorders", *AMIA Annual Symposium Proceedings*American Medical Informatics Association, , pp. 11.
- Qamar, R., Kola, J.S. & Rector, A.L. 2007, "Unambiguous data modeling to ensure higher accuracy term binding to clinical terminologies", *AMIA Annual Symposium Proceedings*American Medical Informatics Association, , pp. 608.
- Rafiei, D. 2005, "Effectively visualizing large networks through sampling", *Visualization, 2005. VIS 05. IEEE*IEEE, , pp. 375.
- Randorff Højen, A. & Rosenbeck Goeg, K. 2012, "SNOMED CT Implementation. Mapping Guidelines Facilitating Reuse of Data", *Methods of information in medicine*, vol. 51, no. 6.
- Randorff Rasmussen, A. & Rosenbeck, K. 2011, "SNOMED CT implementation: Implications of choosing Clinical findings or Observable entities", .
- Rector, A.L., Qamar, R. & Marley, T. 2009, "Binding ontologies and coding systems to electronic health records and messages", *Applied Ontology*, vol. 4, no. 1, pp. 51-69.
- Rector, A. & Iannone, L. 2012, "Lexically suggest, logically define: Quality assurance of the use of qualifiers and expected results of post-coordination in SNOMED CT", *Journal of Biomedical Informatics*, vol. 45, no. 2, pp. 199-209.
- Rector, A.L. 1999, "Clinical terminology: why is it so hard?", *Methods of information in medicine*, vol. 38, no. 4-5, pp. 239-252.
- Rogers, J. & Bodenreider, O. 2008, "SNOMED CT: Browsing the browsers", *KR-MED 2008*, , pp. 30-36.
- Rogers, J. 2006, "Quality assurance of medical ontologies", *Methods of information in medicine*, vol. 45, no. 3, pp. 267.
- Rosenbeck, K.H., Rasmussen, A.R., Elberg, P.B. & Andersen, S.K. 2010, "Balancing centralised and decentralised EHR approaches to manage standardisation", *Studies in health technology and informatics*, vol. 160, no. Pt 1, pp. 151.

- Rosenbloom, S.T., Brown, S.H., Froehling, D., Bauer, B.A., Wahner-Roedler, D.L., Gregg, W.M. & Elkin, P.L. 2009, "Using SNOMED CT to represent two interface terminologies", *Journal of the American Medical Informatics Association*, vol. 16, no. 1, pp. 81-88.
- Schulz, S. & Cornet, R. 2009, "SNOMED CT's Ontological commitment", Citeseer, .
- Schulz, S., Cornet, R. & Spackman, K. 2011, "Consolidating SNOMED CT's ontological commitment", *Applied Ontology*, vol. 6, no. 1, pp. 1-11.
- So, E. & Park, H. 2011, "Exploring the possibility of information sharing between the medical and nursing domains by mapping medical records to SNOMED CT and ICNP", *Healthcare informatics research*, vol. 17, no. 3, pp. 156-161.
- Spackman, K.A., Campbell, K.E. & CÄ, R. 1997, "SNOMED RT: a reference terminology for health care.", *Proceedings of the AMIA annual fall symposium*American Medical Informatics Association, , pp. 640.
- Spackman, K.A., Dionne, R., Mays, E. & Weis, J. 2002, "Role grouping as an extension to the description logic of Ontylog, motivated by concept modeling in SNOMED.", *Proceedings of the AMIA Symposium*American Medical Informatics Association, , pp. 712.
- Spackman, K.A. 2001, "Normal Forms for Description Logic Expressions of Clinical Concepts in SNOMED RT", *AMIA.Annu.Symp.Proc.*, .
- Statens Serums Institut 21st june 2013, , *SNOMED CT på dansk*. Available: <http://www.ssi.dk/Sundhedsdataogit/National%20Sundheds-it/TerminologiOgKlassifikationer/SNOmedCT/SNOmedCTpaaDansk.aspx> [2013, July 1st].
- Stroetmann, V.N., Kalra, D., Lewalle, P., Rector, A.L., Rodrigues, J.M., Stroetmann, K.A., Surjan, G., Ustun, B., Virtanen, M. & Zanstra, P.E. 2009, "Semantic Interoperability for Better Health and Safer Healthcare", .
- Sundvall, E., Qamar, R., Nyström, M., Forss, M., Petersson, H., Karlsson, D., Åhlfeldt, H. & Rector, A. 2008, "Integration of tools for binding archetypes to SNOMED CT", *BMC Medical Informatics and Decision Making*, vol. 8, no. Suppl 1, pp. S7.
- Wade, G. & Rosenbloom, S.T. 2008, "Experiences mapping a legacy interface terminology to SNOMED CT.", *BMC medical informatics and decision making*, vol. 8.
- Wasserman, H. & Wang, J. 2003, "An applied evaluation of SNOMED CT as a clinical vocabulary for the computerized diagnosis and problem list", *AMIA Annual Symposium Proceedings*American Medical Informatics Association, , pp. 699.

Paper A

SNOMED CT implementation: Implications of choosing Clinical findings or Observable entities

Anne Randorff Højen^a, Kirstine Rosenbeck Gøeg^a

^a Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

Published in: Stud Health Technol Inform. 2011;169:809-13. PMID: 21893859 [PubMed - indexed for MEDLINE].

A part of this thesis has been removed from this publication due to
copyright.

Paper B

SNOMED CT implementation: Mapping guidelines facilitating reuse of data

Anne Randorff Højen^a, Kirstine Rosenbeck Gøeg^a

¹ Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

Published in: *Methods Inf Med*. 2012 Dec 4;51(6):529-38. doi: 10.3414/ME11-02-0023. Epub 2012 Oct 1. PMID: 23038162.

A part of this thesis has been removed from this publication due to
copyright.

Paper C

Methods and applications for visualization of SNOMED CT concept sets.

Anne Randorff Højen^a, Kirstine Rosenbeck Gøeg^a, Erik Sundvall^b

^a Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

^b Department of Biomedical Engineering, Linköping University, Sweden

Submittet for: Applied Clinical Informatics

A part of this thesis has been removed from this publication due to
copyright.

Paper D

SNOMED CT in Denmark. Why is it so hard?

Anne Randorff Højen^a, Pia Britt Elberg^a and Stig Kjær Andersen^a

¹ Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

Submitted to: Medical Informatics Europe (MIE) 2014 conference.

A part of this thesis has been removed from this publication due to
copyright.