Medical Informatics Group

Bridging Centralized and Decentralized Clinical Modelling through Semantic Comparison Methods

PhD Thesis by
Kirstine Rosenbeck Gøeg

River Publishers

Bridging Centralized and Decentralized Clinical Modelling through Semantic Comparison Methods

Bridging Centralized and Decentralized Clinical Modelling through Semantic Comparison Methods

PhD Thesis by

Kirstine Rosenbeck Gøeg

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

ISBN 978-87-93237-56-8 (Ebook)

Published, sold and distributed by:
River Publishers
Niels Jernes Vej 10
9220 Aalborg Ø
Denmark

Tel.: +45369953197
www.riverpublishers.com

Mandatory page including thesis publications

Thesis title
Bridging centralised and de-centralised clinical modelling through semantic comparison methods

Name of PhD student
Kirstine Rosenbeck Gøeg

Name and title of supervisor
Associate professor, Stig Kjær Andersen

Thesis publications

Paper I: Rosenbeck KH, Rasmussen AR, Elberg PB & Andersen SK (2010). *Balancing centralised and decentralised EHR approaches to manage standardisation*. Studies in health technology and informatics, 160(Pt 1), 151.

Paper II: Højen AR & Gøeg KR (2011). *SNOMED CT Implementation. Mapping guidelines facilitating reuse of data*. Methods of Information in Medicine, 50(5), 472-478.

Paper III: Karlsson D, Gøeg KR, Örman H, Højen AR (2014). *Semantic Krippendorffs Alpha for measuring interrater variabillity in SNOMED CT coding studies*. MIE2014. Accepted for publication.

Paper IV: Gøeg KR, Chen R, Højen AR & Elberg PB (2014). *Content analysis of physical examination templates in electronic health records using SNOMED CT*. International Journal of Medical Informatics. In Press . DOI: 10.1016

Paper V: Gøeg KR, Cornet R, Andersen SK (). *Clustering local electronic health record content based on semantic similarity*. Journal of Biomedical Informatics. Submitted August 2014

Related publications by the author

Papers

Højen AR, Sundvall E, Gøeg KR (2014). *Methods and applications for visualization of SNOMED CT concept sets*. Applied Clinical Informatics 2014;5(1):127-152

Gøeg KR, Elberg PB (2012). *Global applicability of a local physical examination template. Scandinavian Conference on Health Informatics*, SHI2012, 2-3 October 2012, Linköping, Sweden. Linköping University Electronic Press, 2012. p. 1-7 (Linköping Electronic Conference Proceedings; No. 70).

Rasmussen AR & Rosenbeck K (2011). *SNOMED CT implementation: implications of choosing clinical findings or observable entities*. Stud Health Technol Inform, 169, 809-13.

Rosenbeck KH & Rasmussen AR (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 AR, Rosenbeck KH & Andersen SK (2010). *Pitfalls when integrating terminology systems and EHRs.* In Scandinavian Conference on Health Informatics. Tapir Akademisk Forlag 2010, p. 34-38.

Abstracts

Lilholt L, Haubro CD, Møller JM, Aarøe J, Højen AR, Gøeg KR (2013). *Developing an acute-physical-examination template for a regional EHR system aimed at improving inexperienced physician's documentation.* MEDINFO 2013: Proceedings of the 14th World Congress on Medical and Health Informatics. ed. / Christoph Ulrich Lehmann; Elske Ammenwerth; Christian Nøhr. IOS Press, 2013. p. 1129 (Studies in Health Technology and Informatics; No. 192).

Højen AR, Gøeg KR, Elberg PB (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).

Rosenbeck KH, Rasmussen AR, & Elberg PB (2009). *Bridging Danish EHR systems by sharing, reusing and standardising Clinical Content.* SHI2009 Proceedings: 7th Scandinavian Conference on Health Informatics, 24-26 August 2009, Arendal, Norway. ed. / Rune Fensli; Jan Gunnar Dale. TAPIR Akademisk Forlag, 2009. p. 56.

Acknowledgements

A PhD project depends on so much more than the PhD student. I could not have traveled all this way alone, and I owe thanks to contributors, supervisors, colleagues, co-authors, collaborators, family and friends.

To begin a PhD project you need a good idea, and people who believes enough in the idea to support it financially and academically. I would like to express my gratitude to CSC Scandihealth and Region Northern Jutland for financial support and to my supervisor Stig Kjær Andersen for agreeing to supervise my work and helping me with my employment at Aalborg University. You have been a valuable resource and you know the university better than your back pocket.

I am grateful for colleagues with open doors regardless of whether my current interest has been MatLab code, bootstrapping, report layout, the precise meaning of “ectopic heart”, teaching methods or company in the lunch breaks. The Medical Informatics group has an open culture that I have enjoyed to be a part of. Medical Informatics is a broad research field, and therefore it has been valuable to be part of the subgroup for Terminology and Models in Clinical information Systems. Thank you for our nerdy discussions focused on Electronic Health Records, information organization, interoperability and SNOMED CT. Special thanks should go to Pia Elberg -- co-supervisor, mentor and friend. You have a methodological way of thinking and a strategic talent out of the ordinary. You are the Gandalf to my Frodo. Anne Randorff Højen, without you my PhD project would have been something completely else. Not many PhD-students have the opportunity to truly be part of a team. I am grateful that I am one of them. I foresee that team Randorff-Rosenbeck has yet to accomplish their greatest achievements.

I have had the pleasure to work with international experts within my research field. My thanks go to Rong Chen, Daniel Karlsson and Ronald Cornet. Developing research ideas, conducting studies and writing papers with you have been a great learning experience for me. International collaboration is not only about writing papers. I have visited University College London, and Linköping University in the beginning of my PhD project. I would like to thank the the research groups for welcoming me, and participating in discussions that brought me valuable insight. I also joined the Student Editorial Board of Methods of Information in Medicine, and my thanks go to the editors that gave me insight on the other side of the review process.

Studying local clinical models have only been possible because generous organizations in Denmark and Sweden have shared their clinical models with us. I would like to thank the EHR units in Östergötland Län, Uppsala University Hospital, Odense University Hospital, Randers Regional Hospital and Region Northern Jutland.

Last but not least, I would like to thank my family and friends for their love and support. Especially, thanks go to my husband Pelle who kept our family together when I had to work long hours. Alfred and Maren, nothing ever compares to hearing you laugh.

Kirstine Rosenbeck Gøeg

Abstract

[Introduction] Electronic Health Record (EHR) systems are a much anticipated solution to improve flow of information in health care, but current development in EHR-system research is fragmented. Studies have shown that in a de-centralized setting innovative systems can be developed with close user involvement, but scaling the solutions with its local storage models, clinical models and terminology has proven difficult. Inability to scale results in patient information locked in local systems with limited ability to support interoperability. Taking a centralised viewpoint, researchers and standardisation organisations have developed semantic interoperability frameworks based on standardised reference information models, clinical models and terminologies for EHRs, but the progress towards semantic interoperability is slow.

[Objectives] The aim of this PhD-thesis has been to bridge de-centralised and centralised clinical models by using semantic methods for content comparison has been. The development of comparison methods has been directed by two research questions: Q1: What are the prerequisites of applying semantic methods for local clinical model comparison? Q2: Which semantic approaches can be applied to local clinical model comparison? The questions have been addressed in the five enclosed papers.

[Q1: Paper I,II and III] Harmonisation was a prerequisite for comparison of local clinical models because the models were expressed using vendor-dependent formalisms and local interface terminology. In paper I, II and III artefacts were developed and evaluated that help ensure a consistent representation of local clinical models both with respect to model formalism and terminological representation using SNOMED CT. The focus of paper I was to develop a model formalism based on an analysis of Danish clinical models. The focus of paper II was formulation of a set of guidelines for using SNOMED CT to express the interface terminology of clinical models. The guidelines were developed iteratively while the two authors mapped 14 Swedish and Danish local clinical models to SNOMED CT. The focus of paper III, was to propose a semantic inter-rater agreement measure and evaluating it against a nominal measure using a Swedish set of procedures mapped to SNOMED CT by two coders.

[Q2: Paper IV and V] Two methods for semantic clinical model comparison was developed and evaluated in Paper IV and V respectively. Both methods used a consistent representation of content drawn from paper I and paper II. The focus of paper IV was to analyse four clinical models for physical examination from four different organisations in Sweden and Denmark. The content analysis was based on pair-wise analysis of number of full matches, terminology matches and non-matches between clinical models. In addition, a comparison with textbook material and a common content analysis was performed. The focus of paper V was to give an overview of multiple clinical models using hierarchical clustering. This required a quantification of similarity and automation compared to paper IV. Different quantification techniques were compared using a combination of different similarity estimates and aggregation techniques. An evaluation was performed on 15 Swedish and Danish local clinical models from five different organisations.

[Conclusion] The clinical model comparison studies showed that an analytical approach based on a consistent representation using SNOMED CT can be used as a basis for identifying common content i.e. semantic overlap and providing an overview of multiple templates. In conclusion, the developed methods contribute to bridge centralised and de-centralised approaches because they support harmonisation clinical models.

Abstract in Danish

[Introduktion] Elektroniske patient journal (EPJ) systemer forventes at kunne forbedre informationsflowet i sundhedsvæsentet, men den nuværende forskning indenfor EPJ-systemer er fragmenteret. Studier har vist at det er muligt at udvikle innovative systemer med tæt bruger-involvering på decentralt niveau. Skalering af disse systemer har dog vist sig at være et problem, blandt andet fordi systemernes datamodeller, kliniske modeller og terminologi er lokale. Manglende skalering betyder at patientinformationen bindes til lokale systemer med begrænset interoperabillitet. Frameworks for interoperabillitet, der er baseret på standadiserede reference informationsmodeller, kliniske modeller og terminologier der kan anvendes i EPJ, er blevet foreslået af forskere og standadiseringsorganisationer, men udviklingen frem mod semantisk interoperabillitet går langsomt.

[Formål] Formålet med denne PhD-afhandling har været at bygge bro mellem de-centrale og central kliniske modeller ved at anvende sematiske metoder til sammenligning af indhold. Metodeudviklingen har været guidet af to forskningsspørgsmål. Q1: Hvilke forudsætninger er der for at anvende semantiske metoder til sammenligning af lokale kliniske modeller. Q2: Hvilke semantiske tilgange kan anvendes til sammenligning af lokale kliniske modeller? De to spørgsmål er besvaret i de fem inkluderede artikler.

[Q1: Artikel I,II og III] Harmonisering er en forudsætning for sammenligning af lokale kliniske modeler, fordi modellerne er udtrykt ved hjælp af leverandør afhængige formalismer og lokal interface terminologi. I artiklerne I, II og III blev der udviklet og evalueret artefakter der sikrede en konsitent repræsentation af lokale kliniske modeller både med hensyn til formalisme og terminologi. Formålet med artikel I var at udvikle en informationsmodel-formalisme baseret på en analyse af danske kliniske modeller. Formålet med artikel II var at udvilke guidelines, der gjorde det muligt at udtrykke interface terminologi i kliniske modeller ved hjælp af SNOMED CT. Guidelines blev udviklet itterativt i en proces, hvor de to forfattere mappede 14 svenske og danske lokale kliniske modeller til SNOMED CT. Formålet med artikel III var at foreslå et semantisk mål for inter-rater agreement og evaluere det i forhold til et nominelt mål. Studiet blev lavet på baggrund at et sæt af svenske procedurer, der var blevet mappet til SNOMED CT af to forskellige kodere.

[Q2: Artikel IV and V] I artikel IV og V blev der udviklet to metoder til semantisk sammenligning af kliniske modeller. Begge metoder anvendte en konsistent repræsentation af indhold som beskrevet i artikel I og II. Formålet med artikel IV var at analysere fire kliniske modeller vedrørende objektiv undersøgelse fra fire forskellige organisationer i Danmark og Sverige. Indholdsanalysen blev baseret på par-vis analyse af modellerne med hensyn til antallet af ens dataelementer, terminologisk ensartede dataelementer og uens dataelementer. Desuden indeholdt analysen en sammenligning med tekstbogsmateriale samt en analyse af fælles indhold i de kliniske modeller. Formålet med artikel V var at give et overblik over et større antal kliniske modeller ved at anvende hierakisk clustering. Forskellige kvantificeringsteknikker, der anvendte en kombination af similaritets estimater og aggregeringsteknikker blev undersøgt. Evalueringen blev udført på 15 svenske og danske lokale kliniske modeller fra fem forskellige organisationer.

[Konklusion] Studierne vise at en analytisk tilgang til sammenligning af kliniske modeller baseret på en consistent repræsentation vha. SNOMED CT kan bruges som basis for at identificere fælles indhold, samt til at give overblik over mange modeller. Det konkluderes at de udviklede metoder bidrager til at bygge bro mellem centraliserede og de-centraliserede tilgange fordi de på forskellig vis understøtter harmonisering af kliniske modeller.

Contents

Definitions and Abbreviations

EHR	Electronic Health Record: A repository of patient data in digital form, stored and exchanged securely, and accessible by multiple authorized users. It contains retrospective, concurrent, and prospective information and its primary purpose is to support continuing, efficient and quality integrated health care. EHR has multiple subtypes. Some of these specify EHRs that contain all or most of a patient's clinical information from a particular hospital. The hospital specific subtypes are Hospital Electronic Medical Record, Electronic Patient Record and Computerized Patient Records [1] . Most of the mentioned EHRs in this thesis are of this subtype. However, a few are Inter-hospital EMR's. Inter-hospital EMRs contain information from two or more hospitals [1] . Consequently, the term EHR is used throughout the thesis.
EHR system	We distinguish between EHR and EHR system. The EHR is the repository i.e. the actual patient information. An EHR system is an information system that can be used to input and access the EHR.
Semantic interoperability	Ensuring that the precise meaning of exchanged information is understandable by any other system or application not initially developed for this purpose [2].
Information model	Model of data structure, e.g. in databases, messages or interfaces [3].
DCM	Detailed Clinical Model. DCMs are *"the structure of clinical data that are stored and managed in electronic patient records, send between clinical systems, and referenced in decision support rules."* [4]. However, as oppose to Clinical Models, the term DCM implies some degree of standardisation and reusability [4].
Clinical model	A collective name for detailed clinical models and local EHR templates. They are information models that express clinical concepts including possible values and datatypes to be used for e.g. data entry, messaging or information retrieval.
Template	Commonly used name for clinical models in local EHR systems, especially for data entry purposes. "openEHR template" is a template constructed by combining and constraining openEHR archetypes. In this PhD-thesis it is stated explicitly if referring to openEHR templates. Else, template refers to the common meaning of the word.

1 Introduction

Patient information is a valuable resource in clinical care and analytics. In clinical care, information about past medical history and the patient's current encounter make it possible for clinicians to make informed decisions about patient treatment. In analytics, information from multiple patients facilitates statistics regarding disease prevalence and incidence, outcomes associated with specific treatment regimes etc. Analytics is important in quality of care monitoring, health management and research. However, patient information is not a consistent and well-organized entity. It is distributed between different health organisations and professionals, held partially on paper and partially in clinical information systems and represented in different modalities e.g. structured forms, narratives and images [1]. Information organisation in this complex health environment is a challenge, and a longitudinal patient-centred electronic health record (EHR) is an anticipated solution to this problem [1]. More so, policy documents from England, Denmark and the U.S. state that the introduction of EHR systems will make health care better, safer, cheaper and more integrated, and that problems like missing files, double registration, medication errors and ideosyncratic clinical decisions will be solved [2-4].

Successful innovations have been made were EHR systems have been developed to support small-scale, practically focused work in well-defined clinical environments, but often the solutions do not scale [5,6]. In other words, many local EHR-systems are departmental or hospital centred rather than longitudinal and patient-centred. They succeed in meeting close-to-user needs, but fail to meet shared clinical care and analytics requirements.

Supporting shared care and analytics has inspired research and development in the field of semantic interoperability i.e. preservation of meaning across systems [7]. Semantic interoperability requires the combined use of standardised information models and terminology [8]. Researchers have developed, refined and evaluated information models, e.g. [9-12] and terminologies, e.g. [13,14] or both [15-17], and investigated the mismatch between different approaches within standardisation [18]. However, a review of semantic interoperability in eight European countries plus Australia, Canada and USA concluded that there was limited progress towards full semantic interoperability [19].

The aim of the introduction of this PhD thesis is to further explore the fragmented EHR content research i.e. the seemingly contrasting approaches to content development:

> Supporting day-to-day clinical work requirements, but risking local systems that do not scale.
> Supporting the information need of many and developing interoperable frameworks, but risking that the frameworks prove difficult to implement.

Figure 1 illustrates the fragmentation within the areas of content requirements, content development methods, information models and terminology. Each of these areas is expanded on in the following sections.

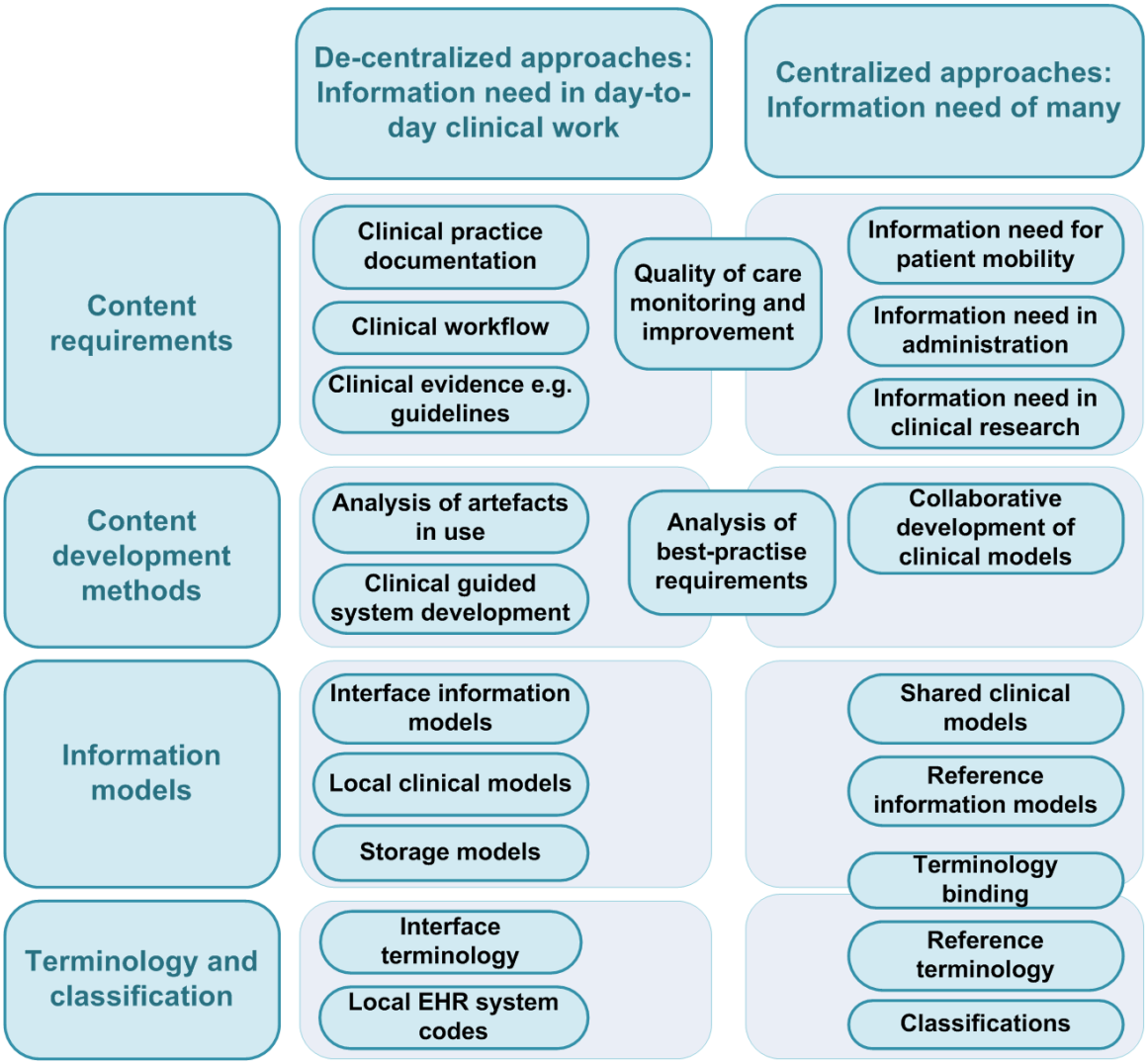

Figure 1 - Centralised and decentralized EHR system development

1.1 Content requirements for EHR-systems

EHR-systems are implemented in complex healthcare settings, and consequently, requirements for EHR-systems are many. A review by Hoerbst et al. categorizes EHR-system requirements into the following 11 areas: global requirements, data security, content, usability, interoperability, reliability, privacy and data protection, maintainability, performance/efficiency, portability and general functional requirements. Specific functional requirements e.g. content of medication lists or requirements that concern order entry were excluded in the review [20]. The 11 requirement areas reflect the need for at comprehensive IT infrastructure in health, which calls for multidisciplinary efforts. Hoerbst et al. differentiate between content, interoperability, maintainability and portability related requirements, the following analysis will argue that these are to some extend intertwined.

Understanding EHR-system content requirements presuppose a clear distinction between knowledge and information. Knowledge is accepted truth as discovered by research and information is linked to the single

patient [12,21-23]. The knowledge-information distinction is evident in EHR-system information modelling, for example*: ""Gina Smith has a resting BP of 110/80" (information) and "Blood Pressure consists of two quantitative data items, called 'systolic' and diastolic' (each with units =mmHg), and an optional 'protocol' indicating position of patient, instrument used and type of cuff" (knowledge)"*[12]. Whereas these can be referred to as knowledge-level and information-level [12], other definitions are also accepted e.g. type level (knowledge) and instance level (information) [24] .

Knowledge-level and information-level related requirements constitute core functionality of EHRs and EHR-systems. The requirements are not only complex, they are also contrasting. The contrasting requirements of flexibility, stability and interoperability are explained below and illustrated in Figure 2.

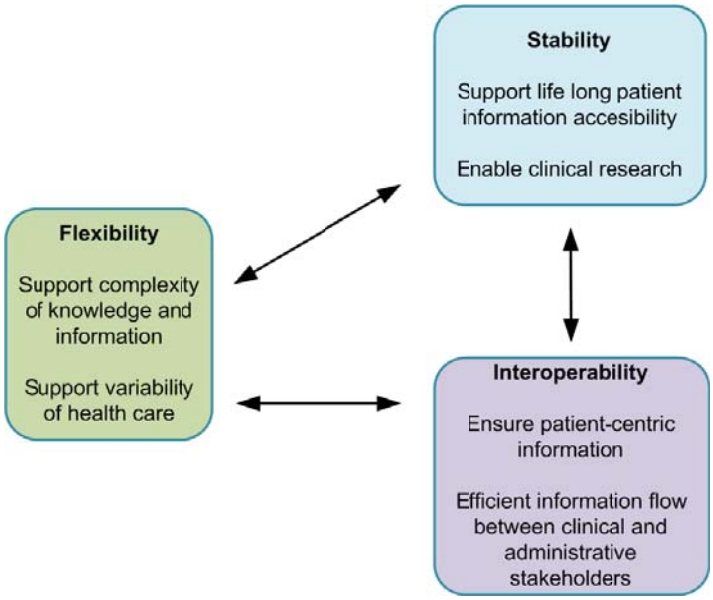

Figure 2 - Contrasting requirements for EHR-systems

Stability is a requirement because it should be possible to retrieve patient information throughout the patient's lifespan and possibly beyond [23]. Moreover, patient information is a valuable source in clinical research [25]. Clinical research requires stable definitions because this is a prerequisite of obtaining comparable patient information in considerable time spans. Required stability is probably one of the reasons why (core) IT-systems are allowed to grow old in health care. In addition, EHR-system changes should be considered carefully because software errors are potentially a matter of life and death in health care [23]. Consequently, stability is more than a requirement. Stability is also a characteristic of the current EHR-system landscape that makes it difficult to move towards interoperable and/or flexible systems.

Flexibility is a requirement because EHR-systems have to support the complexity of medical knowledge and information and the variability of health care [23]. The complexity of clinical knowledge is described by the ever evolving knowledge base in medicine *"Medicine is not only big, it is also open-ended in depth, breath and complexity"* [14]. Depth means that new details of known entities are discovered, breath means that new phenomena are discovered and complexity means that new relations between known entities are discovered. The big and open ended nature of clinical knowledge requires flexible knowledge-level modelling in EHR-systems. For this flexibility to be meaningful there should be a stable link between

knowledge models and patient information. Patient information adds a complexity of its own due to different modalities e.g. unstructured clinical notes, filled clinical forms, images and ECGs [23]. In addition, the variability of treating patients is high [23]. Consequently, knowledge level modelling should be flexible both in terms of data definition and work flow support. Flexibility is a challenge because of the before-mentioned stability requirements, but also because it can be considered as conflicting with interoperability requirements. Whereas flexibility allows EHR-systems to support the micro details of clinical work, interoperability frameworks are more concerned with the shared information need of multiple stakeholders. The tension between interoperability and flexibility is a well-known challenge [26].

Interoperability is a requirement because information should follow the patient, not systems or organisations so that each of the Health Care Professionals treating the patient has updated information available [23]. Interoperability would allow aggregation of patient information for administration, quality monitoring and research [25], enhance information use in clinical workflow and help avoid double registration [27]. Consequently, semantic interoperability that allows seamless integration is a much desired characteristic of EHR-systems [8,28]. However, in most countries, health care IT is characterised by heterogeneity that does not allow information sharing [23,24]. Add to the complexity of current information systems, the complexity of the domain and stakeholders. There are stakeholders employed in primary, secondary and tertiary health care and within each sector there are different health professionals with different information preferences. In addition, local as well as national health administrations have information needs with different management objectives such as monitoring hospital productivity and supplying information to national cause-of-death registries. Interoperability should be prioritized in EHR-systems and architectures throughout the health care sector for patient information to flow efficiently between stakeholders. Interoperable architectures face the challenge of taking into account different organisational needs while also accounting for legacy of the heterogeneous system landscape [27].

Whereas stability is a characteristic and requirement that affect all aspects of information modelling and EHR infrastructure development, flexibility and interoperability is directly related to the fragmentation between the centralized and decentralized approaches in EHR system development. This is illustrated in Figure 3.

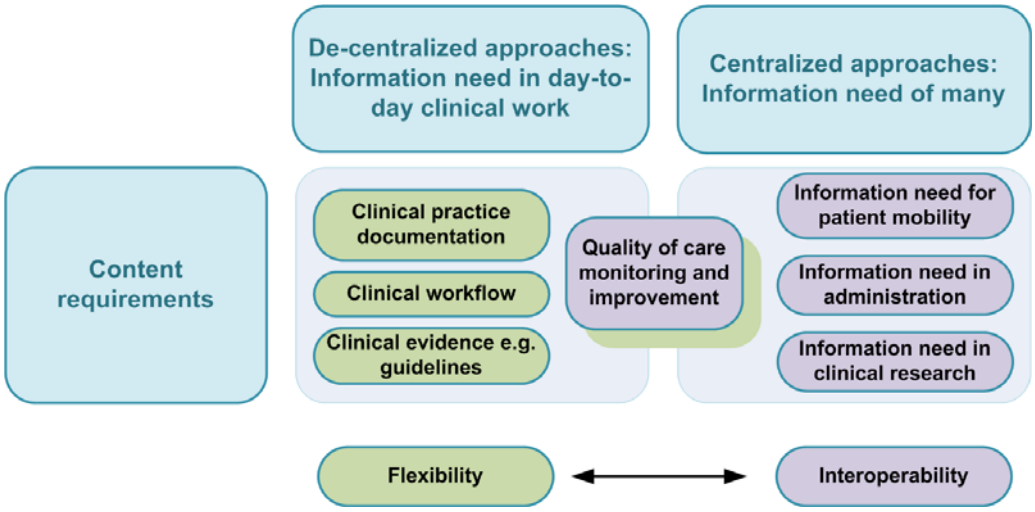

Figure 3 - The contrast between flexibility and interoperability in content requirements.

1.2 Content development methods for EHR-systems

The complex and contrasting content requirements of EHR-systems underlines the need for well-established content development methods. These methods are summarized in Figure 4. Content development methods are presented in more details in the following.

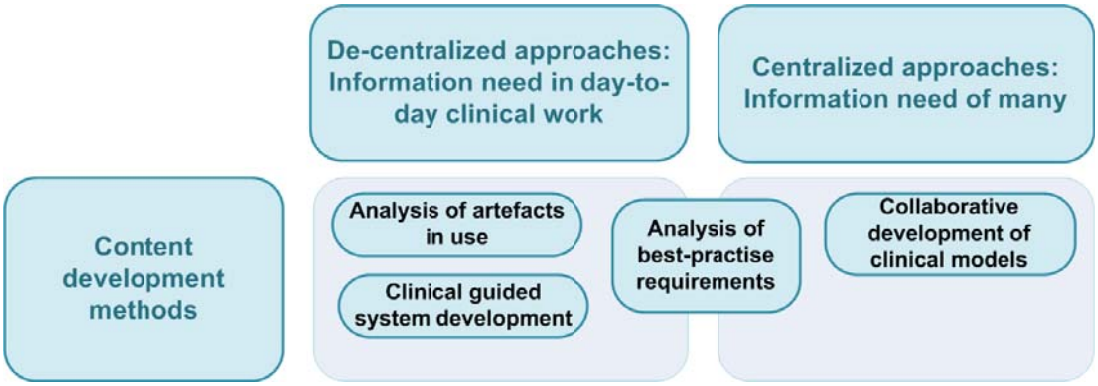

Figure 4 - Content development methods for EHR-systems

Analysing artefact in use has been explored scarcely in the scientific literature. Existing artefacts like EHR-system, patient information captured in IT or paper-based systems or other documentation instruments can be seen as a resource when collecting content requirements. For example, Elisabeth Chen et al. made a multi-site content analysis of unstructured notes regarding social history information to develop a structured social history template [29], and Håkonson et al. analysed existing nutrition status instruments before developing a nursing minimum dataset related to nutrition [30].

Clinically guided system development is mostly evident in participatory design research. It is important to note that it is not the main focus of researchers in this area to separate usability and user-interface requirements from content requirements, because they focus on how they can form a whole that support users in their workflow. Participatory design of EHR systems is an extensive research fields and the following are merely illustrative examples. In 1994, Nowlan published an early example of a participatory approach [31]. An EHR-system was developed iteratively and consisted of design-cycles of rapid prototyping and formative evaluation in which user participation was an integral part [31]. Another approach was described by Garde and Knaup. During an eight year period they iterated on a clinical information system exploring users' needs in participatory meetings, by surveys and interviews [23]. Both Nowlan's and Garden and Knaup's approaches are interesting because they have a multicenter view and rely on systematic design approaches. On contrary, multiple studies use more or less structured methods to involvement of users in EHR system design without this being the primary focus of their papers e.g. [32] and [33]. Some authors have argued that a multi-disciplinary participatory method is a prerequisite when aiming at supporting the team-oriented nature of modern health care [34]. The risk of the user-centric iterative development is that the bigger and more inflexible the system gets the more difficult and time consuming a design-cycle gets [31], and there is a risk that only the users involved in the design will use the system [23].

Analysis of best-practise requirements can be applied as a method when supporting evidence-based care is a goal of the EHR-system. Supporting evidence-based care might very well be a goal because the

compliance of evidence based care in clinical practise is far from perfect [35]. Promising results suggest that implementation of guidelines or decision support in computerized systems could improve the compliance [36], and there are numerous examples of dedicated decision support systems and systems where guidelines can be viewed e.g. to avoid prescription errors [37]. However, integration of Clinical Decision Support (CDS) with EHR system is more uncommon, even though attempts are made [23,38]. Chen et al. analysed and a chemotherapy guideline and represented it using EHR information models [38]. The before mentioned study by Garde and Knaup focused on the evidence-based care by systematically reviewing trial protocols in the clinical area of interest and included the result in the EHR-system design [23].

It may be argued that there is not much difference between analysing guidelines and protocols, as described above, and performing content analysis of nutrition screening instruments [30], because both methods are based on an analysis of knowledge-level representations. However, "nutrition screening instruments" implies that design choices has already been made in terms of what patient information should be collected in a specific context, whereas "chemotherapy guidelines" describes the clinical practise and does not necessarily include specific design choices.

Multi-site collaborative development is another way of approaching user's requirements, but the scientific evidence is scarce. One study evaluated a method where health care professionals could join, and contribute with requirement using a web page for collaboration [39]. In addition to this example from the scientific literature, multi-site collaborative development is one of the ways that the information modelling organisation, the openEHR foundation, creates content using a clinical knowledge management system where members can contribute to and review existing models [40,41].

1.3 Information models

Objectives of the international modelling initiatives are to reduce duplication of effort and support semantic interoperability. The international work on health information modelling is typically found within standardisation organisations and academia. Figure 5 summarizes information models that contribute to the E-health landscape. In the following, the models and how they are mentioned in the scientific literature will be presented. Local model implementation efforts are typically not mentioned in the scientific literature, but to give a full picture these are added to Figure 5.

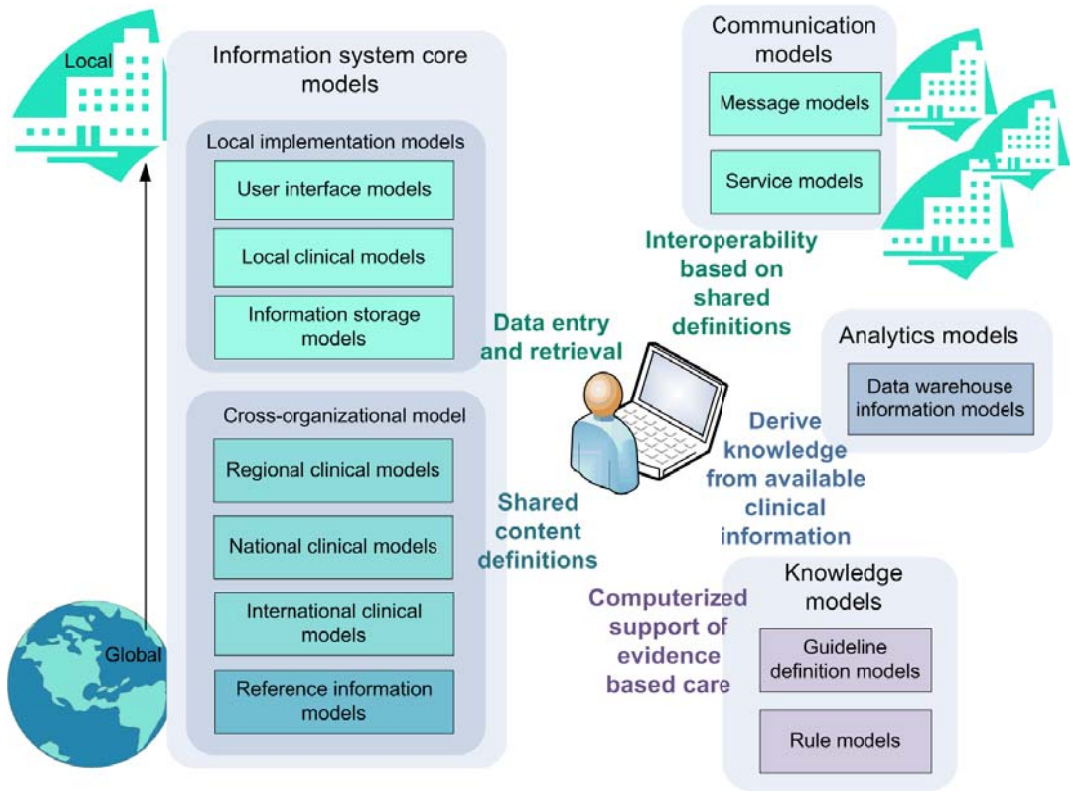

Figure 5 - Overview of information models

Consistent and valid cross-institutional EHR-systems require standardization to be achieved at different levels. Several papers have provided overview of available EHR standards [1,24,27,42,43]. In 2007, Knaup et al. provided the following overview of artefacts that could be standardized [27]:

> Basic data types including general applicable datatypes such as ISO 11404 and health specific data types such as HL7v3 data types.
> Messages and services e.g. general standards such as HL7v2.x messages and speciality specific such as Digital Imaging and Communications in Medicine (DICOM).
> Architectures for EPR systems and knowledge-enabled computing such as the HL7v3 in combination with CDA and the openEHR framework.
> Clinical content such as openEHR archetypes with reference to clinical terminologies

The basic data types are what underlie the reference information models, but the basic data types are not illustrated specifically in Figure 5. The messages and services are illustrated as the models that provide interoperability based on shared definitions. In the architecture category, Knaup et al. placed the two-level modelling approaches. Dual modelling is a paradigm where information-level and knowledge-level modelling is separated i.e. [12] Reference information models store data (information-level) and clinical models semantically describe the data structures (knowledge- level). Since Knaup et al. wrote their review, two-level modelling has dominated the way research environments and standardisation organisation organisations think about EHR-architectures. Consequently, a later categorization would probably split reference information models and clinical models (what Knaup et al. refers to as clinical content) into two

different categories, knowing that they together constitute the core models of EHR architectures. This categorization is used in Figure 5.

In addition to the model-oriented standards presented above, Knaup et al. also mention ISO work on requirements of EPR systems, clinical terminologies such as ICD10 and SNOMED CT and presentation of clinical concepts. These are not modelling standards per se, and they will not be considered in this section.

Two additional types of standards/specifications which are relevant for EHR-systems are mentioned by Kalra and Ingram [1]:

> Knowledge models i.e. models for reminders, alerts and decision support. Rule syntaxes such as Arden Syntax, proForma and Prodigy might be seen as de facto standards that cover this area. However, Kalra and Ingram see the tight integration with the EHR as most important in terms of linking guidelines and decision support to individual patients [1]. Innovative solutions such as GDL developed to fit with the clinical models in the openEHR framework are promising in this respect [44]. The knowledge models are illustrated in Figure 5 as the models that allow computerized support of evidence based care.
> Implementation profiles such as the IHE profiles [1]. The reason for mentioning these is that in recent years there has been a growth in developing resources and patterns that makes implementation easier. In this category, initiatives such as HL7 FHIR [45] could also be placed. FHIR has a strong focus on implementability even though, in its core, it is a two-level modelling approach requiring a reference information model and clinical models specific to HL7 to function. Implementation profiles are not illustrated specifically in Figure 5 because they can be seen as a special type of EHR system core models and communication models.

In addition to the models and standards mentioned in the scientific literature, analytics models might be a new type of models that would be worth to add to the EHR-model family. This would be a logical consequence of the increased awareness on big data and data mining in health care e.g. [46]. Analytics models are added to Figure 5.

The modelling initiatives with the closest relation to the EHR content are the clinical model initiatives because their focus is to engage health care professionals in developing the information artefacts needed in health care.

1.3.1 Clinical models

Reviewing clinical models require the distinction between the model formalism or format, and the model instantiation for different clinical situations. For example, openEHR archetypes have the formalism: archetype definition language (ADL). Each openEHR archetype is an instantiation of this underlying formalism for a clinical situation or entity e.g. laboratory data. The research and development within the field of clinical models is both within the area of formalism and the area of instantiation.

Clinical model formalisms have been developed by standardisation organisations, researchers and cross-organisational foundations, as well as private vendors. In the following, the focus is on three international standardisation initiatives.

'Archetype-template approach' is used as a collective name for the GEHR projects , openEHR archetypes and templates and ISO 13606 archetypes. GEHR stands for Good European Health Record, later called Good Electronic Health Record because of extended participation from non European countries. The GEHR projects aimed at developing a future-proof architecture for EHR systems. In this project, the two-level modelling approach was developed by utilising a small generic reference model constrained by a number of clinical archetypes [47], an idea based on the work of Rector et al.[48] , Johnson [49] and Beale [12]. The GEHR two-level modelling approach was adopted by the openEHR foundation and later CEN/ ISO13606. openEHR archetype entry-types can either be observations, evaluations, instruction or actions. In the modelling community, ISO13606 archetype formalism is seen as a subset of the openEHR archetypes i.e. not all archetype-types are adopted in the international standardisation work in ISO. In addition, some differences exist in the openEHR and ISO13606 reference models that affect interoperability [50]. In openEHR, archetypes can be composed and constrained to form templates, which are e.g. used to define which patient information that should be entered at different points of care [51]. ISO13606 focus on communication, and does not imply any specific architecture [52](and hence do not define templates).

Clinical models in HL7 include HL7v3 messages, CDA and FHIR resources and profiles. HL7 is a standardisation organisation in the United States which has been very successful in terms of promoting message communication in health care and HL7 v2 messages are used worldwide. The development of the HL7 v.3 reference information model (RIM) signified that the HL7 organisation has moved beyond message standards into more broad architectural standards. In the context of HL7v3 a clinical statement model was developed that can express almost any clinical content in a RIM consistent manner. The clinical statement model is the basis of the clinical content of both HL7 message fragments (known as CMETs) and CDA [53,54]. CDA (clinical document architecture) is the most used HL7 v.3 standard [53]. CDA is known for its different implementation levels. At the one end it can mark up simple documents and at the other end use the clinical statement model to express clinical information structured and RIM consistent [55]. FHIR (Fast Health Interoperable Resources) is published as a Draft Standard for Trial Use, which means that it has not yet gone through the needed ballots to be considered a standard [56]. FHIR resources and profiles are based on the RIM. Resources are the underlying building blocks that can be composed and extended to form profiles for specific clinical purposes [57]. As oppose to CDA and CMET, FHIR do not use the clinical statement model, and modellers do not need detailed knowledge of the RIM to develop and use resources and profiles [53,54].

Clinical Information Modelling Initiative (CIMI) is a collaborative initiative with the aim to harmonize existing information models. The participant are from non-profit organisations, national authorities and industry e.g. EN 13606 foundation, HL7, openEHR foundation, IHTSDO, Canada infoway and Intermountain Healthcare. Participants in CIMI have proposed a harmonized reference model and clinical model syntax. Additionally, they have analysed similarities and differences between existing clinical models to propose harmonized content for a shared content library [58,59].

In addition to these three international collaborative initiatives, the clinical model harmonisation effort of Goosen et al is also worth mentioning. They have developed the Detailed Clinical Models (DCM) as a close-

to-practise model which at a later stage can be expressed using a standardised framework [60]. Experiments have been carried out where DCMs bridge the difference between models specific to HL7 and EN13606 respectively [18]. Clinical models can also be national models e.g. [39] or vendor specific models[61] that do not primarily refer to any standards or international collaborative work.

Studies have shown that clinical model formalisms mainly differ from each other with regard to the presence of an underlying reference model [62], and the expressiveness of the model formalism [62,63].

Clinical model implementation: Wollersheim et al. have reviewed research relevant for model implementation. They understand implementation in a broad sense meaning that both examples of implementation as well as different kinds of analysis of the implementation potential or comparisons with other models are included in the review. They included only openEHR/13606 archetypes. Their main result is that it is possible to develop archetypes within different clinical fields and geographical contexts. Wollersheim et al. included 47 studies in their review. When adding to this the clinical models developed within standardisation organisation, national authorities and local/regional organisations, the share number of models suggests the need for quality criteria. Recently, such criteria have been suggested by Ahn et al. They reviewed other author's suggestions for quality metrics and used a Delphi survey as a basis for suggesting eight domains and 29 specific quality metrics [64]. The eight domains are:

Scope and purpose
Stakeholder involvement
Rigor of development
Clarity of presentation
Compliance to standards
General methodology
Metadata
Management and maintenance

Notably, Wollersheim et al and Ahn et al both report that clinical models should be developed using well-defined design processes [11,64]. The design processes associated with different modelling initiatives have recently been criticized in a review by Blobel et al. Their main result is that most modelling initiatives risk violating good modelling practice by failing to allow the domain requirements to influence modelling because top-down modelling is the most obvious choice [65].

Clinical model usability increase when they are linked to clinical terminologies [27], and the use of clinical terminologies with clinical models are a prerequisite for semantic interoperability [8]. Some research exist that have studied the link between information models and terminology i.e. terminology binding. Terminology binding researchers have focused on automation [17,66], and interoperability between models with different granulation level, coordination level and context [67].

1.4 Clinical terminology

Clinical terminology systems have been focus of attention in research because they can support a consistent and unambiguous representation of information. The unique identifiers and the conceptual structure representing each concept in a terminology system allow an unambiguous interpretation of the concept's meaning across EHR-systems. This is beneficial for the use of information in both clinical practice

and for secondary purposes [68]. Different types of terminologies are summarized in Figure 6. These are expanded on in the following.

Clinical reference terminologies and clinical classifications are two different entities. Existence of rich reference terminologies like SNOMED CT or LOINC is justified in capture of clinical information at the level of detail relevant to a specific clinical context. In contrast, classifications are characterized by summarized disjunctive categories, which make them fit for statistical purposes [21,69]. In many countries this means that administrative and research databases as well as reimbursement rely heavily on the International Classification of Disease v. 10 (ICD10) or v. 9 whereas local use of EHR-systems needs the details of reference terminologies and/or a local terminology. Consequently, clinical models have to link to both classifications and terminologies meaningfully. A labor saving, but largely unexplored, scheme is to capture detailed information using reference terminologies like SNOMED CT and leave it to computers to recode the data as needed for administrative, quality assessment and research purposes e.g. using ICD10 [70]. This requires cross mappings between ICD10/11 and SNOMED CT, which is a theme in WHO and IHTSDO respectively as well as in the research community [71-73].

At an interface level, reference terminologies or classifications may not be appropriate. Consequently, interface terminologies that reflect users' language have been developed. Research within the field of interface terminology is e.g. concerned with mapping and maintainability. Experiments have shown that mapping to reference terminologies is possible [74], but maintainability of such a map should be considered [75]. In some cases, interface terminologies are shared [74] in other cases they are proprietary [76].

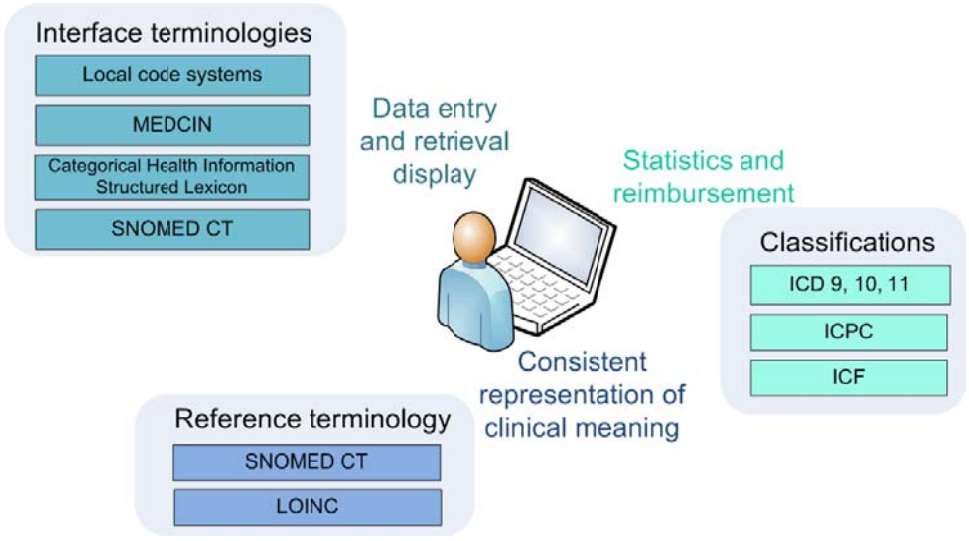

Figure 6 - Clinical terminologies and classifications and their purpose

1.4.1 SNOMED CT flexibility and consistency

When multiple clinical specialties are covered with the same clinical terminology, a consistent representation of clinical information can be obtained. Consistent representation is most important when querying patient information across different departments, hospitals, regions or nations. Cross-departmental EHR-system will therefore require reference terminologies that cover the clinical domain as a

whole, and allow flexibility i.e. it should be possible to produce extensions when specific concept cannot be identified in the terminology. SNOMED CT fulfill these requirements because it has proven flexible compared to other terminologies [77-80], and because it has shown a high level of coverage in different clinical fields [76,81,82].

SNOMED CT flexibility has its roots in its complex multi axial and compositional structure. Compositional means that if a given term cannot be mapped to SNOMED CT, two or more SNOMED CT concept can be combined to form the meaning of the term. This is referred to as post-coordination whereas terms that can be represented with one SNOMED CT concept are pre-coordinated [2].

Post-coordination means that SNOMED CT becomes expressive and flexible from a clinical viewpoint because terms can be combined to form relevant clinical expressions. An example is the 788 different subtypes of pain that can be qualified with the seven types of severities of the pain. This means that composite expression: 65761003 | inflammatory pain |: 246112005 | severity | = 255604002 | mild |can be used to express “mild inflammatory pain”. Numbers of post-coordinated terms in mapping studies varies from 2.3% [83]–78%[76] in different clinical fields using different mapping guidelines. Even though post-coordination improves flexibility, a recent survey on SNOMED CT implementations, found that implementation of post-coordinated terms remains a challenge [84].

The defining relationships of SNOMED CT concepts allow retrieval of patient information based on meaning. Basic semantic queries utilize the hierarchies of SNOMED CT. E.g. searching for ‘Cardiac finding’ and descendants could retrieve information like ‘blood pressure finding’, ‘pulse rate finding’ and ‘finding of peripheral pulse’ from various EHR templates. This would be of interest in overview of single patients or, if queried for multiple patients, hypothesis-formulation in clinical research.
More sophisticated methods use the defining characteristics of a concept expressed in a common form to query and retrieve stored information. So, instead of being limited to only retrieve information based on the hierarchical structure, each of the defining characteristics of a concept can be used as entries for detailed queries [85]. For example, the body structure associated with a finding could be used to illustrate place of present illness on a sketch of a human. This type of queries requires consistent use of SNOMED CT.

Consistency of SNOMED CT representations remains a challenge. Studies show that when using SNOMED CT redundant representations of identical clinical information occur [86,87]. Redundancy problems are highlighted in coding variability studies [87,88]. In one of the studies performed by Andrews et al. [87] the coding variability found was partly due to different levels of contextualizing the data, and partly due to different approaches to perform post-coordination. Overall, the study emphasizes the need for consensus and communication regarding the use of SNOMED CT. Redundancy is not the only shortcoming in SNOMED CT that hinders efficient use, also, gaps in the terminology, shortcomings in compositional structure and consistency problems are reported issues [89].

1.5 Challenges in content development for EHR systems

The introduction has documented that research in the field of EHR content development has been evolving in recent years. The complexity of the health care domain and EHR system requirements has lead to a situation where researchers from different disciplines have focused on different aspects of EHR content development. Understanding the health care domain and developing requirement collection methods have been developed under a social science paradigm, where the micro details of clinical work, different

stakeholders and clinically guided system development have been in focus. An engineering paradigm has been driving the modelling community to discover which information models, terminologies and terminology bindings would be needed to achieve semantic interoperability to support use cases such as analytics and shared care. Meanwhile, vendors, local information modellers and clinicians have continued to implement proprietary EHR systems. The practical experiences are only scarcely mentioned e.g. [61], and the content of local clinical models have, to my knowledge, not been the focus of any research.

From a research perspective, focus on specific objectives is a prerequisite in achieving an appropriate depth. However, when a modelling community review criticizes the clinical models ability to reflect domain requirements [65] cross-disciplinary efforts may be needed. At least, efforts that aim at bridging the gap between the information need of local work practice and the information need of shared care and analytics may be worth exploring.

2 Objectives

The objective of this PhD study was to bridge de-centralized and de-centralized approaches in EHR-system content development. Given the fragmented nature of existing research, many different specific research objectives could have been chosen. Typically, scientific literature mentions “the islands of information” within proprietary systems as the problem that limits interoperability. Taking the opposite stand i.e. “standardization can learn from local clinical models” has been the point of departure of this PhD-study. Local clinical models could be analyzed with respect to their formalism using a research design similar to that of Goossen [62]. Alternatively, they can be regarded as artefacts in use and analyzed based on their content using a content comparison methodology [29,30,90]. The latter has been chosen in this PhD-study because comprehensive standardized modeling frameworks have already been developed, or are under development. However, analyzing clinical models from local EHR systems might help bridge the gap between the information need of many and supporting local users, which is an unsolved and largely unexplored area.

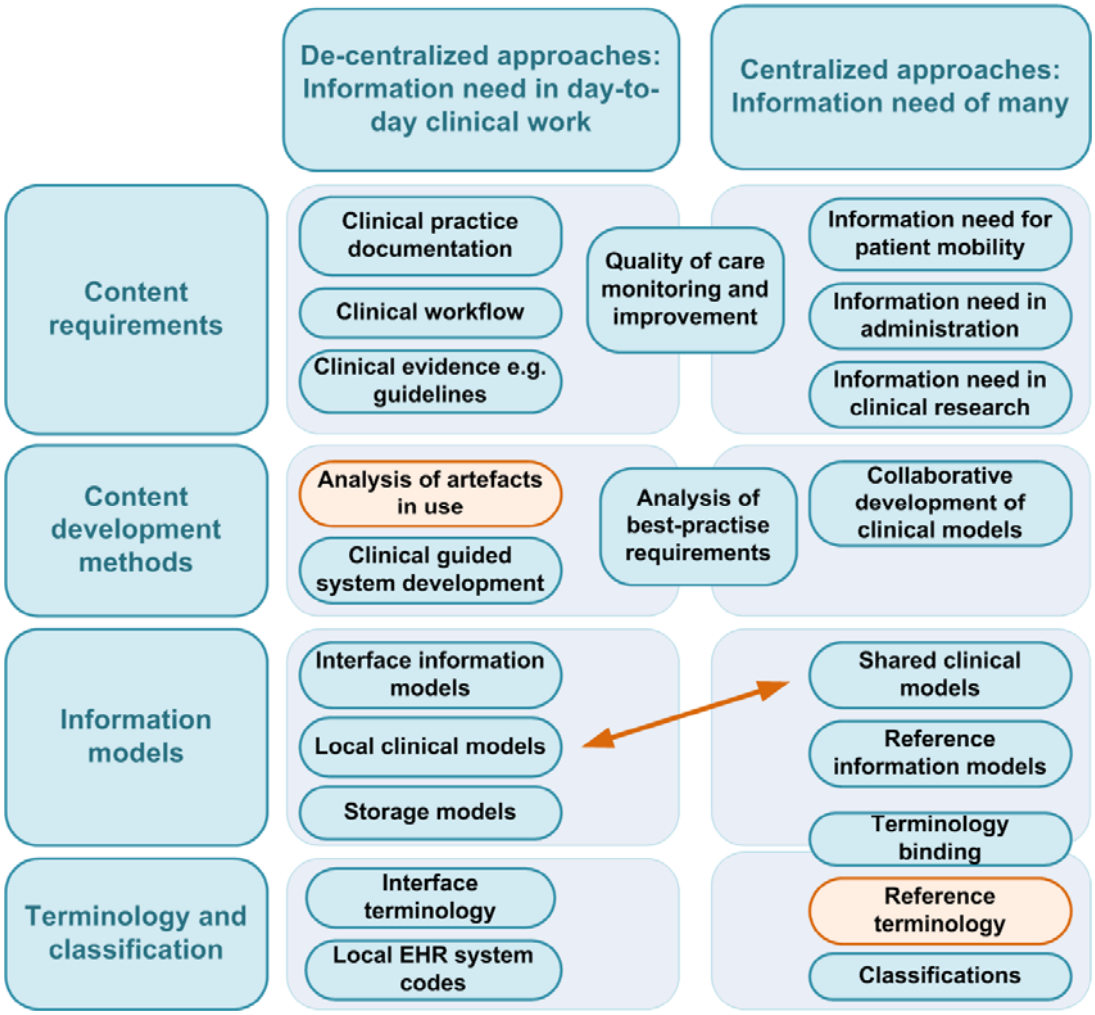

Figure 7 – Bridging centralized and de-centralized content development for EHR systems

Focusing on artefacts in use means that the PhD-study also contributes to a scarce research area. The few existing studies have used either qualitative content analysis methods [29,30] or analysis methods based on semantics using a clinical terminology as a reference [90]. Improving semantic methods for content comparison was another aim of this PhD-study. The research questions that have guided this work are:

> What are the prerequisites of applying semantic methods for local clinical model comparison?
> Which semantic approaches can be applied to local clinical model comparison?

In Figure 7, the specific research objective of bridging local and shared clinical models is illustrated, and the key elements in the methodology is marked in orange i.e. **analysis of artefacts in use** and applying a **reference terminology** in the effort. The method is described in the following chapter.

3 Material and methods

An overview of research questions for this PhD-thesis is illustrated in Figure 8.

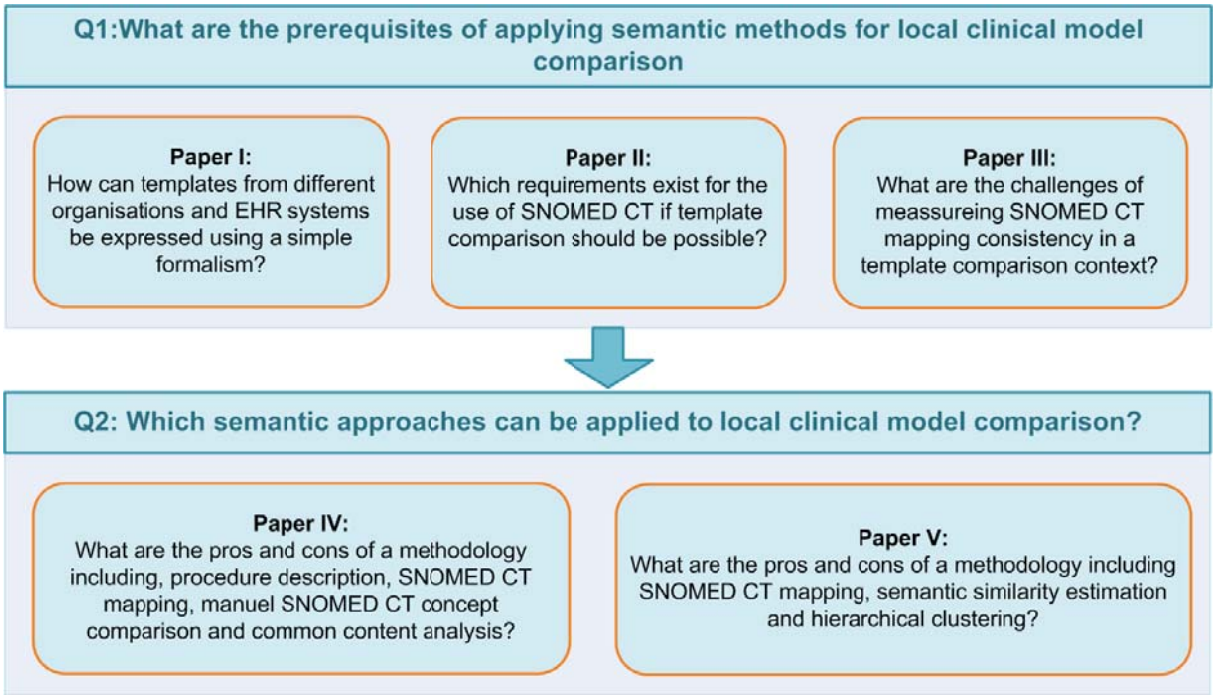

Figure 8 - Methodological framework

The research questions are closely related to the Thesis-papers. The following refers to Q1 and Q2 in Figure 1:

> **Q1: What are the prerequisites of applying semantic methods for local clinical model comparison?** is addressed in Paper I, II and III respectively.
> **Q2: Which semantic approaches can be applied to local clinical model comparison?** is addressed in Paper IV and Paper V respectively.

The questions were answered using a design-science approach, as described by Hevner et al. [91], see Appendix **Error! Reference source not found.**. The specific methods are described in the method-section of each paper. However, the methods of paper IV and paper V have subtle similarities and differences, therefore a comparative description of these methods are presented in the following.

3.1 Template comparison methods

The analysis methods used in paper IV and paper V respectively are illustrated in Figure 9.

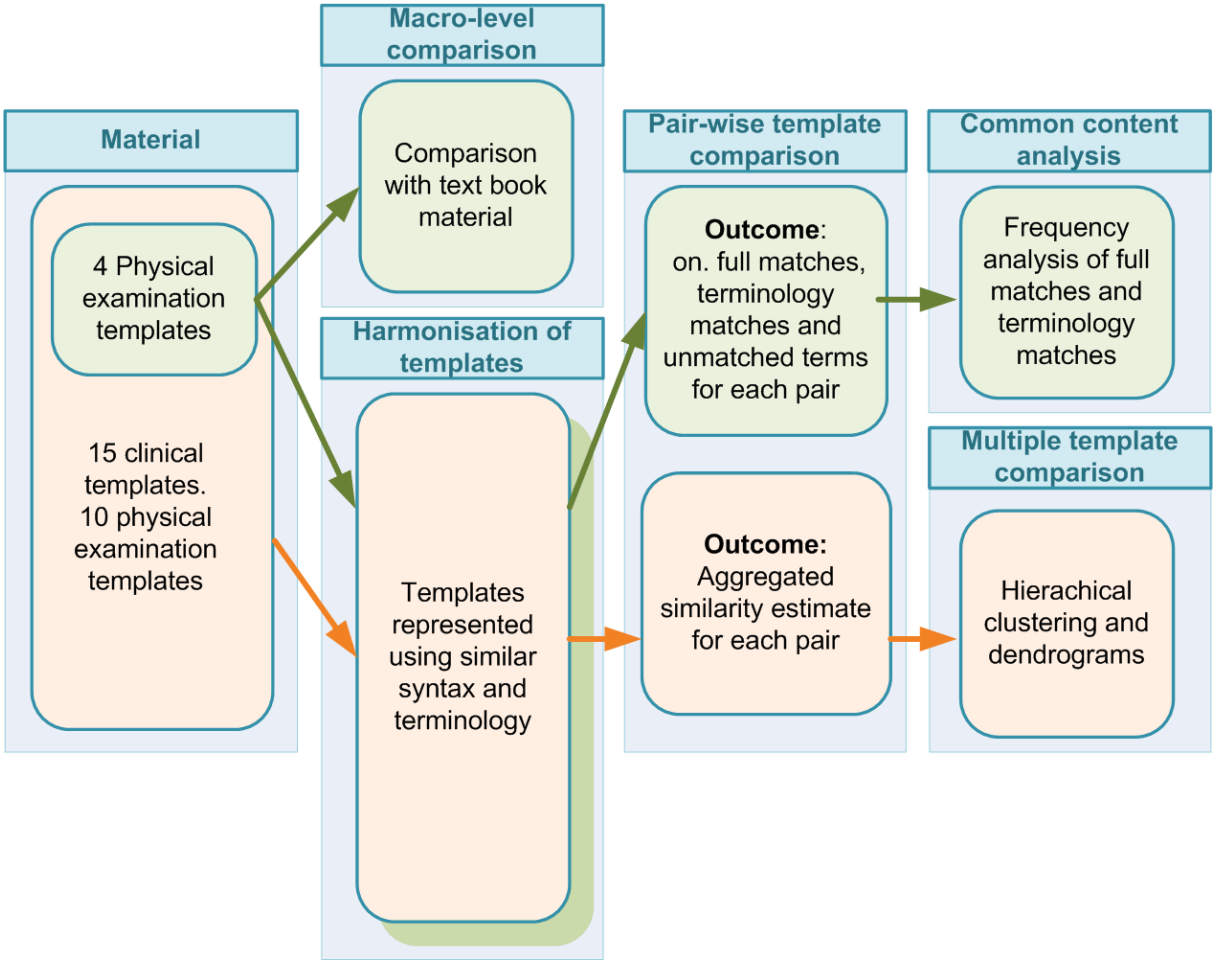

Figure 9 – Analysis methods used in paper IV (green arrows) and V (orange arrows) respectively.

In paper IV, four physical examination templates from different organizations were analyzed on a macro level to study the clinical understanding of physical examinations compared with EHR-templates. We performed a template harmonization using SNOMED CT and a clinical content format (Developed in paper I and II). The templates were compared in terms of number of full matches, terminology matches and unmatched terms. At last, common content was analyzed on the basis of full matches and terminology matches. As a result, we were able to show the overlap between different organizations documentation of physical examination. In addition, we suggested what semantic elements a common physical examination could consist of.

In paper V, 15 physical examination templates were harmonized (as above) and compared pair-wise using four different semantic similarity estimation techniques[1]. Each of the four techniques resulted in a hierarchical clustering of the 15 templates illustrated as dendrograms. Consequently, it was possible to compare the four techniques and to get an overview of similarities and differences.

[1] Technical details regarding semantic similarity estimation is described in the background section of paper V.

In paper V, we refined some of the methodological elements of paper IV. The pair-wise comparison was automated using semantic similarity estimation. This made it possible to compare larger set of templates than in paper IV and using hierarchical clustering and dendrograms. Moreover, the manual labor was reduced compared to the analysis method in paper IV.

4 Papers

In the following chapter, the scientific papers that constitute the core part of the PhD study is enclosed in the full version of the thesis.

5 Conclusion

This PhD study contributes to the medical informatics knowledge domain by answering the two research questions:

Q1: What are the prerequisites of applying semantic methods for local clinical model comparison?
Q2: Which semantic approaches can be applied to local clinical model comparison?

A presentation of future applications and further research concludes this PhD thesis.

5.1 Prerequisites of applying semantic methods for local clinical model comparison

The studied local clinical models are expressed using vendor-dependent formalisms and local interface terminology. Consequently, if model comparison should be possible, it required a consistent representation of the clinical models on the model-formalism level as well as on the semantic level. Methods for consistent representation at formalism and semantic level were developed in paper I, II and III. In Paper I, a model formalism was developed. In paper II, a guideline for consistent semantic representation of interface terminology with SNOMED CT was developed. In paper III, a semantic inter-rater agreement measure was developed. Further concluding remarks are presented in the following.

In Paper I, it was shown that local clinical models, i.e. local templates, have a formalism which is dependent on the EHR systems. In some cases, the EHR owners had access to their templates via a configuration system, and in other cases they had access to the templates via XML files. This meant that the local clinical models had different formalisms and in some cases, they were inaccessible. Consequently, collecting template information and expressing it uniformly to allow template comparison was a challenge. Taking on the challenge, internal requirement specification material and dumps of user interfaces from different EHR units at regional or hospital level were collected, and a simple template model expressed as UML class diagrams was developed and evaluated [92]. The formalism was named the clinical content format. The idea was not to create an information model to replace local formalisms or standardised clinical models like openEHR/13606 templates and archetypes or Goosen DCMs [60] which are intended to specify the data and architecture related information necessary to build an EHR system. Merely, the clinical content format specifies a use case view of a template. The format express the order of fields on the user interface, the general data type, the hierarchy of the fields, the content of lists (e.g. drop down menus) and the interface terminology with the possibility to link to standardised terminology. A clinical content format will be obviated in a setting where standardised clinical models can be exported directly from EHR systems as in [93]. However, to this date, none of the EHR owners have access to their local clinical models expressed in a non-vendor specific manner. This lack of standardisation means that the clinical content format is still relevant in a template comparison context. In this PhD study, the format made it possible to build a database that could hold information about templates from multiple organisations and multiple EHR systems in a uniform fashion. Clinical model comparison depends on formalism as explained above, but it also depends on semantics.

In Paper II and throughout the PhD study, SNOMED CT has been used as a semantic reference making it possible to bridge between the context dependent ways of expressing the meaning of fields and list-items in templates. Firstly, context dependent refers to language as both Swedish and Danish templates have been analysed. Secondly, context dependent refers to the variability of clinical terminology regardless of

language, rooted in the simple fact that clinicians disagree on what to call things. An example of Alan Rector goes as follows:

"...over the past three decades the use of the word 'neoplastic' and its cognates in various European languages has meant at various times and places either literally 'new growth', i.e. 'any cellular proliferation, benign or malignant', or 'malignant proliferation'. Interminable discussions have occurred. However, at no time has there been any question that both underlying concepts were valid, only of which linguistic construct ought to be applied to each."[14]

Rector continues with the argumentation that separating language and concepts would solve the problem, and this is exactly the aim of mapping the interface terminology of all template content to SNOMED CT in this PhD study. However, mapping interface terminology to SNOMED CT is not a trivial task. Mapping consistency is crucial because template comparison findings would be misinterpreted if variability due to mapping inconsistency were mistaken for template differences. In paper II, mapping guidelines which support mapping consistency were developed. In the mapping guideline study it is argued that due to the current completeness of SNOMED CT, it is not meaningful to base concept comparison and information retrieval on fully defined terms. Consequently, the primary SNOMED CT hierarchy expressed by IS-A relations is used, which gives rise to a set of guidelines which ensures that interface terms representing similar concepts are mapped to the same SNOMED CT hierarchies regardless of the template in which the interface terms occur. Similar concept does not mean "the same" concept, but concepts which are linked through the relationships in SNOMED CT. Using an example from Paper II [94]: 'blood pressure finding', 'pulse rate finding' and 'finding of peripheral pulse' are linked by IS-A relations because they are all children of 'Cardiac finding'. If some of the concepts were drawn from other hierarchies the mapping would be inconsistent e.g. 'observation of blood pressure' from the 'observable entities hierarchy'. The guideline described in Paper II, could be understood as a terminology binding approach because it links the clinical models to SNOMED CT. Our early work shows, that the interface between models and terminology is a known challenge [95]. Concurrent design of clinical models and terminology might improve SNOMED CT coverage and consistency further.

Paper III proposed a semantic inter-rater agreement score. Mapping consistency can be evaluated and improved only if meaningful inter-rater agreement scores are available. Inter-rater agreement is important in a template comparison context because it measures mapping consistency between coders. As previously stated, template comparison findings would be misinterpreted if variability due to mapping inconsistency were mistaken for template differences. The proposed semantic interrater agreement score was based on Krippendorff's α [96] as an alternative to nominal interrater agreement scores which have been used in earlier SNOMED CT mapping studies. The difference between a nominal and a semantic score is that the nominal score will evaluate whether two codes are the same or not, whereas the semantic score introduces a difference function expressing how different two codes are. In a SNOMED CT context, 'How different' can be expressed dependent on closeness of SNOMED CT concepts based on IS-A relations. In Paper III, a path length based measure was chosen, but other semantic similarity estimates could be considered as alternatives e.g. those proposed by Sanchez et al. [97]. In the study, it was shown that when using partially constructed datasets the proposed semantic α was more sensitive to semantic inconsistencies than a nominal agreement score (nominal Krippendorff's α). This means that the semantic α is promising in terms of evaluating the consistency of SNOMED CT coding. Moreover, the dependency on IS-A relations means

that the semantic α is consistent with the mapping guidelines. This means that coders mapping in agreement with the mapping guidelines theoretically would yield a high semantic alpha score. Inter-rater agreement has not been applied for the compared templates because the interface terms of the templates were mapped at the same time as the guidelines were developed. Therefore, the way terms were mapped was an ongoing discussion and the dataset could not be subjected to interrater agreement calculations.

5.2 Semantic approaches to local clinical model comparison

In this PhD study, two methods for semantic clinical model comparison were developed and evaluated. The methods both included structuring local clinical models in accordance with the developed model formalism and in accordance with the SNOMED CT mapping guidelines, as explained in Section 5.1. In addition to consistent representation, the quality and applicability of semantic template comparison approaches relies on the quality of SNOMED CT, the comparison method, and the level of automation. Our studies showed that semantic approaches to local clinical model comparison is possible and can lead to insight into the extent of the semantic overlap between local clinical models. The specific methodological differences are explained in section 3.1, and they are not repeated here. Further concluding remarks are presented in the following.

In paper IV, we showed that the data elements of four physical examination templates from four different organisations in Denmark and Sweden could be compared using the proposed method. All six pair-wise comparisons had exact matches, terminology matches and unrelated fields, and cross-country comparisons revealed more unrelated content than in-country comparisons. The semantic overlap between templates was analysed on the basis of exact matches and terminology matches. The semantic overlap i.e. a core set of concepts representing the analysed template could be utilized in clinical model standardization efforts, or to reveal unintended overlap in clinical model libraries, as in [98].

In paper V, we analysed 15 clinical templates hereof 10 physical examination templates from five different organisations. In paper V, the manual identification of exact matches, terminology matches and unrelated fields from paper IV was replaced by an automatic pair-wise comparison based on semantic similarity estimates. Two different node-based similarity estimates and two different aggregation methods were evaluated in terms of their ability to cluster physical examination templates in a hierarchical clustering approach. The results showed that the aggregation method meant more for the clustering than which similarity estimate was used. The methods that performed best had the ability to cluster Danish and Swedish physical examination templates in two distinct clusters, which were joined with each other before clinical models covering other clinical areas (e.g. COPD follow up) were included. We concluded that hierarchical clustering of templates could be a valuable tool for comparison and summarization of multiple templates.

5.2.1 Comparison of semantic approaches in paper IV and V

Even though there are many common elements between the two analysis studies, it should be kept in mind that the focus and aim was different. Actually, the study disseminated in paper V came from concerns raised after finishing paper IV. The concerns were that the amount of manual work would discourage others from applying the method and that pair-wise comparison was not enough to get an overview of similarities and differences of templates because the number of comparisons would increase exponentially with the number of templates compared. These shortcomings were addressed in the second study.

Automation was possible because of the similarity estimation approach, and overview of a higher number of templates became possible by using hierarchical clustering and dendrograms. However, this does not mean that the second study is a better version of the first study, merely that they each have elements that solve different problems. The similarity estimation approach makes it possible to get an overview of large amounts of diverse templates and choose a relevant subset of these for further processing. The semi-automatic method provides qualitative insight on a reduced number of templates and makes it possible to do common content analysis. Rethinking the methodology so that both overview and detailed analysis/common content analysis can be done in one workflow and providing adequate tooling will be a future research objective. When rethinking methodology, the use of SNOMED CT as semantic reference in template comparison should also be considered.

5.2.2 The role of SNOMED CT and similarity estimation

The SNOMED CT related concerns are the quality and completeness of SNOMED CT, handling of post-coordination and the underlying methodology of similarity estimation

Regarding the first concern of quality and completeness of SNOMED CT, it applies for both methods that the comparisons cannot be of a better quality than the model on which it is based. This means that missing or wrongful concepts and relationships in SNOMED CT affect the result. One example taken from paper IV [99] is that 'electrocardiogram finding' does not have a relationship to 'cardiovascular finding'. In paper IV, this means that 'electrocardiogram finding' does not participate in any of the terminology-similarities even though the concept 'cardiovascular finding' occurs in other templates. In paper V, it means that for templates with the concept 'electrocardiogram finding', the similarity estimation method would be likely to underestimate the actual similarity. Terms not mapped to SNOMED CT because an adequate concept could not be found also affects the result. Quality and completeness of SNOMED CT has been an area of concern [100,101]; however, the idea of a complete terminology is unrealistic and costly [102]. Consequently, the interesting question to answer in this context is probably not whether SNOMED CT is complete, but if it is comprehensive enough to be used as a semantic reference in a template comparison perspective. The two comparison studies have shown that despite the shortcomings of the model and a simplistic view on post-coordinated expressions valuable insight on template similarity and differences can be gained from a SNOMED CT based analysis.

We have chosen very simple ways of dealing with the post-coordinated expressions. In paper IV, post-coordinated expressions were counted as one of the types of terminology similarities. In paper V, they were atomized and each concept was treated separately. The latter could be a theoretical concern because a post-coordinated expression representing one field in a template would count for two or more concepts in the analysis whereas pre-coordinated expression would only count for one. This could be solved by simply weighing the similarity estimates accordingly. The problem is that neither of these techniques, in an accurate manner, places the post-coordinated expression in a meaningful place in the SNOMED CT hierarchy, and consequently, the similarity estimate will not be accurate because it is based on IS-A relations. Alternatively, similarity in a SNOMED CT context could be based on the defining relationships of an expression, which would equate post-coordinated expressions with pre-coordinated expressions because each of them would be a summation of the concepts that defines them. Even though using other relationships than IS-A has been proposed by others, e.g. Batet et al. in their discussion section [103], no relatedness or similarity estimates exists yet that can be based on differences in defining relationships.

Other alternatives do exist to the intrinsic IC-based similarity estimation used in Paper V e.g. those based on other taxonomical features like path and taxonomical depth or those based on corpora [97,104], and evaluating more of these in a template comparison context could be valuable. Especially when having in mind that the current evaluation using the Lin similarity estimate and the Sokal and Sneath similarity estimate was in-conclusive in terms of which similarity estimate was most suitable. Recently, McInnes and Pedersen evaluated which similarity estimate was most suitable in a word sense disambiguation context [105]. A similar method could be used for the template comparison context.

5.3 Tooling to allow application of clinical model comparison

The focus of paper V was on getting an overview of many templates whereas the focus of paper IV was concerning details of a few templates. Both overview and detailed analysis would be important elements if the method should be applicable in local, national and international information management organisations[2]. Furthermore, the analysis should be supported by adequate tooling so that the workflow will be acceptable in information management organisations. This includes that the artefacts developed throughout the PhD study (see Appendix C) ideally should be redesigned as one software package.

Re-design would require that one database was implemented where both template information and SNOMED CT information could be included. Moreover, calculated similarity estimates and possibly other features from the template comparison could be considered implemented in the database. The benefit of storing features would be better response time in the software because the similarity calculations are resource intensive and thus should not be repeated. Shortcomings would be the risk of inconsistent data due to derived features being stored. Consequently, implementation of new SNOMED CT releases in the database would also require update of the derived features. In addition to making *one* database, *one* software component should be developed.

One approach would be to further develop the existing template similarity estimation software. Key features are described in Figure 10. The features that need to be developed or refined are:

Multiple template comparison can be based on the existing m-function, but it has to be refined so that template subsets can be chosen from the dendrogram. The dendrogram functionality might have to be improved for large sets of templates. One idea could be the dendrogram stacking approach [106] .

Common content analysis based on similarity estimates should be developed. The existing method is described in paper IV, but the method is manual. Developing software support based on similarity estimates would provide users with ideas for semantic content of new designs based on a quantitative analysis of multiple templates. The common content analysis can be used as point of departure for comparing templates that work in clinical practise with clinical evidence of best-practice.

Visualisation of pair-wise template similarity can be based on the similarity estimates and shown as a matrix. However, extending the functionality with SNOMED CT visualisation of the two templates would improve the insight in the actual semantic differences in a set. Visualisation could be inspired by our earlier work regarding set-of-concept visualisation [107].

[2] The potential application in a local, national and international setting is described in the introduction of paper V

The overview of an information system for clinical model comparison illustrated in Figure 10 is explained in the following. The greenish-blue boxes denote system functionality. The violet boxes denote user-activities which can be supported by system functionality. The gray boxes are elements already developed, but which are not adding value to the user i.e. the information modeller. The green boxes are elements which would be valuable in a bottom-up analysis and design process. Overview of and insight in existing templates can be gained from the two visualisations: Multiple and Pair-wise. Based on the dendrogram the user can choose to delimit the templates, e.g. only include templates that form well-defined clusters in further analysis. When a subset is chosen, in depth analysis can be made using the common content analysis and comparing common content with textbook and guidelines.

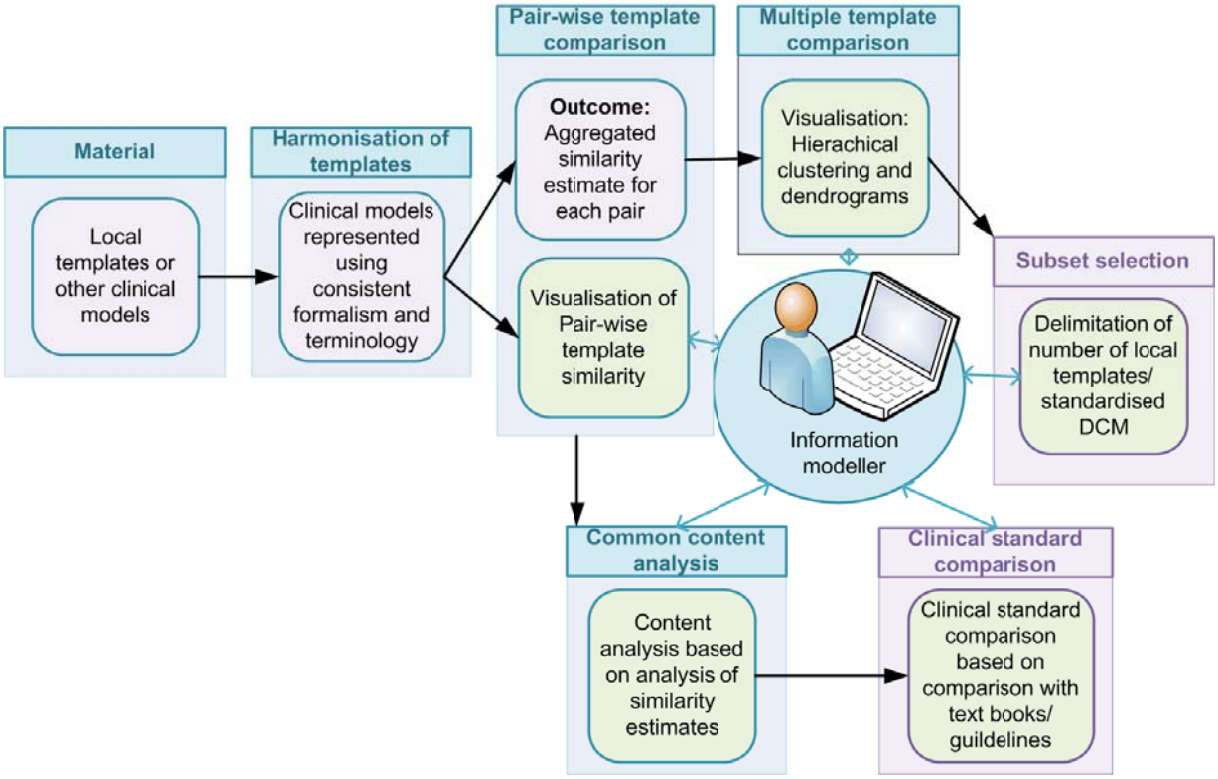

Figure 10 - Use of clinical model comparison software.

5.4 Further research perspectives

Evaluation of the clinical model comparison methods in a theoretic and a real-life setting is the logical continuation of this PhD study. A theoretical feasibility evaluation would include modelling standardised clinical models on the basis of local clinical model comparison e.g. openEHR archetypes could be modelled on the basis of a semantic comparison of local models. In such a clinical model development project, it should be further analysed how to handle granulation differences between existing models. In a practical setting, an EHR implementation project or standardisation initiatives, such as CIMI, could try to use a library of existing clinical models and the comparison methods to support design of new clinical models. This type of evaluation would show how users perceive usefulness of the clinical model comparison methods.

Evaluation would further contribute to solve the challenge that initiated this PhD-study i.e. the fragmentation of centralized and de-centralized EHR-system initiatives. The clinical model comparison methodologies contribute to bridge centralised and decentralised approaches because they provide a mechanism for harmonization of multiple clinical models so that they become relevant in a standardization context.

6 References

[1] Kalra D, Ingram D. Electronic health records. Information Technology Solutions for Healthcare 2006:135-181.

[2] Connected Digital Health in Denmark. the National Strategy for the Digitalization of the Health Care Sector 2008-2012. 2007.

[3] Department of Health. Health informatics review report
. 2008.

[4] Schaeffer LD, Schultz AM, Salerno JA. HHS in the 21st Century: Charting a New Course for a Healthier America. : National Academies Press; 2009.

[5] Greenhalgh T, Potts HWW, Wong G, Bark P, Swinglehurst D. Tensions and Paradoxes in Electronic Patient Record Research: A Systematic Literature Review Using the Meta-narrative Method. The Milbank quarterly 2009;87(4):729.

[6] Ingram D, Kalra D, Austin T, Darlison MW, Modell B, Patterson D. Towards an interoperable healthcare information infrastructure—working from the bottom up. BT Technology Journal 2006;24(3):17-32.

[7] Information Society - ICT for Health - Information Centre. Available at: http://ec.europa.eu/information_society/activities/health/glossary_of_terms/. Accessed 7/3/2012, 2012.

[8] European Communities. Semantic Interoperability for Better Health and Safer Healthcare. 2009.

[9] Lopez DM, Blobel BG. A development framework for semantically interoperable health information systems. International journal of medical informatics 2009;78:83-103.

[10] Lopez DM, Blobel B. Enhanced semantic interoperability by profiling health informatics standards. Methods of information in medicine 2009;48:170-7.

[11] Wollersheim D, Sari A, Rahayu W. Archetype-based electronic health records: a literature review and evaluation of their applicability to health data interoperability and access. The HIM journal 2009;38(2):7-17.

[12] T. Beale. Archetypes: Constraint-based domain models for future-proof information systems. OOPSLA 2002 workshop on behavioural semantics; 2002.

[13] Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods of information in medicine 1998;37:394-403.

[14] Rector AL. Clinical terminology: why is it so hard? Methods of information in medicine 1999;38:239-52.

[15] Rector AL. The interface between information, terminology, and inference models. Studies in health technology and informatics 2001;84:246-50.

[16] Rector A, Qamar R, Marley T. Binding ontologies and coding systems to electronic health records and messages. Applied Ontology 2009;4(1):51-69.

[17] Sundvall E, Qamar R, Nyström M, Forss M, Petersson H, Karlsson D, et al. Integration of tools for binding archetypes to SNOMED CT. BMC Medical Informatics and Decision Making 2008;8(Suppl 1):S7.

[18] Goossen WT, Goossen-Baremans A. Bridging the HL7 template - 13606 archetype gap with detailed clinical models. Studies in health technology and informatics 2010;160(Pt 2):932-936.

[19] Stroetmann V, Jung B, Rodrigues J, Hammerschmidt R. Infrastructure, connectivity, interoperability – inventory of key relevant Member States and international experience. European Commission 2007.

[20] Hoerbst A, Ammenwerth E. Electronic Health Records. A Systematic Review on Quality Requirements. Methods of information in medicine 2010 Jul 6;49(4).

[21] Ingenerf J, Giere W. Concept-oriented standardization and statistics-oriented classification: continuing the classification versus nomenclature controversy. Methods of information in medicine 1998;37:527-39.

[22] Chen R. Towards Interoperable and Knowledge-Based Electronic Health Records Using Archetype Methodology. PhD thesis 2009.

[23] Garde S, Knaup P. Requirements engineering in health care: the example of chemotherapy planning in paediatric oncology. Requirements Engineering 2006;11(4):265-278.

[24] Lenz R, Beyer M, Kuhn KA. Semantic integration in healthcare networks. International journal of medical informatics 2007;76(2-3):201-207.

[25] Prokosch HU, Ganslandt T. Perspectives for medical informatics. Reusing the electronic medical record for clinical research. Methods of information in medicine 2009;48:38-44.

[26] Hanseth O, Monteiro E, Hatling M. Developing information infrastructure: The tension between standardization and flexibility. Science, technology & human values 1996;21(4):407-426.

[27] Knaup P, Bott O, Kohl C, Lovis C, Garde S. Electronic patient records: moving from islands and bridges towards electronic health records for continuity of care. Yearb Med Inform 2007:34-46.

[28] Blobel B, Pharow P. Analysis and evaluation of EHR approaches. Methods of information in medicine 2009;48:162-9.

[29] E. S. Chen, S. Manaktala, I. N. Sarkar and G. B. Melton. A Multi-Site Content Analysis of Social History Information in Clinical Notes. AMIA Annual Symposium Proceedings: American Medical Informatics Association; 2011.

[30] Håkonsen S, Madsen I, Bjerrum M, Pedersen PU. Danish National Framework for collecting information about patients' nutritional status. Nursing Minimum dataset (N-MDS). Online Journal of Nursing Informatics (OJNI) 2012;16(3).

[31] Nowlan WA. Clinical workstations: identifying clinical requirements and understanding clinical information. International journal of bio-medical computing 1994 Jan;34(1-4):85-94.

[32] Essin DJ, Dishakjian R, deCiutiis VL, Essin CD, Steen SN. Development and assessment of a computer-based preanesthetic patient evaluation system for obstetrical anesthesia. Journal of clinical monitoring and computing 1998;14(2):95-100.

[33] Hyun S, Johnson SB, Stetson PD, Bakken S. Development and evaluation of nursing user interface screens using multiple methods. Journal of Biomedical Informatics 2009;42(6):1004-1012.

[34] Scandurra I, Hägglund M, Koch S. From user needs to system specifications: Multi-disciplinary thematic seminars as a collaborative design method for development of health information systems. Journal of Biomedical Informatics 2008 8;41(4):557-569.

[35] Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. The Lancet 2003;362(9391):1225-1230.

[36] Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al. t physicians follow clinical practice guidelines? A framework for improvement. JAMA : the Journal of the American Medical Association 1999;282:1458-65.

[37] Pearson SA, Moxey A, Robertson J, Hains I, Williamson M, Reeve J, et al. Do computerised clinical decision support systems for prescribing change practice? A systematic review of the literature (1990-2007). BMC health services research 2009 Aug 28;9:154.

[38] Chen R, Georgii-Hemming P, Ahlfeldt H. Representing a Chemotherapy Guideline Using openEHR and Rules. Studies in health technology and informatics 2009;150:653.

[39] Hoy D, Hardiker NR, McNicoll IT, Westwell P, Bryans A. Collaborative development of clinical templates as a national resource. International journal of medical informatics 2009;78 Suppl 1:S95-100.

[40] Sebastian Garde, Rong Chen, Heather Leslie, Thomas Beale, Ian McNicoll and Sam Heard. Archetype-based knowledge management for semantic interoperability of electronic health records. MIE; 2009.

[41] Clinical Knowledge Manager. Available at: http://www.openehr.org/ckm/. Accessed 8/8/2014, 2014.

[42] Kalra D. Electronic health record standards. Yearb Med Inform 2006:136-44.

[43] Eichelberg M, Aden T, Riesmeier J, Dogac A, Laleci GB. A survey and analysis of electronic healthcare record standards. ACM Computing Surveys (CSUR) 2005;37(4):277-315.

[44] Anani N, Chen R, Moreira TP, Koch S. Retrospective checking of compliance with practice guidelines for acute stroke care: a novel experiment using openEHR's Guideline Definition Language. BMC medical informatics and decision making 2014;14(1):39.

[45] FHIR Specification Home Page - FHIR v0.0.81. Available at: http://www.hl7.org/implement/standards/fhir/index.html. Accessed 7/3/2014, 2014.

[46] Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nature Reviews Genetics 2012;13(6):395-405.

[47] Bird L, Goodchild A, Tun ZZ. Experiences with a two-level modelling approach to Electronic Health Records. Journal of Research and Practice in Information Technology 2003;35:121-138.

[48] Rector AL, Nowlan WA, Kay S, Goble CA, Howkins TJ. A framework for modelling the electronic medical record. Methods of information in medicine 1993 Apr;32(2):109-119.

[49] Johnson SB. Generic data modeling for clinical repositories. Journal of the American Medical Informatics Association : JAMIA 1996 Sep-Oct;3(5):328-339.

[50] Martínez-Costa C, Menárguez-Tortosa M, Fernández-Breis JT. An approach for the semantic interoperability of ISO EN 13606 and OpenEHR archetypes. Journal of Biomedical Informatics 2010;43(5):736-746.

[51] Welcome to openEHR - openEHR :: future proof and flexible EHR specifications. Available at: http://www.openehr.org/home.html. Accessed 6/21/2012, 2012.

[52] The CEN/ISO EN13606 standard. Available at: http://www.en13606.org/the-ceniso-en13606-standard. Accessed 8/12/2014, 2014.

[53] Relationship between FHIR and CDA - FHIR v0.0.81. Available at: http://www.hl7.org/implement/standards/fhir/comparison-cda.html. Accessed 8/13/2014, 2014.

[54] Relationship between FHIR and v3 Messaging - FHIR v0.0.81. Available at: http://www.hl7.org/implement/standards/fhir/comparison-v3.html. Accessed 8/13/2014, 2014.

[55] Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, et al. HL7 Clinical Document Architecture, Release 2. Journal of the American Medical Informatics Association : JAMIA 2006;13:30-9.

[56] DSTU Considerations - FHIR v0.0.81. Available at: http://www.hl7.org/implement/standards/fhir/dstu.html. Accessed 8/13/2014, 2014.

[57] Summary (Glossy) - FHIR v0.0.81. Available at: http://www.hl7.org/implement/standards/fhir/summary.html. Accessed 8/13/2014, 2014.

[58] The Clinical Information Modeling Initiative | AMIA. Available at: http://www.amia.org/the-standards-standard/2012-volume3-edition1/clinical-information-modeling-initiative. Accessed 4/17/2013, 2013.

[59] About | CIMI. Available at: http://opencimi.org/node/20. Accessed 8/18/2014, 2014.

[60] Goossen W. Using detailed clinical models to bridge the gap between clinicians and HIT. Stud Health Technol Inform 2008;141:3-10.

[61] Huff SM, Rocha RA, Coyle JF, Narus SP. Integrating detailed clinical models into application development tools. Studies in health technology and informatics 2004;107(Pt 2):1058-1062.

[62] Goossen W, Goossen-Baremans A, Van Der Zel M. Detailed clinical models: a review. Healthcare informatics research 2010;16(4):201.

[63] Chen R, Enberg G, Klein GO. Julius--a template based supplementary electronic health record system. BMC Medical Informatics and Decision Making 2007;7:10.

[64] Ahn S, Huff SM, Kim Y, Kalra D. Quality metrics for detailed clinical models. International journal of medical informatics 2013;82(5):408-417.

[65] Blobel B, Goossen W, Brochhausen M. Clinical Modeling–a Critical Analysis. International journal of medical informatics 2013.

[66] R. Qamar and A. Rector. Semantic mapping of Clinical Model Data to Biomedical Terminologies to Facilitate Data Interoperability. proceedings of the Healthcare Computing 2007 Conference: 19–21 March 2007 Harrogate; UK; ; 2007.

[67] MARTÍNEZ-COSTA C, KALRA D, SCHULZ S. Improving EHR Semantic Interoperability Future Vision and Challenges.

[68] Elkin PL, Trusko BE, Koppel R, Speroff T, Mohrer D, Sakji S, et al. Secondary use of clinical data. Studies in health technology and informatics 2010;155:14-29.

[69] Bowker GC, Star SL. Sorting things out: Classification and its consequences. : MIT press; 1999.

[70] Cimino J. High-quality, standard, controlled healthcare terminologies come of age. Methods of information in medicine 2011;50(2):101.

[71] Amy Y. Wang, Jeremiah H. Sable and Kent A. Spackman. The SNOMED clinical terms development process: refinement and analysis of content. Proceedings of the AMIA Symposium: American Medical Informatics Association; 2002.

[72] Tu SW, Bodenreider O, Çelik C, Chute CG, Heard S, Jakob R, et al. A content model for the ICD-11 revision. 2010.

[73] Nyström M, Vikström A, Nilsson GH, Åhlfeldt H, Örman H. Enriching a primary health care version of ICD-10 using SNOMED CT mapping. J.Biomedical Semantics 2010;1:7.

[74] Rosenbloom ST, Brown SH, Froehling D, Bauer BA, Wahner-Roedler D, Gregg WM, et al. Using SNOMED CT to represent two interface terminologies. Journal of the American Medical Informatics Association : JAMIA 2009;16:81-8.

[75] Wade G, Rosenbloom ST. The impact of SNOMED CT revisions on a mapped interface terminology: terminology development and implementation issues. Journal of Biomedical Informatics 2009;42:490-3.

[76] Wade G, Rosenbloom ST. Experiences mapping a legacy interface terminology to SNOMED CT. BMC medical informatics and decision making 2008;8(Suppl 1):S3.

[77] H. Wasserman and J. Wang. 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; 2003.

[78] J. C. McClay and J. Campbell. Improved coding of the primary reason for visit to the emergency department using SNOMED. Proceedings of the AMIA Symposium: American Medical Informatics Association; 2002.

[79] S. H. Brown, S. T. Rosenbloom, B. A. Bauer, et al. Direct Comparison of MEDCIN® and SNOMED CT® for Representation of a General Medical Evaluation Template. : American Medical Informatics Association; 2007.

[80] Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR. The content coverage of clinical classifications. Journal of the American Medical Informatics Association 1996;3(3):224-233.

[81] P. L. Elkin, S. H. Brown, C. S. Husser, et al. Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. Mayo Clinic Proceedings: Mayo Clinic; 2006.

[82] S. H. Brown, B. A. Bauer, D. L. Wahner-Roedler and P. L. Elkin. Coverage of Oncology Drug Indication Concepts and Compositional Semantics by SNOMED-CT®. AMIA Annual Symposium Proceedings: American Medical Informatics Association; 2003.

[83] Lee DH, Lau FY, Quan H. A method for encoding clinical datasets with SNOMED CT. BMC medical informatics and decision making 2010 Sep 17;10:53.

[84] Lee D, Cornet R, Lau F, De Keizer N. A survey of SNOMED CT implementations. Journal of Biomedical Informatics 2012.

[85] R. H. Dolin, K. A. Spackman and D. Markwell. Selective retrieval of pre-and post-coordinated SNOMED concepts. Proceedings of the AMIA Symposium: American Medical Informatics Association; 2002.

[86] K. A. Spackman. Normal forms for description logic expressions of clinical concepts in SNOMED RT. Proceedings of the AMIA Symposium: American Medical Informatics Association; 2001.

[87] Andrews JE, Richesson RL, Krischer J. Variation of SNOMED CT coding of clinical research concepts among coding experts. Journal of the American Medical Informatics Association 2007;14(4):497-506.

[88] M. F. Chiang, J. C. Hwang, C. Y. Alexander, D. S. Casper, J. J. Cimino and J. Starren. Reliability of SNOMED-CT coding by three physicians using two terminology browsers. AMIA Annual Symposium Proceedings: American Medical Informatics Association; 2006.

[89] Smith B, Brochhausen M. Putting biomedical ontologies to work. Methods Inf Med 2010;49(2):135-140.

[90] Dugas M, Fritz F, Krumm R, Breil B. Automated UMLS-Based Comparison of Medical Forms. PloS one 2013;8(7):e67883.

[91] Hevner AR, Salvatore TM, Park J, Ram S. Design Science in information Systems Research. MIS Quarterly 2004;28.

[92] Rosenbeck KH, Randorff Rasmussen A, Elberg PB, Andersen SK. Balancing centralised and decentralised EHR approaches to manage standardisation. Studies in health technology and informatics 2010;160(Pt 1):151-155.

[93] Chen R, Klein GO, Sundvall E, Karlsson D, Ahlfeldt H. Archetype-based conversion of EHR content models: pilot experience with a regional EHR system. BMC Medical Informatics and Decision Making 2009;9:33.

[94] Højen AR, Gøeg KR. SNOMED CT Implementation. Mapping Guidelines Facilitating Reuse of Data. Methods of information in medicine 2011;50(5):472-478.

[95] Pape-Haugaard LB, Rasmussen AR, Rosenbeck KH, Andersen SK. Pitfalls when integrating terminology systems and EHRs. In: Ann Bygholm, Pia Elberg, Ole Hejlesen, editors. SHI2010 Proceedings,: TAPIR Akademisk Forlag; 2010. p. 34-38.

[96] Krippendorff K. Estimating the Reliability, Systematic Error and Random Error of Interval Data. Educational and Psychological Measurement 1970 April 01;30(1):61-70.

[97] Sánchez D, Batet M. Semantic similarity estimation in the biomedical domain: An ontology-based information-theoretic perspective. Journal of Biomedical Informatics 2011;44(5):749-759.

[98] Allones JLI, Taboada M, Martinez D, Lozano R, Sobrido MJ. SNOMED CT module-driven clinical archetype management. Journal of Biomedical Informatics 2013;46(3):388-400.

[99] Gøeg KR, Chen R, Højen AR, Elberg PB.
Content analysis of physical examination templates in electronic health records using SNOMED CT. Submitted September 2012. International Journal of Medical informatics 2014.

[100] Schulz S, Cornet R, Spackman K. Consolidating SNOMED CT's ontological commitment. Applied Ontology 2011;6(1):1-11.

[101] Rector A, Iannone L. Lexically suggest, logically define: Quality assurance of the use of qualifiers and expected results of post-coordination in SNOMED CT. Journal of Biomedical Informatics 2012;45(2):199-209.

[102] Rogers J. Quality assurance of medical ontologies. Methods of Information in Medicine-Methodik der Information in der Medizin 2006;45(3):267-274.

[103] Batet M, Sánchez D, Valls A. An ontology-based measure to compute semantic similarity in biomedicine. Journal of Biomedical Informatics 2011 2;44(1):118-125.

[104] Pedersen T, Pakhomov SV, Patwardhan S, Chute CG. Measures of semantic similarity and relatedness in the biomedical domain. Journal of Biomedical Informatics 2007 Jun;40(3):288-299.

[105] McInnes BT, Pedersen T. Evaluating measures of semantic similarity and relatedness to disambiguate terms in biomedical text. Journal of Biomedical Informatics 2013;46(6):1116-1124.

[106] Gilles Bisson and Renaud Blanch. Improving visualization of large hierarchical clustering. Information Visualisation (IV), 2012 16th International Conference on: IEEE; 2012.

[107] Højen AR, Sundvall E, Gøeg KR. Methods and applications for visualization of SNOMED CT concept sets. Applied Clinical Informatics 2014;5(1):127-152.

[108] Orlikowski WJ, Barley SR. Technology and institutions: What can research on information technology and research on organizations learn from each other? MIS quarterly 2001;25(2):145-165.

[109] Kuhn TS. The structure of scientific revolutions. : University of Chicago press; 1962.

[110] Burrell G, Morgan G. Sociological paradigms and organisational analysis. : Heinemann London; 1979.

[111] Hookway C. Pragmatism (Stanford Encyclopedia of Philosophy). 2008; Available at: http://plato.stanford.edu/entries/pragmatism/. Accessed 11/1/2010, 2010.

[112] McKay J, Marshall P, Hirschheim R. The design construct in information systems design science. Journal of Information Technology 2012;27(2):125-139.

[113] Orlikowski WJ, Iacono CS. Research Commentary: Desperately Seeking the" IT" in IT Research—A Call to Theorizing the IT Artifact. Information Systems Research 2001;12(2).

[114] Knaup P, Garde S, Haux R. Systematic planning of patient records for cooperative care and multicenter research. International journal of medical informatics 2007;76:109-17.

[115] Turner P, Milne G, Kubitscheck M, Penman I, Turner S. Implementing a wireless network of PDAs in a hospital setting. Personal and Ubiquitous Computing 2005;9:209-217.

[116] Trappenburg JC, Niesink A, de Weert-van Oene, Gerdien H, van der Zeijden H, van Snippenburg R, Peters A, et al. Effects of telemonitoring in patients with chronic obstructive pulmonary disease. Telemedicine and e-Health 2008;14(2):138-146.

A. Appendix: Philosophy of science underpinning and its implications

This PhD project is placed in between the IT-engineering and social sciences with a strong emphasis on IT-design. Introducing a new design can be understood as:

> What would be possible from a technological viewpoint, if a certain design is chosen?
> How it is possible to implement the design technically and organizationally?
> What practical advantages and disadvantages arise?

Answering these questions would not be possible if either a technological *or* a sociological viewpoint was chosen. The argument resembles Orlikowski's temporal observation – having a technology oriented viewpoint will cause you to deduce future use from the material properties of the technology. From a sociological viewpoint you deduce future use of technology from earlier or contemporary circumstances.[108] Since design is about forming a future use of a new technology different from the current use of old technology but still acceptable for users - neither of these perspectives on their own will create a useful framework for designing IT. Consequently, IT-system or artefact design traditionally use a mixture of methods from both fields, creating confusion regarding the paradigmatically background of the research conducted.

In this PhD project the "technology – sociology" dilemma is:

IT-standardization research develops ideal models which are mostly evaluated on a theoretical level – the feasibility of introducing the standards in real systems and creating benefits for those using the systems are typically speculations deduced from the material properties of the IT-standards.

Research forming future use of technology based on current practice, interacts with users or analyses current paper-based systems and develops technology in iterative and participatory design-processes. This type of research tends to ignore the macro-scale IT-requirements i.e. scalability and interoperability with other systems. Moreover, the creative design solving the users' problems smarter than they can think of themselves are typically ignored.

Improving the insight on this dilemma requires an overview of the paradigms that exist and their implications for IT-design, as well as a deeper understanding of IT-design as a concept.

Paradigms and their implications in IT-design

In 1962 Thomas Kuhn formed the notion "paradigm" to describe the way science goes through different phases with paradigms replacing each other. His main argument was that one paradigm will replace another, when the assumptions forming the paradigm can no longer be defended [109]. A paradigm, the way Kuhn saw it, should not be mistaken with colleagues having all the same opinions regarding their research. A paradigm is merely agreeing on what quality criteria characterize good research.

In contrast to Kuhn's idea of paradigms replacing one another, Burrell and Morgan believe that within the social sciences the paradigms can co-exist, forming different traditions at the same point in time. Burrell and Morgan's matrix of paradigms consist on deciding upon four basic assumptions namely ontology, epistemology, human nature and methodology. [110]. However, in this thesis the methodology or research approach is handled in the method in section 3.

In design science the paradigmatic background can basically be either conventional and have roots in positivism or participatory and have roots in interpretivism. In conventional design the ontological assumption is, that multiple, contextually situated alternative world-states constitutes reality. In participatory design and interpretivism reality is constructed, which means that the socially useful design is formed through negotiation and sense-making [5].

The interpretive researcher's assumption about the socially-constructed reality leads to an epistemology where causality is rejected whereas the positivist assumes causality. In the discussion about human nature, the interpretivist assumes free-will and the positivist assumes determinism [110]. In a design science perspective, this has implications for the possibility of the artefact to be used in other contexts. The conventional designer will assume that the contextually situated world-state in which an artefact is used can be characterized and used to foresee the usefulness of an artefact in other similar characterized realities. The interpretivist will limit the use of the artefact to the organization in which it is designed. The conventional designer will view the design process as rational – an organizational problem leads to a certain solution and that solution operates the same way in similar organizations. An interpretivist would say that one organizations problem can be solved by multiple designs, dependent on the choices that the involved users make, and there is no guarantee that the solution can be used in other organizations.

The contrast between positivist and interpretivist is the theoretic explanation that the basic "technology – sociology" dilemma of this PhD-thesis even exist.

An alternative paradigm is pragmatism. The original pragmatic paradigm is linked to pragmatist thinkers like Peirce, James and Dewey. A later version was formulated in 1994 by Putnam, who identified four

characteristics of pragmatism namely the rejection of skepticism, the willingness to embrace fallibilism, the rejection of sharp dichotomies, and the primacy of practice. The rejection of skepticism was formulated as an objection to the Cartesian "method of doubt" which means that ideas can be proposed, without being obliged to critically doubt all the assumptions linked to the idea. Concequently, researchers within this paradigm accept the possibility that they might be wrong which is linked to fallibilism, but the pragmatist incorporates the idea that further investigations will lead to correction of error. The rejection of sharp dichotomies is a criticism of philosophy of science as a whole, as opposite paradigms i.e. positivists and interpretivists are fighting instead of finding pragmatic solutions. The primacy of practice means that the hypothesis is clarified by identifying its practical consequences [111].

The pragmatic paradigm is in a design science perspective because:

> It would be possible to acknowledge the free-will of users and design-team, which also means the creative power of these to bring forward IT-design ideas. The rejection of skepticism makes "creation" possible – as long as it acknowledged that the designers might be wrong.
> The rejection of sharp dichotomies makes it possible to comprehend the "technology – sociology" dilemma, since interpretivism and positivism is not seen as strong opponents. It would be possible to form a pragmatic design concept, where one organizations problem can be solved by multiple designs, dependent on the ideas and choices of design-team and users. Some of these choices are linked to the organization and for this part of the solution there is no guarantee that it can be used in other organizations. Some choices and ideas have a general value, and this part of the solution can be used in other organizations.
> The primacy of practice goes well with the engineering worldview, that is more concerned with "making things work", than building consistent theoretical frameworks. Design choices are made based on the best possible conception of the practical consequences of the design choices. Practical consequences can be understood as consequences to details in local clinical workflows and consequences to scalability and interoperability.

IT-design conceptualization

In information system research, there is a tendency to have a narrow conceptualization of design – typically as problem solving, product and/or process. As summarized by McKay et al. in other disciplines, design has been conceptualized in broader terms adding design as intention, planning, communication, user experience, value, professional practice and service. In an overview perspective, design is impacted from three knowledge domains namely "human sciences", "engineering" and "art and aesthetics". A conclusion from the paper is that more perspectives on design might lead to new valuable research [112].

The engineering perspective is the dominant design conceptualization used in this PhD project, meaning that the classic IS problem solving, product and process as explained in [91] and presented in appendix B provides a meaningful framework with one exception.

In Hevner et al.'s framework, the artefact can be designed with input from the domain to ensure relevance and from the knowledge base to ensure rigor. From that it should be possible to build artefacts, evaluate them and improve them in an iterative process [91]. The rigor of this model ignores the internal ability of the designer to invent creative solutions or to choose from a number of known solutions. The designer or

the design team impacts the design beyond the influence from a domain and knowledge base. In McKay et al.'s conceptualization this means introducing "design as intention" in the PhD project.

Another perspective is linked to the power of users. In a very simple presentation, the design process consists of an oscillation between domain-space and design-space, where requirements are formed in the domain-space and the solution is developed in the design-space. In some (older) literature, the users decide the requirements, and the design team decides the actual solution [31]. In participatory design the borders between users and design-team is blurred – the users influence the design-space much more direct. When designing complex macro-scale solutions such as IT standards the design-team develops the solution without much connection to the domain – or only involving a few experts. The design of complex macro-scale design e.g. design of the SNOMED CT concept model or the openEHR archetype definition language is so abstract that few users will be interested in or knowledgeable enough to participate.

PhD project approach

The paragraph above illustrates that the border between domain-space and design-space should be dependent on the properties of the designed artefact. In this PhD project there is a clash between the participatory and the macro-scale design viewpoint, making it a challenge to define the roles in the design-processes. In the approach chosen, the design of an IT-solution can be guided by a macro-scale design i.e. a standard. The macro-scale design guides all the design-choices constituting the solution or artefact. To manage this, design-choice are split in three phases: Suggest - Reformulate – Object. "Suggest" denotes an initial design-choice or requirement formulated by the user if necessary facilitated by the design-team. "Reformulate" denotes a reformulation of the initial choice by the design team, to make the choice in accordance with macro-scale design. "Object" denotes an evaluation of the choice by the users to validate that the reformulation did not alter the initial meaning. The IT-design framework of the PhD project is illustrated in Figure 11.

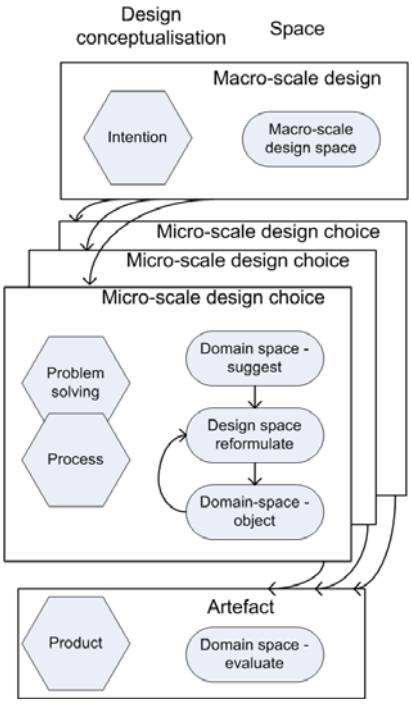

Figure 11 - The design-framework of the PhD-project. The figure illustrates the connection between macro-scale design, the micro-scale design choices and the artefact. The changing dominant design conceptualization in the framework is also described.

B. Appendix: Design Science in medical informatics

Hevner et al.'s approach to Design Science is illustrated in Figure 12. They argue that for design science research to be relevant it should be related to business needs and for it to be rigorous it should draw on the best knowledge available i.e. the scientific knowledge base. They stress that any system development should not aim at developing finished applications. On contrary, researchers should design like an artifact, which means that each component should be just finished enough to help answer the research questions. Hevner et al. defines the IT artefact as "the constructs, models, and methods applied in the development and use of information systems" and another place in paper it is added that artefact can be "intellectual or software tools aimed at improving the process of information system development" [91]. This is narrower than earlier definitions e.g. [113] because it is delimited from including people or organization related elements. The artefacts of this PhD project are presented in Appendix C. Constructing artefacts is not enough to comprise a research process. Hevner et al. describe that the research consists of an iterative

process where development, assessment, evaluation and refinement are key elements, see Figure 12.

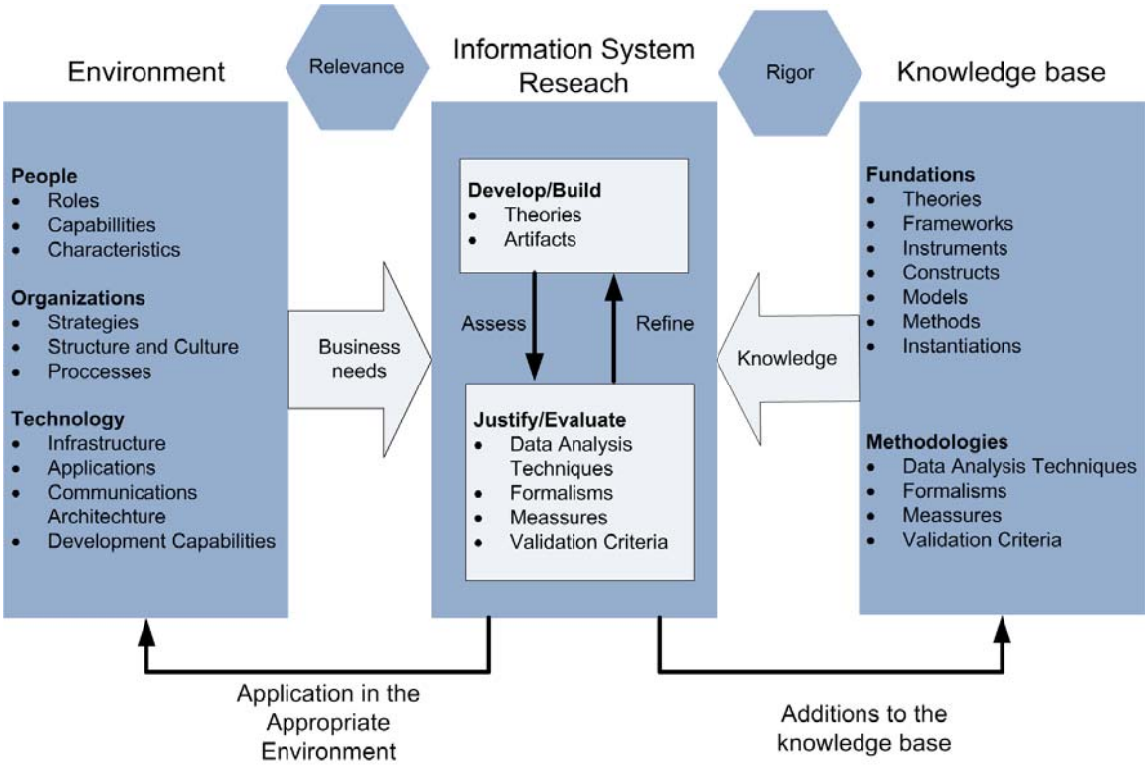

Figure 12 - Characterisation of Design Science in an Information System Research context (adopted from [91]).

Hevner et al.'s description of Design Science comes from the scientific field known as Information Systems Research where there is a strong tradition of explicating the underlying research methodology. The same tradition is not evident in medical informatics research, which makes it difficult to assess whether the Design Science methodology is widely-applied in clinical information systems research. What can be determined is that some medical informatics studies, including one of ours, *have* referenced to the design science approach [92,114,115] . One could argue that the unique complexity of medical informatics e.g. [23] might hamper the usability of general information system methodology or that the methodology clashes with a clinical research understanding of rigor, which includes the randomized controlled trial as a gold standard. For now, clinical study designs can be chosen for evaluation, when it is relevant for the clinical situation in which the artefact is used e.g. Trappenburgs et al's evaluation of telemonitoring of COPD patients, where the use of the monitoring device is regarded as an intervention, and improved outcomes can be shown [116]. It should be noted that clinical evaluation cannot replace classic technology evaluation methods e.g. usability testing or software testing. However, some medical informatics artefacts cannot (and should not) be regarded as clinical interventions, because they do not have a well-defined clinical setting in which the evaluation makes sense e.g. two different clinical model formalisms might express the same clinical content so that when implemented in software there is no difference between the entry screen that the physician see. Any model or terminology could also be implemented in good and bad software respectively, and the software not the underlying model would determine physicians' attitude. As stated by Rogers, it is difficult to distinguish the success or failure of the information system

from the success or failure of the terminology [102]. Evaluation of applicability for underlying models, does not have one well-established methodology in a medical informatics context and therefore evaluation studies in the field is still very context dependent. It remains a challenge to establishing consistent methods for evaluation approaches in medical informatics that takes into account the clashes between design science and clinical research.

C. Appendix: Developed artefacts

To summarize from the enclosed papers, the artefacts designed in the PhD study is:

Template database: Implemented in MS Access with a schema developed in the study described in paper I.

Mapping guidelines: Not implemented in software. Conforming to the guidelines is more a question of training than of software support even though semi automatic mapping would probably be possible. The mapping guidelines are described in paper II.

Inter-rater agreement function: Implemented as a re-write of an existing Krippendorff's α matlab-function. The re-write consisted of introducing a semantic difference function. I was responsible for the matlab function which was one of two implementations of the semantic Krippendorff's α, we did in the study reported in paper III. The rest of the implementation related task was performed by Daniel Karlsson.

Template similarity estimation software: Implemented in Java. Taking as input templates extracted from the template database and outputting a similarity matrix and aggregated similarity estimates for each template-pair. Reported on in paper V. Some user-interface support has also been implemented in the software e.g. visualisation of pair-wise comparison. This has not been reported on in any papers. Template similarity estimation software is dependent on access to a SNOMED CT database including concepts, relationships and description tables as described by IHTSDO technical implementation guide. Furthermore, a transitive closure table and a leaves table was implemented. The database was implemented in a MySQL database in cooperation with Anne Randorff Højen. The transitive closure table was implemented by Anne Randorff Højen using a script originally developed by Rahil Qamar. I implemented the leaves table adding the number of leaves to each concept in SNOMED CT to improve performance on semantic similarity calculations.

Template dendrograms: m-function implemented in matlab taking as input aggregated similarity estimates for a set of templates, mostly putting together build-in clustering techniques from matlabs library. Specialized to ensure correct labelling.