Norwegian Centre for
E-health Research

PROJECT REPORT

Ontology-based terminologies for healthcare

Impact assessment and transitional consequences for implementation

Marco-Ruiz L, Malm-Nicolaisen K, Pedersen R, Makhlysheva A, Bakkevoll PA.

Ontology-based terminologies for healthcare

Impact assessment and transitional consequences for implementation

Project report

Version 1.0

Report number
08-2017

Project manager
Kristian Malm-Nicolaisen

Authors
Luis Marco-Ruiz

Kristian Malm-Nicolaisen

Rune Pedersen

Alexandra Makhlysheva

Per Atle Bakkevoll

ISBN
978-82-8242-077-8

Date
24.10.2017

Number of pages
78

Summary
This report describes the findings and results of an impact assessment of the transition to ontology-based terminologies. The report is focused primarily on infrastructure implications surrounding transitioning to and implementation of ontology-based terminologies and this report describes mostly technical factors.

Our aim is for the report to be used as a reference for future strategic development and decision-making processes.

Publisher
Nasjonalt senter for e-helseforskning
Postboks 35
9038 Tromsø
Email: mail@ehealthresearch.no
Web: www.ehealthresearch.no

Keywords
Ontologies, terminologies, ontology-based, interoperability, standardization, knowledge representation, SNOMED CT, Semantic Web

Table of Contents

List of Figures

List of Tables

Executive summary

This project has investigated the foundations of ontology-based terminologies and the implications on their adoption in Electronic Health Records in particular, and Health Information Systems in general. The project focuses on SNOMED CT as an example of a biomedical ontology, but the descriptions of the challenges and opportunities in transitioning to systems capable to use different types of ontologies follows a generic view.

The project has identified clear benefits in the transition to health information systems that can use different biomedical ontologies. The formal semantics of ontologies allow for automating the knowledge management of complex terminologies, facilitate the maintenance of Clinical Decision Support artefacts, and increase semantic interoperability. Therefore, the adoption of biomedical ontologies has the potential to improve quality of healthcare by increasing the quality and availability of clinical data in different organizations. Another benefit from the adoption of biomedical ontologies is the secondary use of clinical data since ontologies enable expressive phenotyping queries and improve the accuracy of data analysis.

Positive contextual aspects influencing the adoption of biomedical terminologies by Health Information systems are the existence of an established health IT sector in Norway, the access to advise and training from organizations such as SNOMED International, the efforts in the development of top-level semantic architectures to integrate different types of ontologies, the availability of W3C standards to facilitate the gradual adoption of ontologies by legacy systems, and the running initiatives in the context of the European Union that are studying strategies for the adoption of biomedical ontologies such as SNOMED CT and semantic web architectures.

Several challenges exist related to the adoption of ontology-based terminologies. The most relevant challenges are the complexity of their governance; the lack of large scale implementations making use of different types of ontologies in healthcare; and the lack of knowledge on how to approach, from an organizational and technical point of view, the adoption of biomedical ontologies by legacy systems. In addition to these challenges, there are important external threats that may hamper the adoption of biomedical ontologies. These threats include the lack of technological and organizational experts, the lack of experiences in large scale projects to learn how to draw a clear long term path towards the adoption of biomedical ontologies, the vast amount of existing bio-medical ontologies with uncorrelated developments.

The transition to ontologies should be gradual and incorporate formal semantics to Health Information Systems on an "as-need" basis. The first step should be to plan the adoption of a reference biomedical ontology that later allows to incorporate other ontologies. This should be done following a bottom-up approach gradually enriching legacy systems with formal semantics. In addition to terminology guidelines, W3C standards to develop a semantic layer over existing technologies should be considered for scalability and standardization purposes. These standards should facilitate the gradual transition by extending legacy systems with an ontological layer. Since large-scale implementations leveraging several biomedical ontologies are scarce, it is recommended to define a clear strategy for gaining experience and knowledge on the development of semantic architectures that leverage the use of different ontologies. Such strategy should consider running pilot projects that allow for gaining experience and knowledge on how to approach the semantic enrichment of Health Information Systems in use.

1 Introduction

This project's main objective is to identify and describe the consequences of transitioning to ontology-based terminologies. Since the first of January 2017, Norway has been a member of SNOMED International, and the plan is to pilot parts of the SNOMED terminology in the dental healthcare field, following the recommendations in the report "*Felles standardisert terminologi – Vurdering av SNOMED CT*"[1]. Prior to this report, SNOMED CT had already been reviewed in the report referenced above. The report mainly assessed three focus areas of SNOMED CT: i) clinical suitability; ii) implementation and ICT considerations; and iii) organization and management/governance.

The report's final recommendations were to:

- Join International Health Terminology Standards Development Organisation (IHTSDO (now SNOMED International));
- Establish an expert group to work exclusively with SNOMED CT and share standardized terminologies;
- Make efforts to increase the knowledge of and competency in SNOMED CT and standardized terminologies;
- Establish an investigative project to assess SNOMED CT as part of an overall architecture for structured electronic health records (EHRs) and process and decision support;
- Follow and monitor the development of ICD-11;
- Make SNOMED CT accessible to users with immediate needs – especially dental care and openEHR archetype development;
- Commence a new evaluation of SNOMED CT after the 3-year exploration phase and decide on future implementation.

SNOMED CT, unlike other terminologies and clinical codes in use today, was built on an ontology structure with formal logic. This offers great potential, but also poses potential challenges. Next-generation terminologies, such as ICD-11 from the World Health Organization and ICHI (common international terminology for procedures/measures), are also (to a greater extent) ontology-based. Terminologies that will help contribute to realizing areas such as precision medicine and gene-based treatments were also built using ontologies that allow for some form of reasoning through logics.

This project has been conducted in light of other national initiatives and processes, and this project report aims to provide new knowledge on several of the points above. In addition, this report contributes relevant knowledge to the processes described in the document *National Strategy for E-health 2017-2022* ("*Nasjonal e-helsestrategi og mål 2017-2022*"), especially Chapter 2.5 *Critical ICT infrastructure and shared building blocks*:[2]

- Facilitate modern ICT solutions by preparing the introduction of new clinical codes, terminologies and standards.

The Norwegian Directorate for E-health is tasked with planning a shared national EHR in line with the whitepaper *One Citizen – One Journal* (Meld. St. 9 [2012-2013] *Én innbygger – én journal*) (EIEJ), and

---

[1] https://ehelse.no/Documents/Nasjonale%20utvalg/NUFA/Beslutningsgrunnlag%20SNOMED%20CT%20sammendrag%20v1.0.pdf

[2] https://ehelse.no/Documents/Nasjonal%20e-helsestrategi%20og%20handlingsplan/Nasjonal%20e-helsestrategi%20og%20m%C3%A5l%202017-2022.pdf

based on the experiences gained from Helseplattformen. The process is intended to independently assess the needs and possibilities for shared national EHR and related services and functionalities, and should be conducted in close collaboration with relevant actors. In light of this, the National Strategy for E-health emphasizes the importance of standards, clinical codes and terminologies in the processes and initiatives described. At the same time, the potential challenges should not be underestimated:

> *"Today, standards, codecs and terminologies are to a limited extent based on a common coherent framework, and reuse of definitions across terminologies are few. In addition, they are characterized by local customizations, lack of updating, and unsatisfactory management and governance of versions. The use of codes is highly variable, with a lot of erroneous coding and double registration. A more coordinated approach is needed to apply common definitions based on international standards, and disseminate them effectively. There is also a large and growing need for standards, codecs and terminologies in new areas"* (1 pp: 18).

Several measures for the future process are identified and given in the document. This report is particularly suited for contributions to the following measures (1 pp: 18):

- *"Selecting international standards, codes and terminologies for areas that with health-related importance*
- *Govern and implement standards, codes and terminologies more effectively*
- *Establish principles and frameworks for shared information models as part of architectural management"*

The measure further specifies applying ontology-based terminologies, such as SNOMED CT and ICD-11, which should be evaluated for national use with subsequent implementation and application.

1.1 Note to the reader

We suggest that Chapters 3 through 12 be read as interrelated thematic chapters, as they describe contextually correlated aspects of ontology-based terminologies and implementation. Chapters 13 through 15 (including the attachments and literature review) describe related topics useful for further reading. Our aim is for the report to be used as a reference for future strategic development and decision-making processes.

2 Background and approach

The Norwegian health authorities have set a joint sectorial objective for the coming years in the form of *a patient-centred healthcare service.* The health sector is investing substantial resources into ICT to achieve this goal. To ensure that resources are used as effectively as possible, new knowledge to support political and strategic decisions in this process is needed. The Norwegian Centre for E-health Research is cooperating with the Norwegian Directorate for E-health and other actors to find the best e-health solutions for the sector. The aim of this project is to gain knowledge and a common understanding of the consequences of implementation in the transition to ontology-based terminologies in relation to strategic plans for the healthcare sector.

The Norwegian Directorate for E-health is planning a shared national EHR solution for the municipal healthcare sector in line with the recommendations made in the evaluation report in the whitepaper *One Citizen – One Journal*. The whitepaper represents the strategic development direction for the healthcare sector, and the project is conducted in line with the following principles (2 pp: 10):

- *"Healthcare personnel shall have user-friendly and secure access to patient information;*
- *Citizens shall have access to user-friendly and secure digital services;*
- *Data shall be accessible for quality improvement, health monitoring, management and research."*

Multiple processes have been initiated to achieve the goals presented in the whitepaper. Among other things, Norway became a member of SNOMED International on 1 January, and will pilot parts of the terminology in the dental care field.

The project is also presented in the context of the mandate of the Norwegian Ministry of Health and Care Services' Directorate for E-health for 2017, in which the primary aims are described:

> *"ICT standards and clinical terminologies and codes have to be developed, governed and maintained through the Directorate's regulatory role. The management of terminologies and codes needs to be conducted in close collaboration with healthcare professionals and the Directorate for Health. The regulations on ICT standards for the sector need to be further developed and updated in correlation to arising needs in the sector. The goal is that the concept definitions that form the basis of the registration of medical and patient data are included in the national management of terminologies and codes" (3).*

2.1 About the project

Through this project, the Norwegian Centre for E-health Research will identify and assess the potential consequences of the transition and implementation of ontology-based terminologies. The objective of this report is to provide a general overview of the features of ontologies for representing clinical knowledge and to later analyze the benefits, requirements and consequences derived from their adoption.

The project will focus especially on investigating and describing two objectives:

- The project will describe the terminologies and their metamodels.
- The project will identify and assess the possible consequences of different metamodels/concept models in clinical systems in relation to ontology-based terminologies.

Hence, the project has focused primarily on infrastructure implications surrounding transitioning to and implementation of ontology-based terminologies and this report describes mostly technical factors. For organization perspectives, see ASSESS-CT[3][4] and "*Felles standardisert terminology – Vurdering av SNOMED CT*"[5]

New knowledge on what the practical implications of a transition will be, and which important factors require special attention, is needed in order to map the long-term consequences. This report constitutes the final project delivery, and aims to contribute to meeting that need.

The research group has chosen a project approach with three sub-reports that each correspond to different aspects of the overall objective. The sub-reports have delivered chapters and contents for this final report.

The project group at the Norwegian Centre for E-health Research have held regular meetings with the project group at the Norwegian Directorate for E-health where each sub-report has been followed by discussion workshops, in addition to more frequent status meetings. This structure has contributed to keeping the aim of the project on track as new knowledge has been developed. Some changes were made to the subjects of interest throughout the course of the project, but this process was carried out in collaboration with the Norwegian Directorate for E-health. This approach was chosen as it allowed for flexibility as the project progressed.

2.2 About the report

This report is the product from the project, and contains the findings and discussions in relation to the above mentioned project objectives.

Chapter 3 presents a thematic introduction to terminologies, ontologies and formal semantics. Chapter 4 describes the hierarchies and attributes in the SNOMED CT concept model. Chapter 5 discusses how the formal semantics given in SNOMED CT can be exploited through post-coordination methods. Chapter 6 discusses how the level of expression in descriptive logics challenges practical use, while Chapters 7 and 8 describe information models in health information systems and ontologies. Chapter 9 describes different methods of knowledge modelling and representation. Chapter 10 describes semantic web architectures and top-level ontologies in more detail, while Chapter 11 contains a summary of the questionnaire responses. Chapter 12 contains a summary of findings and a SWOT analysis, while Chapter 13 describes the literature review and presents the relevant literature. Chapter 14 lists the references, while the appendix can be found in Chapter 15.

This report is produced in English, partially to reach a broader audience, but also because the project group is composed of international contributors.

---

[3] http://assess-ct.eu/fileadmin/assess_ct/final_brochure/assessct_final_brochure.pdf

[4] http://assess-ct.eu/fileadmin/assess_ct/deliverables/assess_ct_d1_2_report_from_focus_groups_and_questionnaires.pdf

[5] https://ehelse.no/Documents/Helsefaglig%20kodeverk/Felles%20standardisert%20terminologi_Vurdering%20av%20SNOMED%20CT.pdf

2.3 Helseplattformen

The project developed and submitted an abstract to the research conference on Helseplattformen, held in Trondheim in May 2017[6]. The abstract was based on the thematic and subject orientation of this project, as well as other adjoining processes. The abstract was approved for a poster presentation at the conference and the project group was represented there.

The proposed research project focuses on the practical applications of ontology-based terminologies, information models and clinical decision support — separately, but also in relation to each other. The abstract proposes investigating the possibilities of implementing and using ontology-based terminologies in both Helseplattformen and EIEJ.

The abstract is included in Chapter 15 (Appendix).

2.4 Questionnaire and respondents

We have developed and distributed a questionnaire in order to get input from vendors, developers and experts. A central activity of this project has therefor been to identify vendors, researchers, developers and other expertise in order to map the use and prevalence of ontology-based terminologies. The Norwegian Centre for E-health Research has previously conducted a similar project on national terminology servers, and through that process has established some contacts that have also been relevant to this project. The aim of the recruitment process has been to include both vendors and other researchers and developers. We have also chosen to include the vendors in the bid for EHR in Helse Midt, as Helseplattformen is adopted as a regional pilot for the objectives in EIEJ. The experiences from Helse Midt will be normative for other national processes in the healthcare sector.

The aim of including interviews and a questionnaire in the project has been to gain more thorough insight into technical approaches, architectural designs, implementation experiences and prerequisites for transition.

In contact with respondents, the project members have clearly stated the projects aim, scope and mandate, and that participation in no way will give advantages in any procurement processes.

2.4.1 Vendors from Helseplattformen

All four suppliers[7] competing to provide a solution to Helseplattformen were invited to submit a response to the questionnaire. Of these, three indicated that they would participate, while eventually only DXC Technology and Epic Systems Corporation were able to deliver responses.

2.4.2 Other vendors and expert participants

The project has also contacted and recruited international experts on terminology and medical informatics. The project is being conducted in close cooperation with a project on clinical decision support, and some of the same respondents have been used for both projects. Vendors were identified either from literature or industry benchmarking, such as Best of KLAS[8]. Vendors and experts were also contacted based on our familiarity and prior knowledge in the field.

---

[6] https://helse-midt.no/arrangementer/sokekonferanse-for-utvikling-av-forskningsprosjekter-i-helseplattformen-2017-05-11

[7] At the time of writing, four vendors remained in the bid for tender in Helseplattformen.

[8] http://www.healthcareitnews.com/news/epic-tops-2017-best-klas-awards-securing-top-spot-7th-straight-year-see-complete-winners

2.4.3 Questionnaire development

The questions were developed in a team effort with all of the core project members. The backdrop and context for several of the questions is the national Norwegian processes related to the stated strategies in "*One citizen – one journal*". All questions have gone through hearings and revisions before being finalized. Many of the topics are quite technical and specific, and the quality control for these has been especially challenging.

The questionnaire used in the project was developed through an iterative process, and the first version was included in the sub-report for phase 1. Feedback on the questions have been taken into consideration and included for the final version of the questionnaire. The questionnaire is a joint document for the project and report "*Klinisk beslutningsstøtte – vurdering av standard og arkitektur".*

The questionnaire is attached in Chapter 15 (Appendix).

3 Ontologies

3.1 Introduction

3.1.1 The need for interoperability

The design and development of large IT enterprise architectures is always challenging, particularly because of coordinating many components for data processing, transmission, persistence and user interaction. However, some domains introduce even more complicated challenges as a consequence of their inherent complexities. That is the case of life sciences in general, and medicine in particular. Among the many challenges in today's healthcare field (4, 5), the need for health data communication and interpretation across different health information systems is the one that affects national IT infrastructures (6, 7). The large-scale communication of data will only be useful if its interpretation allows different actors in different organizations to draw the same conclusions. Otherwise, it would lead to the wrong inferences. Consequently, data from many sources must conform to a common representation that faithfully specifies certain aspects, such as the context of data (when were the data reported, who reported the data, which parties were involved in the data communication, etc.), the structure of the information (constraints over data entities that ensure that data complies with a particular canonical information schema) and its precise meaning (a conceptual model of the domain of discussion) (8). This requires not only sharing data in a common format, but also preserving the context and meaning of all data entities transmitted across stakeholders. In other words, not only is the interoperability of data needed, but so too is a significant level of semantic interoperability (SIOp) (8, 9). It is commonly believed that three components are needed to enable SIOp (8): reference models, clinical information models and clinical terminologies. The roles of these three components in defining semantically interoperable frameworks are as follows (8):

- Common reference models (RM) define a reduced set of generic entities that are constant across the data models of different health information systems Examples are the openEHR Reference Model (RM) and the HL7 Reference Information Model (RIM).
- Clinical information models (CIMs) (10) constrain the reference model to define the information structure of clinical content in different application domains. For example, in most cases, the EHR content and clinical decision support systems (CDSS) information structures will use different CIMs. Examples of CIMs are those defined by the Norwegian CKM, HL7 CDA-CCD, opencimi.org and, to some extent, HL7 FHIR resource profiles.
- Clinical terminologies are used to annotate CIMs by attaching a standard definition to their elements. Therefore, they endow the CIM with a certain level of semantics. Examples of terminologies are LOINC, for laboratory observations, ICD-10, for disease classification, and SNOMED CT, for general use. 3

In recent years, several national initiatives have defined frameworks for the standardization of clinical information (11, 12). In the US, the meaningful use initiative has propelled the adoption of HL7 CDA for the specification of clinical documents. Particularly relevant is the broad adoption of the HL7 CDA-CCD document that defines a patient summary based on the key areas of information to treat patients safely (13). In Norway, the adoption of openEHR has already produced 47 validated clinical information models, a.k.a. archetypes (14). Other European nations, such as Spain, have also organized the development of a national patient summary by building ISO 13606 archetypes (15). These initiatives have defined valuable assets upon which to structure the clinical content of the EHR. Therefore, they provide a common syntactic structure for clinical content. The definition of common CIMs at a national level provides vendors with a set of standard information structures upon which to base their developments. Since the same set of CIMs is shared across different implementations, interoperability is granted. Nevertheless, in order to share clinical information, not only is a common information schema needed, but the meaning of the sections of that information schema (i.e. semantics) must also be shared. Standards for defining clinical information models allow the binding of their elements to standard terminologies with the objective of attaching semantics to their elements. By attaching a standard code to a section of a CIM, it is possible to identify part of the clinical data in a terminology that has an external standard definition. For example, the element at0021 Alvorlighetskategori (severity category, in English) of the archetype openEHR-EHR-CLUSTER.symptom_sign.v1 published in the Norwegian CKM (arketyper.no/CKM) can be bound to the SNOMED CT concept *162465004 / Symptom severity (finding).*

An example from HL7 FHIR can be seen in the observation of blood pressure (BP),[9] where the SNOMED CT concept *368209003/ Right upper arm structure* is set in the coding section of the *bodySite* element indicating the place of the BP measurement. This way, the element in the information structure that the CIM (HL7 CDA template, archetype, FHIR resource, etc.) defines has a definition in an international terminology, thus attaching semantics to the element in two ways. First, the concept definition, which is also translated into several languages, does so in a direct way. Second, the hierarchy of the concept in the terminology (findings for symptom severity) does so in an indirect way. In the case of ontology-based terminologies, the latter has further implications since it provides a formal definition of the concept semantics that will be explained in the following sections.

Previously, it was mentioned that semantic interoperability needs to share not only information schemas, but also the meaning of these schemas and their elements. This involves sharing the knowledge necessary to interpret data contained in the schema. However, the term 'knowledge' is commonly used in the scientific literature, but is attributed different meanings (16). Therefore, it is important to contextualize the role of ontologies and the types of knowledge they represent.

3.1.2 Types of knowledge

Several classifications for the different types of knowledge are present in the computer science milieu (17-19). In this section, the classification provided by Rector and Sottara is presented as the one closest to the clinical domain (18):

***Ontology:** in medical informatics, ontologies represent knowledge about invariable truths about a particular domain of discussion. E.g. H1N1 is a type of *influenza virus.* This statement is always true, regardless of the situation in which the statement is made.

***Contingent knowledge:** statements that refer to a situation with a particular context, in which the context may influence the validity of the statement. E.g. a cough present for longer than three weeks may indicate underlying tuberculosis. Note that the statement refers to a situation in a particular context in which a patient has been coughing over a period of time. For most patients, it will not be true; but, for some of them it will be and there will be an underlying tuberculosis infection present.

---

[9] https://www.hl7.org/fhir/observation-example-bloodpressure.json.html

***Schemas for data structure:** this type of knowledge refers to the structuring and expression of information in health information systems. E.g. the FIHR profile and the archetype to record blood pressure provides knowledge on how to specify a complete record of blood pressure measurements. It sets constraints on data in terms of what is mandatory to record, what is optional and the structure with which it needs to comply in order to be a robust, interoperable data schema.

***Procedural knowledge:** dynamic knowledge that combines ontologies and contingent knowledge to infer a new knowledge entity. E.g. IF *elevated blood pressure is present*; THEN *consider treatment with angiotensin-converting enzyme inhibitors*. This type of knowledge is usually contained in the set of rules that expresses a computerized clinical decision support system (CDSS).

Among these different types of knowledge, ontologies are the ones central to this report. The knowledge expressed by SNOMED CT belongs to the ontology category. SNOMED CT (with its concept model) is an ontology that represents a model of the concepts available in the clinical domain providing a formal representation to facilitate the management and the definition of new terms (post-coordination) not available in the fixed set of terms defined in the release. The term ontology has been extensively used in the research literature and it is important to notice that there may be differences when it is used in the philosophical sense than the computer science sense. In this this report when the term ontology is used independently it refers to the computer science sense, which as explained in the next paragraph represents a formal representation of a model independently of the type of knowledge it represents. When “ontology” is used in the philosophical sense (accepted in medical informatics) we will specify the meaning of the term using “domain ontology”.

In the computer science arena, the first description of ontology was provided by Gruber (20): “An ontology is an explicit specification of a conceptualization”. Later, other authors refined the definition (21). Nowadays, Studer’s definition of ontology as a “**formal**, **explicit** specification of a **shared** conceptualization”(21) is widely accepted.

Several of the terms used in Studer’s definition are key to understanding the role of ontologies in computer science and artificial intelligence. **Formal** means that the specification of the domain being modelled is expressed in a mathematical fashion, thus making it machine-understandable. **Explicit** means that the specification of the concepts, relations and constraints of the domain modelled is made in a complete and direct manner without the need to think about implicit implications. **Shared** implies that the ontology must be agreed upon and valid not only to one actor, but to all those involved in the communication based on the ontology. Therefore, an ontology should not be developed from one point of view alone; instead, it needs to be as generic and useful to as many people as possible. The formal and explicit specification avoids ambiguity and describes the domain of discussion in mathematical terms. This leads to models that are machine-understandable. The software capable of processing and interpreting ontologies is known as a “reasoner”. Reasoners apply a set of mathematical theorems to be able to make inferences over ontologies.

3.1.3 Types of ontologies

The previous section presented the different types of knowledge identified by Rector and Sottara (18) and explained the role of ontologies as knowledge specification of a certain domain that express invariable truths. The different types of ontologies are another source of confusion when one studies the literature on knowledge specification. In fact, within the category of knowledge related to ontologies, the different subtypes previously described are identified based on the authors’ perspectives. A commonly accepted classification of the different types of ontologies is that of Guariano (16). The work of

Guariano on the methodological and architectural implications of formal ontologies for information systems has led to four main types of ontologies (16):

- **Top-level ontologies (a.k.a. upper-level ontologies):** these kinds of ontologies (e.g. Knowledge representation (KR) ontology (22) Basic Formal Ontology (23)) describe general concepts common to most domains. Examples of the concepts contained in top-level ontologies are object, purpose, reason, time, space, etc. See Chapter 10.2 for more.
- **Domain ontologies:** these ontologies represent conceptual models of a particular domain. For example, SNOMED CT is a domain ontology for medicine.
- **Task ontologies:** describe tasks or activities. Examples are the ontologies to specify clinical tasks, such as PROforma define action, enquiry, decision and plans (24).
- **Application ontologies:** these ontologies subsume the two previous categories (domain and task ontologies) and specify the role that entities from the domain ontology acquire when they execute a particular task. For example, an application ontology of clinical tasks may contain terms such as perform *vulvar biopsy*, *Chronic disease management of atrial fibrillation* (25).

Ontology modelling processes always reflect some degree of subjectivity. Therefore, for some ontologies, a clear-cut way to assign them to one category or another cannot be established. That is the case of BioTop (26, 27), which is defined by its implementers as a top-level ontology. However, it is not a top-level ontology in the strictest sense as the KR ontology is, for example (22). BioTop is an upper-level ontology with respect to life sciences, but it is not cross-sectional to any domain of knowledge. Therefore, it can be seen as an intermediate layer between biomedical domain ontologies and top-level ontologies.

According to the examples provided above, ontologies such as SNOMED CT, Gene Ontology (GO) and Human Phenotype Ontology (HPO) are domain ontologies. This means that they provide a formal conceptualization of their domain of discussion, i.e. the medical domain, as in the case of SNOMED CT.

3.2 The role of formal semantics

The previous sections have introduced the type of knowledge that ontologies represent and a classification of the different types of ontologies. The definition of 'ontology' previously explained made reference to its formal nature. This section of the report explains the need for formality in ontological knowledge specification.

The medical domain contains a vast amount of concepts interrelated by complex relationships. For example, according to SNOMED CT, for the concept *Malignant neoplastic disease* alone, there are 56 direct subtypes, which, in turn, have additional subtypes (e.g. *Adenoid cystic carcinoma of oropharynx*). In fact, that concept has a total of 3,783 children when the *is-a* relationships across its children are analysed transitively. To further compound the complexity, the relationships between the concepts are very diverse. The most common relationship is the subsumptive one that allows for specifying that one concept is a subtype of another. For example, Warfarin is a subtype of anticoagulant. Other types of relationships aim to specify the location of a condition or a measurement (e.g. the tracheobronchial lymph node is located near the carina). Moreover, some relationships relate to causality and even refer to a level of certainty about a statement (e.g. allergy dermatitis is probably caused by contact with latex). This leads to three main scenarios that need formal semantics:

Knowledge management

Managing the vast number of concepts and relationships in medicine is a challenging task. In fact, if knowledge management is attempted without using technologies that automate the processing of concepts and relationship classification, it can become unmanageable in some cases. Most terminologies are classifications of codes that specifies different topics. That is the case of ICPC-2, for example, which provides categories of codes for recording the reasons for encounters and problems (28). Another example is LOINC for laboratory observations (29). These terminologies provide some rules for using their coding system. LOINC structures each observation into six axes such as component, kind of property, time, etc. However, these structures consist of basic rules that humans must apply in order to use the coding system, but they are not formally expressed. By contrast, terminologies, such as SNOMED CT or ICD-11, provide a formal specification of their meaning by means of languages based on Description Logics (DL). Most terminologies have a relatively limited set of codes and syntactic naming conventions. Additionally, most of them are designed to support the coding of information in one particular application domain (NIC/NANDA for nursing, ICPC-2 for primary care and LOINC for laboratory observations). However, if we consider general-purpose terminologies such as SNOMED CT, a fixed set of concepts does not suffice, nor is it maintainable to cover all the concepts that are present in the clinical domain. As Lassila and McGuinness point out, syntactic technologies and standardization can only provide partial solutions to problems involving complex knowledge management (30). In the case of SNOMED CT, with more than 300,000 concepts and one million relationships, the manual maintenance of the terminology is not possible. Moreover, not even all of its terms would be enough to form a common ontology that could cover any concept in the clinical domain. In many cases, they need to be combined by using some grammatical rules set by the underlying ontology (i.e. the SNOMED CT concept model), thereby forming a new concept. This process is known as post-coordination (a.k.a. composition).

Semantic interoperability

Another implication of the complexity of the clinical domain affects the use and analysis of clinical data. Semantic interoperability and secondary use of clinical data require the interpretation of relationships among clinical concepts. For example, when sharing information across two organizations using the same CIM, it is necessary to interpret equivalences between concepts in local data representation schemas and the CIM used to interoperate. Dixon et al. documented a case in which two organizations, relying on the same HL7 CDA document and mapping to a common terminology, found major barriers to establishing semantic interoperability (31). The reason was that, even when the same terminology was used, different codes were chosen to represent the same concept. This is something very likely to happen in the case of SNOMED CT, as well. For example, if SNOMED CT is used as a mere set of codes, without relying on its ontology, we may find that organization A assigns the code *75367002/Blood pressure (observable entity)* to identify a blood pressure recording, while organization B chooses the concept *163020007/On examination - blood pressure reading (finding)*. Note that both concepts belong to different hierarchies in SNOMED CT and semantic interoperability will be hampered. Without specific mappings, the information system of organization A will not be able to automatically interpret the information sent by organization B. In this case, the only way to detect problems and truly infer whether two organizations are going to be able to semantically interoperate is not only to assign some codes to the CIM, but to define the relationships among the codes used in a common ontology. Only then can both the information model and the terminology concepts be shared and interpreted by machines without ambiguity. Then, if the code used by one organization is equivalent to the other one, or one is a supertype of the other, the reasoner will be able to analyse and interpret these relationships by relying on their formal specification. For example, if a CDS system for recommending the dosage of anticoagulant drugs specifies its valid inputs as any type of anticoagulant, it is possible to automatically infer that Rivaroxaban is a valid input into the system because it is a subtype of anticoagulant (6). Furthermore, any trade name of anticoagulant may be used to identify the drug and the system can validate it without further developments relying on the drugs' ontologies.

Secondary use of data/expressive queries

Besides their use in establishing SIOp, the need to formally represent concepts and their relationships is even more pressing in the field of secondary use of data for clinical research and public health (e.g. phenotype queries). For example, in a national research infrastructure one may want to investigate tendencies in cardiovascular disorders. Therefore, the national health information infrastructure may need to answer a query similar to "count and group by type all cases of cardiovascular disorders". In order to execute such a query, the information system needs to analyse subsumptive relationships to locate any type of cardiovascular disorder. The software executing the query would need to analyse all the data of the organization executing the query and determine which individuals are annotated with any of the possible 5,789 concepts that are sub-concepts of "cardiovascular disease" (*49601007/Disorder of cardiovascular system (disorder)*). For that, the class–subclass relationships among those concepts would need to be modelled in a database. This effort seems unachievable, even at the national level. Syntactic technologies, such as Relational Database Base Management Systems (RDBMS), are not able to manage that level of expressivity in a scalable way, neither through the use of an independent thesaurus nor through informal hierarchies like those formed by terminologies such as ICPC-2. In that case, a certain level of formality in the relationships among medical concepts is needed in order to allow machines to respond to expressive queries.

3.3 Formality levels

The previous section presented the need for formal semantics. Nevertheless, different levels of formality are available. Lassila and McGuinness established a scale depending on the degree of formality of ontologies (30). Figure 1 shows their scale and adds (in red) the language with the expressivity power to represent the different levels of formality. From left to right, the red arrow represents the increase in the expressivity power of the formalism. The cross-cutting line marks the border to what, according to Lassila and McGuinness, can be considered a formal ontology in computer science.

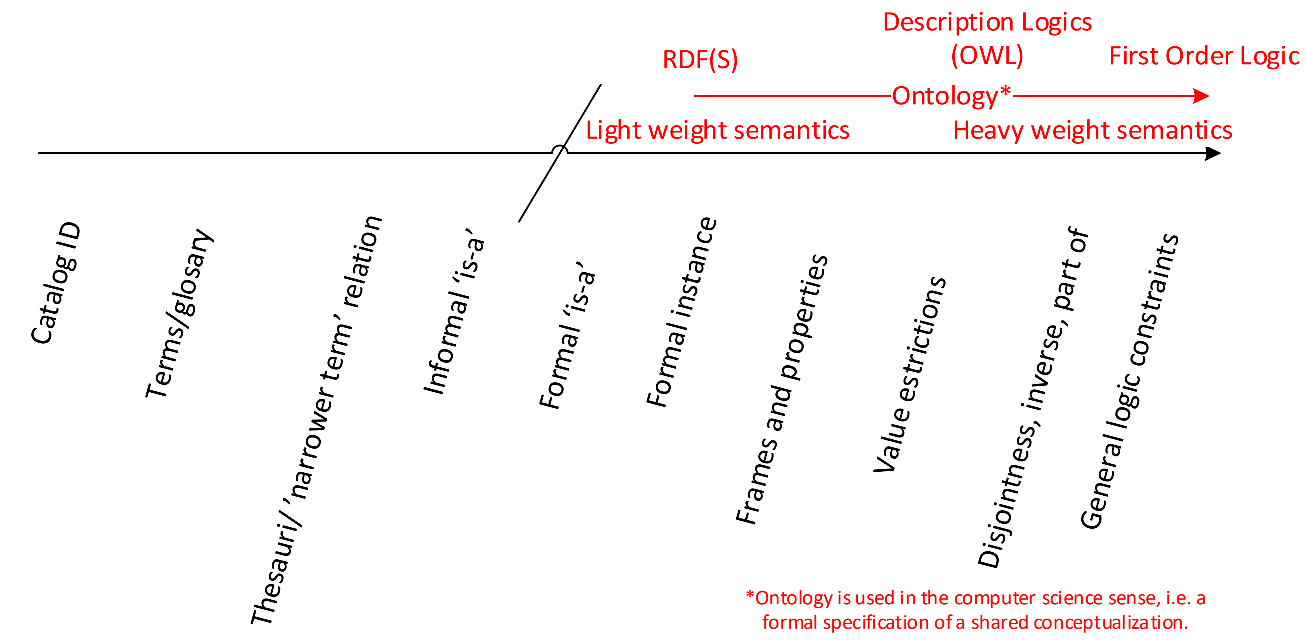


Figure 1: The ontology spectrum adapted from Lassila and McGuinness (29)

According to their scale of formality, the following categories can be established:

- **Catalogue/controlled vocabulary/terms/glossary:** controlled vocabularies are lists of terms used in a particular domain that may assign a unique identifier to each term. They may include a definition attached to each term. Example: HL7 controlled vocabulary.
- **Thesauri/"narrower term relation":** introduce new relationships among terms, such as synonym, antonym, narrower/broader term, etc.
- **Hierarchy:** hierarchies introduce specialization relationships among terms that can be specified with different levels of formality.
    - Informal *is-a* relationship: hierarchies of classifications that are not specified in any formal language. For example, Berlin and Sim developed an informal taxonomy to classify types of Clinical Decision Support Systems (32).
    - Formal *is-a* relationship: includes a formal definition of subclass property. In SNOMED CT, *Lynch syndrome* is a subclass of *primary neoplasm of colon*, which in turn is a subclass of *primary neoplasm of large intestine*. The formal specification of the subclass relationship allows computers to infer that *Lynch syndrome* is also a subclass of *primary neoplasm of large intestine*. The light-weight semantics of RDF(S) allow for this level of expression.
    - Formal "instance": allow defining individuals of classes by appropriately classifying them in the hierarchy of the taxonomy. Again, the light-weight semantics of RDF(S) allow for this level of expression.
- **Frames and properties:** as presented by Brachman (33), frames and properties allow concepts to be defined as something other than nodes and links. Frames and properties allow for endowing concepts with a structure that may be composed of different attributes. For example, a patient class may contain attributes such as social security number, age and ethnicity. In addition, frames and properties allow for defining structural properties of relationships allowing

their specialization. These aspects allow for implementing features such as attributes and property inheritance.
- Value restrictions: allow restricting properties with specific cardinalities and classes. For example, let us present a located at property which aims to specify the anatomical location of a given clinical observation. Value restrictions allow specifying that the destination of that relationship must be a subtype of some anatomical location (e.g. located at left lung) and how many instances may be involved in that relationship (cardinalities) (e.g. a person has exactly one biological father and one biological mother). More expressive mechanisms than those to provided by the light weight semantics of RDF(S) are needed to specify some of these constraints (Description Logics).
- **Disjointness, inverse, part-of:** these kinds of properties are contained in Description Logics (a subset of First-order Logic). They allow, for example, establishing that an individual cannot be member of class A and B at the same time. Description logics and a language to specify them in a computable manner (OWL) allow the definition of this kid of properties.
- **General logical constraints:** contains constraints from First-order Logic. However, ontology reasoners are not able to process models built with this kind of logic in a decidable manner. That was one of the main reasons for developing Description Logics, which are, in fact, tractable[10] fragments of First-order Logic.

According to Lassila and McGuinness, for a model to be considered a formal ontology (in the computer science sense), it must be finite-controlled and have an extensible vocabulary; the model must allow for the unambiguous interpretation of classes and relationships; and it must define strict hierarchical subclass relationships between classes (30).

In the spectrum of complexity shown in Figure 1 the SNOMED CT concept would be situated on the right side of Lassila and McGuinness' schema from the hierarchies that specify formal ´is-a´ relationships and rightwards. The following sections of this report provide a brief note about the computational properties of the language used to express the SNOMED CT concept model.

[10] Processable in polynomial time.

4 SNOMED CT concept model

The SNOMED CT concept model is an ontology that provides a formal representation of the rules to organize the terminology concepts relying on Description Logics. It specifies the directives that determine (34): a) how the hierarchies of the terminology are organized from more general terms to more specific; and b) how relationships between concepts can be specified in a valid manner. These relationships are the ones that will allow for building post-coordinated expressions, as will be explained later.

4.1 Hierarchies

Figure 2 shows the top categories of the terminology encompassed under the general term of SNOMED CT concept. The latter is the parent concept of all the concepts in the terminology. Underneath the SNOMED CT concept, each category contains a poly-hierarchy of concepts (i.e. a concept may have several parents). These concepts become more specialized as one goes down the hierarchy. For example, in the *Clinical finding* hierarchy, the direct child of *Calculus finding (finding)* is a more general concept than the concept *Calculus of cystic duct with acute cholecystitis (disorder)* which is a leaf concept in the hierarchy. The hierarchies follow an is_a specialization in depth. In addition, the relationships in the ontology are transitive. Therefore, if concept B is child of concept A, and concept C is a child of concept B, then concept C is also a child of (is-a) concept A. This means that if a reasoner analyses which concepts are the children of, for example, *Calculus finding (finding)*, not only will the direct children in the hierarchy be considered, but also every concept that is a specialization in the hierarchy, down to the leaves. Therefore, *Calculus of cystic duct with acute cholecystitis (disorder)* will be considered a child of *Calculus finding (finding)*. In total, there are 19 top-level hierarchies. Table 1 contains the textual definitions provided by the IHTSDO guidelines for each hierarchy (34) and the type of hierarchy as defined by Benson (35). Those interested in the complete information about each hierarchy should access the concept model guidelines (34).

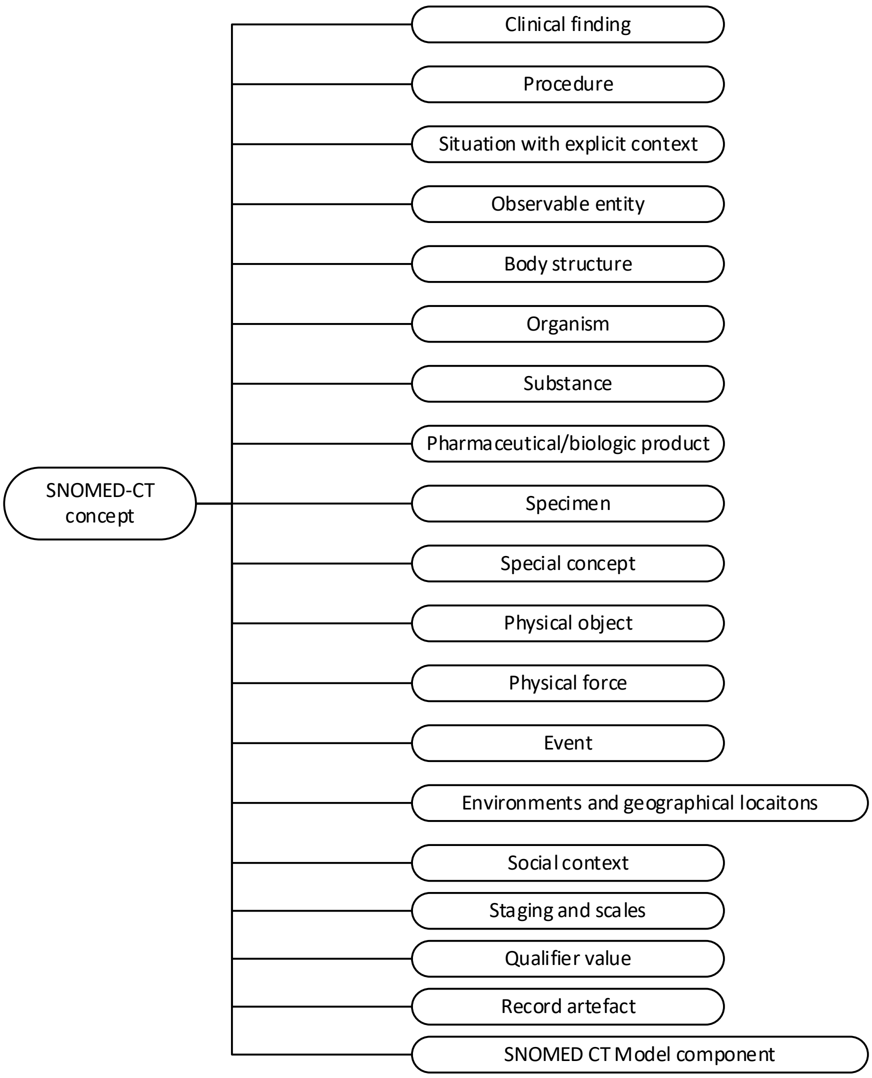


Figure 2: Top hierarchies of SNOMED CT.

Type of Hierarchy	Hierarchy	Description
Object hierarchies (concepts that can be qualified)	**Clinical finding**	“Represents the result of a clinical observation, assessment or judgment and includes normal and abnormal clinical states (e.g. |asthma|, |headache|, |normal breath sounds|). The |clinical finding| hierarchy includes concepts used to represent diagnoses”. (34)
	Procedure	“Represents activities performed in the provision of health care. This includes not only invasive procedures but

		also administration of medicines, imaging, education, therapies and administrative procedures (e.g. |appendectomy|, |physiotherapy|, |subcutaneous injection|)". (34)
	Situation with explicit context	"Represents concepts in which the clinical context is specified as part of the definition of the concept itself. These include presence or absence of a condition, whether a clinical finding is current, in the past or relates to someone other than the subject of the record (e.g. |endoscopy arranged|, |past history of myocardial infarction|, |family history of glaucoma|)". (34)
	Observable entity	"Represents a question or assessment which can produce an answer or result (e.g. | systolic blood pressure|, |color of iris|, |gender|)." (34)
	Event	"Represents occurrences excluding procedures and interventions (e.g. |flood|, |earthquake|)". (34)
	Staging and scales	"Represents assessment scales and tumor staging systems (e.g. |Glasgow Coma Scale|, |FIGO staging system of gynecological malignancy|)". (34)
	Specimen	"Represents entities that are obtained (usually from the patient) for examination or analysis (e.g. |urine specimen|, |prostate needle biopsy specimen|)". (34)
Value hierarchies (concepts used to define values in relationships)	**Body structure**	"Represents normal and abnormal anatomical structures (e.g. |mitral valve structure|, |adenosarcoma|)". (34)
	Organism	"Represents organisms of significance in human and animal medicine (e.g. |streptococcus pyogenes|, |beagle|, |texon cattle breed|)". (34)

	Substance	"Represents general substances, the chemical constituents of pharmaceutical/biological products, body substances, dietary substances and diagnostic substances (e.g. |methane|, |insulin|, |albumin|)". (34)
	Pharmaceutical/biologic product	"Represents drug products (e.g. |amoxicillin 250mg capsule|, |paracetamol + codeine tablet|)". (34)
	Physical object	"Represents natural and man-made physical objects (e.g. |vena cava filter|, |implant device|, |automobile|)". (34)
	Physical force	"Represents physical forces that can play a role as mechanisms of injury (e.g. |friction|, |radiation|, |alternating current|)". (34)
	Social context	"Represents social conditions and circumstances significant to health care (e.g. |occupation|, |spiritual or religious belief|)". (34)
	Environments and geographical locations	"Represents types of environments as well as named locations such as countries, states and regions (e.g. |intensive care unit|, |academic medical center|, |Denmark|)". (34)
Miscellaneous hierarchies	**Qualifier value**	"Represents the values for some SNOMED CT attributes, where those values are not subtypes of other top level concepts. (e.g. |left|, |abnormal result|, |severe|)". (34)
	Record artefact	"Represents content created for the purpose of providing other people with information about record events or states of affairs. (e.g. |patient held record|, |record entry|, |family history section|)". (34)
	Special concept	"Represents concepts that do not play a part in the formal logic of the concept model of the terminology, but which

		may be useful for specific use cases (e.g. |navigational concept|, |alternative medicine poisoning|)". (34)
	SNOMED CT Model Component	"Contains technical metadata supporting the SNOMED CT release". (34)

Table 1: SNOMED CT top hierarchy types and their descriptions from the SNOMED CT concept model (34)

4.2 Attributes

Attributes allow for refining the meaning of concepts by linking them to a value or to other concepts. As mentioned before, the concept model determines the rules to define relationships among concepts. Therefore, attributes cannot be freely applied to any concept, but they need to be assigned a valid domain (origin concept of the relationship) and range (target concept of the relationship). The domain of the attribute determines which concepts can use the attribute (origin of the relationship). For example, the expression below shows how the *Adverse reaction* concept uses the *Severity* attribute to bind the value *Severe*. Since the *Severity* attribute domain is *Clinical finding* (see concept model (34)), and *Adverse reaction* is a subtype of *Clinical finding,* the attribute can be bound to that concept. At the same time, according to the concept model, the range of valid values for the target of the relationship marked is any sub-concept of *Severities / 272141005* (e.g. *24484000 / Severe (severity modifier) (qualifier value)*), thus *24484000 / Severe (severity modifier) (qualifier value)* is a valid value. The former makes the expression comply with the concept model, thus a reasoner can classify it.

```
(281647001|Adverse reaction|: 246112005|Severity|=24484000|severe|)
```

5 Exploiting the SNOMED CT ontology: Post-coordination

5.1 Exploiting formal semantics

The SNOMED CT concept model (i.e. the SNOMED CT ontology) presented determines the valid organization and possible combinations of the concepts and attributes to build new concepts (i.e. post-coordinated concepts) that refine the meaning of the original ones (i.e. pre-coordinated concepts). This process is known as composition in the ontologies development milieu and post-coordination in the terminologies jargon (18). Post-coordination provides a mechanism that allows for combining concepts and attributes to define more specific concepts that could not otherwise be specified using pre-coordinated terms. Although post-coordination is often seen as merely a way of building sentences about clinical statements, it is important to remember that there are further implications of the use of the ontology. When building post-coordinated expressions following the combination rules marked by the ontology, the user is in reality building logical axioms. These axioms conform to Description Logics and therefore provide a formal definition of the clinical semantics specified by such statements. As the first section of this report explained, these formal definitions allow machines to understand these expressions. This means that a reasoner capable of interpreting the ontology will be able to analyse the expressions and classify them correctly in the corresponding SNOMED CT hierarchy.

In order to understand how a reasoner can correctly classify new terms defined as post-coordinated expressions, let us suppose that there is a subset of the SNOMED CT hierarchy (displayed in Figure 3) where the concept *Disease* has already been defined as a subclass of *Clinical_finding*. In addition, the hierarchy of the top-level *Organism* has been defined up to the concept *Virus*. In that situation, we could define the property *causative_agent* to link an organism to a disease to specify the agent that causes it. Using abbreviated statements from the Ontology Web Language (OWL) to specify the axioms, this would look like:

Property *causative_agent.*

We could be interested in creating a concept to define a sub-category for infectious diseases. In OWL, this would look like:

(1) *Infectious_Disease SubClassOf Disease* (see blue SubClassOf relationship 1).

Then we could define more specific concepts, such as *Viral_Disease*, as those that are caused by some type of virus:

(2) *Viral_Disease EquivalentTo (Infectious_Disease and (causativeAgent some virus))* (see blue SubClassOf relationship 2 and green causative_agent relationship 2).

We could also define that *Influenza_Virus* is a subtype of *Virus*:

(3) *Influenza_Virus SubClassOf Virus* (see blue SubClassOf relationship 3).

Then we could rely on the previous axioms to define that *Influenza* is an *Infectious_Disease* for which the causative agent is an *InfluenzaVirus*:

(4) *Influenza EquivalentTo (InfectiousDisease and causativeAgent some InfluenzaVirus)* (see blue SubClassOf relationship 4 and green causative_agent relationship 4).

The reasoner would infer that influenza is a subtype of viral disease without stating it explicitly, provided that it is a viral disease with a causative agent that is a virus:

(5) *Inferred=> Influenza SubClassOf ViralDisease* (dotted red SubClassOf relationship 5).

In this case, the formal semantics of the specified axioms have allowed the reasoner to automatically infer a relationship that was not explicitly stated by the ontology developer. Although this is a basic example, when the number of concepts is as large as it is in SNOMED CT (more than 300,000), and with so many relationships (more than 1,000,000), automatic classification is necessary to maintain the terminology and ensure that all the relationships for a concept have been generated.

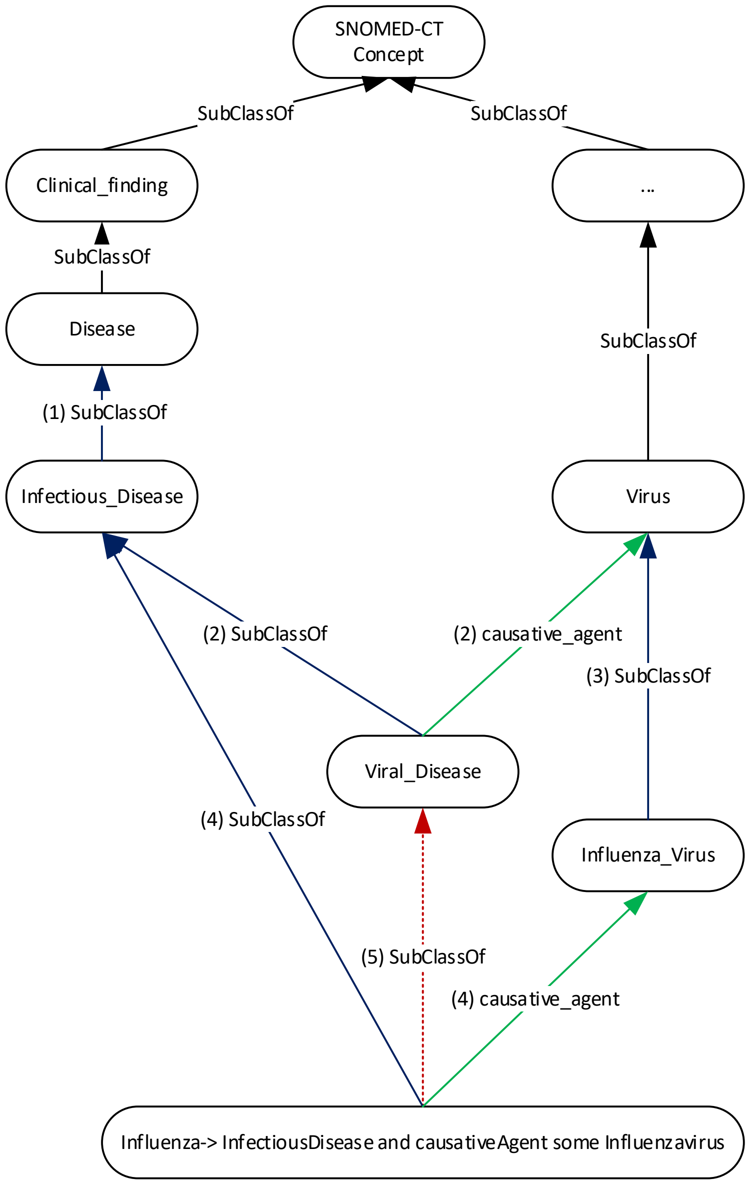


Figure 3: Example of classification by a reasoner.

This example was presented using an abbreviated form of OWL. However, this is not a human-friendly language and it may be cumbersome to users of the terminology who do not have training in DL. To overcome the problem of complexity, IHTSDO identified that the SNOMED CT language, including the one used to build logical expressions, should have the following properties (36):

- Backward compatibility with previous versions of the language;
- Consistency in the meaning of each logical feature across all SNOMED CT languages;
- The language should be sufficiently expressive for defining the clinical concepts that the terminology was designed to support;
- Machine processability: it must be possible to parse SNOMED CT expressions into a language that allows for its processing (e.g. OWL);
- Human readability: the language must be readable by non-technical, human stakeholders.

5.2 SNOMED CT compositional grammar

The language developed for defining post-coordinations and compiling them with the properties previously presented is known as the compositional grammar. The compositional grammar allows for defining logic axioms in the form of human-friendly clinical statements. Let us present some examples of how the compositional grammar allows for exploiting the SNOMED CT concept model to refine the meaning of clinical concepts. The following example presents the specification of a clinical problem, the cause of which is an immune hypersensitivity reaction. The example exploits the *Due to* attribute to specify the cause of the problem.

```
55607006|Problem|:42752001|Due to|=609407009|Immune hypersensitivity reaction by mechanism|
```

The following example exploits two attributes in one expression for defining a blood pressure measurement using a *Blood pressure cuff*.

```
122869004|Measurement|: {
363702006|Has focus|=75367002|Blood pressure|,
                363699004|Direct device|=70665002|Blood pressure cuff|}
```

It is interesting that in the previous example, the post-coordinated expression has a pre-coordinated equivalent:

```
371911009|Measurement of blood pressure using cuff method (procedure)
```

These situations may occur when no proper control of the terms to be used is performed. In addition to the technical details, it is important to consider the appropriate management of the terminology. SNOMED CT is not intended to be used directly. Adequate management and selection of the set of concepts that will be used in an application domain must be performed. More information on the organizational needs for the adoption of SNOMED CT can be found in (36, 37).

Post-coordinated expressions can become very complex, allowing complex refinements to be made to the semantics of a particular concept. For example, the following expression, taken from the compositional grammar guideline (36), specifies a concept that is the salpingo-oophorectomy with laser of the right ovary procedure and an excision of the left fallopian tube. In the latter, the device employed for the procedure is not specified. Grouping using curly braces helps clarify to which procedure the device refers.

(Example taken from (36))

```
71388002 |procedure|:
          { 260686004 |method| = 129304002 |excision - action|,
          405813007 |procedure site - direct| = 20837000 |structure of right ovary|,
          424226004 |using device| = 122456005 |laser device|}
          {260686004 |method| = 261519002 |diathermy excision - action|,
```

405813007 |procedure site - direct| = 113293009 |structure of left fallopian tube|}

The following is another example of how to record the history of a risk factor. The concept specifies a history of an adverse reaction to penicillin.

```
161632004 |History of risk factor|:
        { 246090004 |Associated finding| =
        ( 281647001 |Adverse reaction|:
        { 47429007 |Associated with| = 439772008 |Date of last episode of disorder|,
        246112005 |Severity|= 24484000 |severe|,
        42752001|Due to|=6369005|Penicillin|}))}
```

The compositional grammar avoids the use of logic operators, such as conjunctions or existential ones, that would require having some knowledge of DL. More examples are available in the compositional grammar guideline (34). In addition, a parser to validate post-coordinated expressions is freely available at: http://apg.ihtsdotools.org/SCT-cg.html.

6 The cost of expressivity

In previous sections, the ability of ontologies to be processed by reasoners relying on their mathematical underpinnings was presented. It was also mentioned that ontologies in the computer science arena are usually defined by relying on Description Logics. Description Logics are a family of logics in which axioms can be mapped to some First Order Logic (FOL) predicates, thus they allow for specializing reasoning techniques (improving their computability in comparison to FOL theorem-provers) (38). When relying on logics for the specification of semantics, it is important to consider that the higher the expressivity, the higher the computational cost for computing the logic axioms (39). This is of significant relevance to large enterprise systems since a very expressive logic may lead to models that cannot be fully processed by any computer. This is an issue that pertains to the computability and computational complexity theories. In a nutshell, computer scientists consider a “good” solution to a problem to exist if they can guarantee that: a) it is possible to build an algorithm capable of finishing the computation to establish whether the set of formulas that specify the problem are valid (i.e. the problem is decidable); and b) its computation for certain input will be carried out in polynomial time (i.e. the algorithm is tractable). In the semantic web domain, this means that: a) the reasoner will return a result stating whether a set of axioms is valid at some point; and b) the time to process the ontology, at worst, can be specified by a polynomial function. Figure 4 is a chart showing how the execution times of two sample algorithms (A in orange and B in green) grow in relation to the size of the input data. On the one hand, the figure depicts how, for algorithm A, the polynomial function corresponding to its execution time (in orange) ($DATA\_SIZE^{10}$) is kept at reasonable levels with the increase of input data. Therefore, the problem solved by algorithm A is considered tractable. On the other hand, the exponential function in red (10^{DATA_SIZE}) that corresponds to the computation time spent by algorithm B experiences a boost in the cost of execution times with the increase of the input data size (in green). The problem solved by algorithm B would be considered intractable.

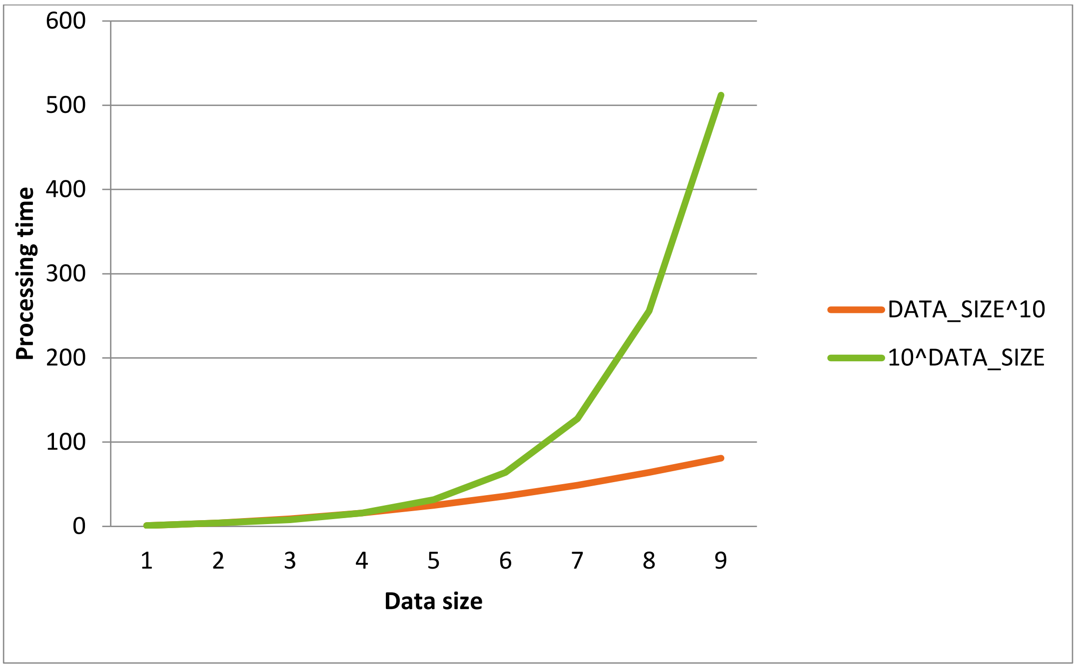


Figure 4: Processing time of polynomial (tractable) vs. exponential (intractable) execution times.

To maintain the tractability of the computation of ontologies, OWL2 DL is currently divided into several profiles with different computational characteristics. These different OWL2 profiles restrict some of the axioms that can be used to define ontologies in order to guarantee tractability. The profile used in the implementation of SNOMED CT is the OWL2 EL profile, based on EL++ logics (40, 41). This profile has some features that make it appropriate for the definition and processing of ontologies that contain a large number of classes. These ontologies are common in life sciences. Figure 5 presents a representation of the combined computational complexity of the logic underlying OWL2 DL (in green) and the profile OWL2 EL++ (in orange). It is possible to appreciate how the exponential complexity of the OWL2 DL (N2ExpTime-complete, to be precise) (42) boosts the processing time in comparison to the polynomial complexity of OWL EL++. The processing time and data size units in Figure 5 are only illustrative. Those interested in the particular logic features of OWL2 DL (i.e. SROIQ) and EL++ may consider reading Hitzler et al. (43).

Although EL++ allows for complex ontologies, such as SNOMED CT, to be defined, it still restricts the use of many axioms that would be beneficial to representing clinical statements (44, 45). Some of these axioms are negation, cardinality restriction and universal quantification, among others (44). This has very important implications since, for example, the absence of negation does not allow to reason about statements such as *absence of fever* or *no internal bleeding*. The implications of such limitations will be treated in the following section in connection with the use of clinical information models.

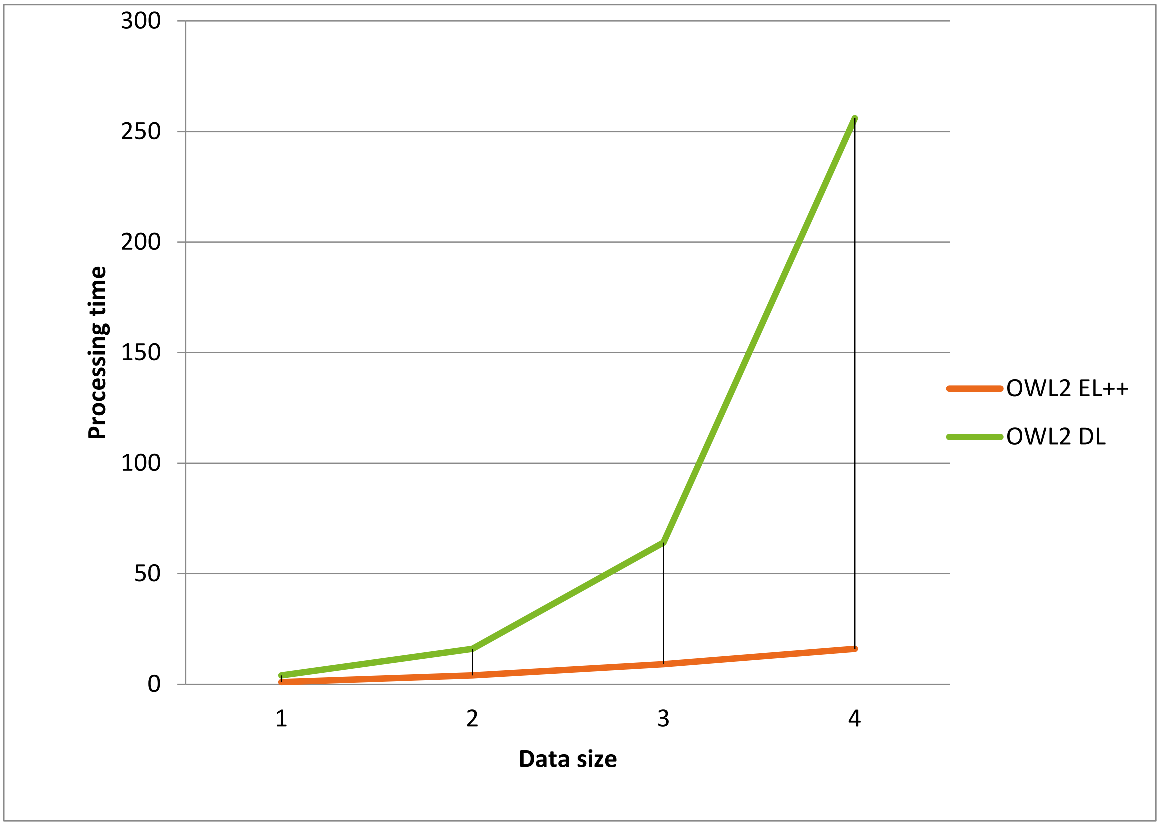


Figure 5: Computational complexity illustration of OWL2 DL and OWL2 EL++.

7 Representing ontology and epistemology in health information systems

The need to build different models in information systems to represent epistemology and ontology dates back to the 1970s, when knowledge representation languages (predecessors of languages such as OWL) were being developed by artificial intelligence researchers (46). At that time Brachman (33) identified the need of developing a semantic model to link the implementation logic (pure logic atoms and predicates) with the conceptual model (concerned with domain ontology aspects). This model was called the epistemology level (33). Brachman presented that level in terms of the need for providing constructs to specify aspects such as inheritance and the structure of concepts. From a health informatics point of view, this is actually the space covered by CIMs since they provide the structure and context of clinical information. In the case of Bachman's levels, at a technical level, the epistemological level allows for specifying the internal structures and features of concepts and relationships (independently of pure ontological knowledge that does not account for structure). In sum, the epistemological level defined by Brachman allows for describing the conceptual sub-pieces of an entity, defining how sub-pieces should be used and the relationships between them (e.g. inheritance, properties of relationships between entities, etc.).

In the 1980s, Brachman's technical view was generalized, adding a more philosophical dimension to the definition of epistemology. This definition considered epistemology as not only the study of the structural aspects of entities, but also as the study of how knowledge is produced. Aligned with that vision and based on the work of Nutter (47), Guariano pointed out that epistemology refers to the nature and sources of knowledge (i.e. knowledge provenance) (46):

> *"Epistemology can be defined as "the field of philosophy which deals with the nature and sources of knowledge" (Nutter 1987). The usual logistic interpretation is that knowledge consists of propositions, whose formal structure is the source of new knowledge. The inferential aspect seems to be essential to epistemology (at least for what concerns the sense that this term assumes in AI): the study of the "nature" of knowledge is limited to its superficial meaning (i.e., the form), since it is mainly motivated by the study of the inference process." (46)*

More recently, in the medical informatics milieu, Bodenreider (48) dealt with this issue from a medical informatics perspective, defining epistemology as (48):

> *"The study of how cognitive subjects come to know the truth about given phenomena in reality – for example that they instantiate given classes or universals. In the sense that is relevant to our present purposes here, epistemology is the study of biological or medical knowledge. Thus it encompasses the ways in which physicians come to know about the existence of given diseases in given patients." (48)*

Leveraging the visions of Guariano, Nutter, Brachman and Bodenreider we define **epistemology as the domain that concerns the study of clinical knowledge provenance, including the structural constructs that allow for specifying such knowledge (i.e. the context in which knowledge is produced and its representation).** Recalling the knowledge classification performed in section 3.1.2, it is possible to observe that the epistemology domain concerns contingent knowledge (certainty, timing aspects, authorship, parties involved in knowledge development etc.) and schemas for data structures. It is possible to see how epistemology is actually the *leitmotif* beyond the specification of CIMs, since they convey these two types of knowledge. CIMs, at a technical level, define both the information structure (representation) and the context to capture the clinical reality in a health information system (HIS). Archetypes, HL7 CDA documents or FHIR resources define constraints on the structure of a clinical document, cardinalities on the relationships between the entities represented, the participants in a clinical process (parties), attributes to record the certainty of an evaluation, the interval of time and date when an observation was performed, etc. CIMs therefore allow for the implementation of the epistemic level of HIS.

8 Alignment of information models and domain ontologies

8.1 Binding models for semantic interoperability of clinical information

As section 3.1.1 explained, three main components need to be combined to enable semantic interoperability, namely reference models, CIMs and biomedical terminologies. On the one hand, Reference Models and CIMs concern the specification of epistemic aspects of clinical reality in health information systems. On the other hand, terminologies pertain to the specification of ontological aspects (i.e. statements covered by domain ontologies) of clinical reality in health information systems. For example, the reference model is present in HL7 (RIM) CDA and FHIR profiles, openEHR and ISO 13606 archetypes, as well. The other component is (biomedical terminologies) is ontological.

Reference models are constrained by certain rules and a multidisciplinary editorial process to build application domain models, called clinical information models (CIMs) (a.k.a. detailed clinical models) (10, 49). This is a process that involves challenges that are mainly related to the coordination of multidisciplinary groups to agree on the definitions of CIMs. However, CIMs by themselves are a necessary — but not sufficient — component to enable SIOp. SIOp needs the binding of CIMs with a standard terminology in order to attach semantics provided by a standard vocabulary to its sections and content. This involves a connection between epistemic and ontological aspects of the clinical reality reflected in HIS. With this aim, different health information standards have defined mechanisms for linking CIMs' elements to terminology concepts.

Terminologies can be used with CIMs that were implemented with different standards (HL7 CDA, FHIR, openEHR, ISO 13606, etc.). This has been the approach taken by most projects approaching interoperability in e-heath over the last two decades, but it is not exempt from challenges as examined in the following (8, 50, 51).

8.1.1 Terminology binding of clinical information models

The linkage of CIMs to a terminology is performed by assigning concepts from one or more standard terminologies to the different sections of CIMs, thus endowing CIMs with a certain level of semantics. This process is known as terminology binding. Terminology binding can be performed on two complementary levels: *semantic binding* and *content binding*. The following explains these two levels.

Semantic binding:
Semantic binding assigns a meaning to each section of the CIM, linking it to a concept from one or more selected terminologies. Figure 6 illustrates this type of binding using red boxes. For the archetype *Symptom* from the Norwegian CKM, the semantic binding assigns: a) the code *19019007* (corresponding to the concept *Symptom (finding)*) to specify the meaning of the archetype itself; b) the concept *16240800* (corresponding to *General symptom description (finding)*) to specify that the section corresponds to the general symptom description; and c) *162465004* (*corresponding to the concept Symptom severity (finding)*) to specify the meaning in SNOMED CT of the section severity. If one searches for these codes in a SNOMED CT browser, it is possible to see how that code has assigned the definition provided in parenthesis. This way, the information structure that the CIM defines is assigned a code that provides meaning in a standard terminology that is also translated into several languages.

Content binding:

Content binding defines the possible values that a particular field of a CIM may be assigned when an instance from it is created. For example, this type of binding is shown in the green box in Figure 6. In the example, content binding specifies that the content of the field severity can contain one code belonging to the set (162468002, 162469005, 162470006) corresponding to the definitions (Symptom mild, Symptom moderate, Symptom severe), respectively. Please note how it differs from semantic binding, which specifies the meaning of the section (i.e. the section for recording the severity of the symptom) rather than the possible values that such a section may contain. Languages, such as the SNOMED CT constraint language and query language, have been designed to facilitate content binding (42, 52).

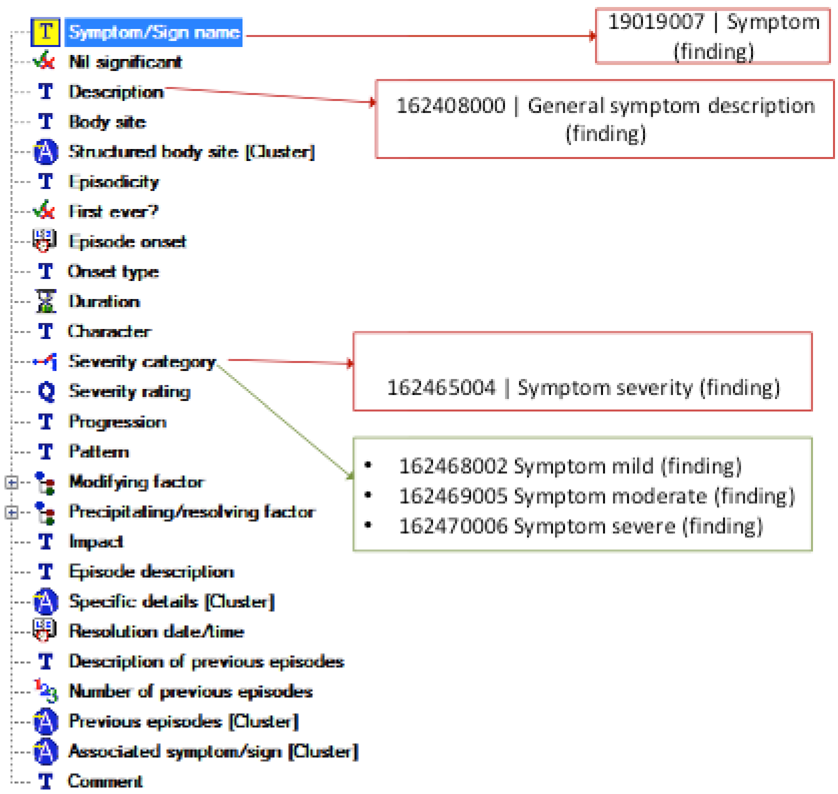


Figure 6: Semantic and content binding of the symptom archetype.

One possible use of the two types of binding is illustrated in Figure 7, where the EHR generates a General User Interface (GUI) form from the archetype using the semantic binding. The system asks a terminology server what concept should be displayed for the severity section of the clinical document. In addition, the GUI generator may query a terminology server to find out the definitions of each of the options that the user is allowed to select based on the content binding. These labels could be automatically translated into any of the languages that SNOMED CT supports.

Another advantage of binding CIMs to standard terminologies is the mapping of different terminological systems. For example, a system using ICD-10 or LOINC can be mapped in a straightforward manner to SNOMED CT by simply checking the public equivalence tables between the two terminologies (53).

This is particularly useful for establishing SIOp across systems that rely on different terminologies. More details on the use of the terminology to bind information are available at the SNOMED CT guidelines (35, 54, 55).

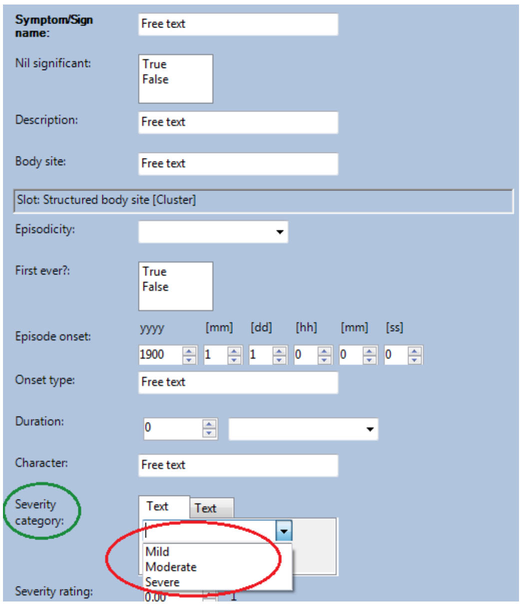


Figure 7: User interface generated using semantic and content binding.

8.1.2 Challenges in terminology binding: epistemology vs. ontology

Terminologies like SNOMED CT provide complete and sound guidelines for terminology binding (35, 54). However, defining a clear separation of concerns to leverage the use of CIMs and terminologies optimally remains a challenge. Although in theory there is a separation between epistemology and ontology, in many cases, what to represent in one model or another is ambiguous and dependent on the use case (9, 25, 51). CIM terminology binding involves the alignment of some of the different models of knowledge explained in section 3.1.2. Specifically, the knowledge models aligned are information schema and contingent knowledge (i.e. epistemological aspects conveyed by CIMs), and ontological knowledge implemented as a biomedical domain ontology. This involves well-known challenges since the border between what is represented in one or the other is often unclear (9, 56). Indeed, this is the key issue when using information models and terminologies to enable SIOp. Theoretically, their differences are clearly stated, but in reality the implementation of both models (CIMs and ontologies) usually enters into the domain of the other (to a certain extent) for practical reasons and because of requirements identified during their development (57).

At this point, the reader with some background in SNOMED CT may wonder why, if SNOMED CT is a domain ontology, and knowledge provenance and context (epistemology) should not be part of the domain ontologies, why does a hierarchy exist in SNOMED CT called *Situation with explicit context* with the purpose of defining SNOMED CT expressions with contextual properties? This is, in fact, a clear example of the intrusion of epistemic aspects into a domain ontology (57). This intrusion of epistemic aspects into terminologies is common in many terminologies, such as ICD (48). In the following, we will document the main problems associated with the terminology binding of CIMs, but those interested in the different types of epistemic intrusion and inconsistencies affecting terminologies independent of the CIM may consider reading Bodenreider et al. (48).

The intrusion of epistemic aspects in terminologies, or ontological aspects in CIMs (e.g. EVALUATION, OBSERVATION and ACTION classes in openEHR) (58), raises the question of where to specify certain information — on an information model level or on a terminology level? An illustrative example is presented in Table 2 that contains a SNOMED CT expression specifying the history of an adverse reaction to penicillin. The example shows aspects such as time, severity (dependent on each particular case) and certainty about the cause of the adverse reaction assigned to penicillin (i.e. "*Probably present*"). These are pure epistemic aspects rather than the invariable truths of a domain. In fact, these aspects are specified in the CIMs for adverse reaction.

```
161632004 |History of risk factor|:
   { 246090004 |Associated finding| =
     ( 281647001 |Adverse reaction|:
        {246112005 |Severity|= 24484000 |severe|,
         42752001|Due to|=6369005|Penicillin|}),
         408729009|Finding context|=410592001|Probably present|}
```

Table 2: SNOMED CT expression with epistemic content

8.1.2.1 Challenges selecting the binding approach

One of the main benefits of using clinical information standards and terminologies is the ability to count on a generic, consistent method of representing clinical information, thus enabling SIOp. SIOp's main objective is to guarantee that the same conclusions are derived when different stakeholders interpret a dataset. However, terminology binding involves challenges that, if not treated carefully, may hamper the unambiguous interpretation of clinical information, thus affecting SIOp. When it comes to ontology-based terminologies that rely on formal semantics, the challenges in terminology binding become exacerbated. The approach of terminology binding shown in Figure 8 endows the CIM with a certain level of semantics by attaching a code that refers to a concept in a standard terminology (e.g. SNOMED CT or ICD-10). However, these semantics correspond to concepts that comply with an external ontology (i.e. the SNOMED CT concept model). Therefore, they are not formalized within the archetype model. This means that the ontology that unambiguously defines the semantics of the concept is detached and external to the conceptual model that the archetype represents. This often causes divergences in the semantics intended between the CIM designers and those represented by the domain ontology, as documented by Markwell (51). For example, Figure 9 shows different alternatives that an information architect may encounter when binding the archetype *Family history* to SNOMED CT to document a case of a relative diagnosed with diabetes mellitus.

Figure 8(a) shows how the root of the archetype holds a semantic binding with the SNOMED CT expression *Family history of: / Associated finding: / Diabetes mellitus*. This alternative for annotating the archetype presents two main challenges. First, part of the expression to annotate the archetype is known at the design time (red box in Figure 8.a), while its specific use to record diabetes mellitus is decided at runtime (see green text). Therefore, the expression would need to be completed at the run time when creating an instance of the archetype, assigning the part decided at the run time to diabetes mellitus (in green in Figure 8.a). Second, only the archetype root is annotated with the SNOMED CT expression. This allows for classifying the expression and performing expressive queries to find instances annotated with it. However, the element *Problem/diagnosis name* of the archetype lacks semantic binding, introducing challenges in the use of the terminology to interpret the data structure when instances are interoperated (e.g. automatically translating sections of the archetype). Alternatively, if the instance were annotated only at the run time, the instance rather than the archetype could rely on the pre-coordinated equivalent expression *430678008/Family history of diabetes mellitus type 1 (situation),* instead of the post-coordinated one. For this approach to guarantee SIOp across different organizations they would need to agree on the way of binding the archetype since at design time the binding is not known. Rather the binding is performed at runtime by the internal logic of each HIS.

In Figure 8(b), the root of the archetype openEHR-EHR-EVALUATION.family_history.v1 holds a semantic binding to the concept 57177007/Family history. The element for recording the *Problem/diagnosis name* of the archetype has a semantic binding specifying that the element contains a diagnosis name. Also, the *Problem/diagnosis name* has been assigned a content binding by using the SNOMED CT constraint language, which indicates that the possible values are descendants of the clinical finding.[11] At the run time, a particular clinical finding (i.e. diabetes mellitus) is assigned as a value of the archetype element. The intended semantics are that a family member of the patient had diabetes. However, that is not formally described following the SNOMED CT concept model, as it was in Figure 8(a). Therefore, if a query on the EHR based on terminological concepts returns the diabetes value and the results of the query are not properly post-processed, one may interpret that the patient, rather than the relative, has diabetes. Interpretation of the position of the bindings in the hierarchy of the CIM elements (*Family history* in the root and then *Diabetes* in the *Problem/diagnosis name*) is necessary in order to understand the intended semantics. However, post-processing based on the CIM structure after performing a sematic query is needed as in the approach presented in(59).

In Figure 8(c), only the element of the archetype is annotated at the run time with the value. The expression is a SNOMED CT expression and therefore could be classified and used in expressive queries based on formal semantics. However, if a semantic query were performed using the term, a mechanism to refer back to the information structure would be needed. Moreover, this pushes the IT infrastructure to work at two levels as described previously, i.e. syntactic and semantic. This introduces some challenges in the definition of expressive queries "on the fly" that, as explained before, are needed in contexts such as clinical research. Again, this situation is equivalent to setting the content binding to the pre-coordinated concept *430678008/Family history of diabetes mellitus type 1 (situation)* at the run time.

[11] The symbol next to the concept corresponds to the valid range of the *associated finding* relationship. The symbol is interpreted as "descendants only (stated) except for super-category groupers and descendants in a qualifying relationship". Therefore it specifies by means of the SNOMED CT constraint language the possible values valid for that node.

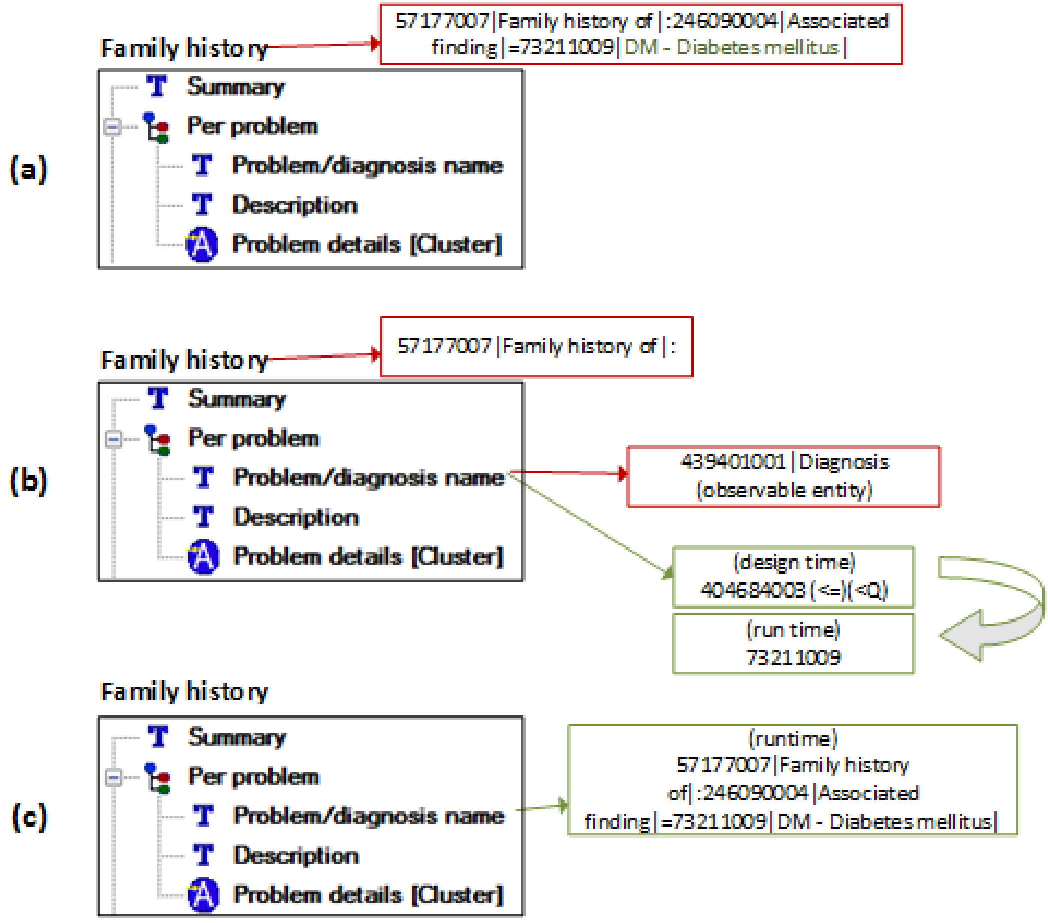


Figure 8: Different alternatives for archetype terminology binding.

At the moment of writing, openEHR does not restrict any of the approaches above except for the limitations that each archetype edition tool may impose to support post-coordinated expressions. However, there is an open discussion on weather post-coordinated expressions should be allowed in the archetype specification or they should be referenced using URIs and linked data principles (60) pointing to an external resource as proposed elsewhere (25). While the first approach allows having both the terminology expression used in the binding and the CIM definition in one artifact; the second one allows keeping the terminology detached and performing separate maintenance over it. Which approach is more appropriate depends on the context of deployment and the availability of implementers to access an external terminology server.

In the case of other standards such as HL7 FHIR, the approach taken for terminology binding is to specify directly the way of performing the terminological binding focusing mostly on content binding.

HL7 FHIR has already predefined sets of codes, known as ValueSets, that are shared by several resources to perform content binding using its CodeableConcept[12] and Coding[13] data types. For example, the FamilyMemberHistory[14] FHIR resource represented in Figure 9 binds the element reasonCode to a the predefined ValueSet clinical-findings for specifying the possible values that can be used to specify why the family history of a relative was recorded. In addition, it binds the code that represents the element conditionOf the relative to the ValueSet condition-code. Both ValueSets define the same set of codes but are kept separated for management purposes. In specific they define the set <<404684003[15] (Clinical finding and its descendants). The advantage of maintaining ValueSets is that they can be reused among different FHIR resources. As explained, the element “condition” of such resource refers to the disorder of the relative and contains a CodeableElement to perform the terminology binding. Note that, in opposition to openEHR, it restricts the possible ways of performing terminology binding since it specifies explicitly that the CodeableConcept condition must be bound to codes that specify the “condition suffered by the relation”[16]. This means that this code will refer to a condition that is the one sufferend by the family relative. This disables the use of the context model to post-coordinate using the concept *57177007 / Family History of*. This way the CIM itself is telling the terminologist that an approach similar to the one of Figure 9 (b) using the specified ValueSet should be followed rather than leave it open for the terminologist to decide. An additional example is displayed in Figure 10 where the FHIR profile MedicationAdministration contains two content bindings. The first binds the element site to a predefined FHIR ValueSet that contains all the possible values in SNOMED-CT that can be used to specify an anatomical location (<<91723000). The second value set allows binding the element route to a ValueSet that contains all the *allowed SNOMED-CT* codes to specify a route of administration *(<<284009009, the concept Route of administration (qualifier value) and its descendants*). As it occurred in the openEHR examples, the infrastructure that performs queries over clinical data should work at two levels, one at the semantic level checking the terminology ontology and a second one applying filtering criteria over the resource properties. Therefore post-processing to derive consistent inferences is needed as explained before.

[12] https://www.hl7.org/fhir/datatypes.html#CodeableConcept

[13] https://www.hl7.org/fhir/datatypes.html#coding

[14] https://www.hl7.org/fhir/familymemberhistory.html

[15] In SNOMED-CT constraint language <<[code] means the set of code defined by all the descendants of [code] and [code] itself.

[16] https://www.hl7.org/fhir/familymemberhistory.html

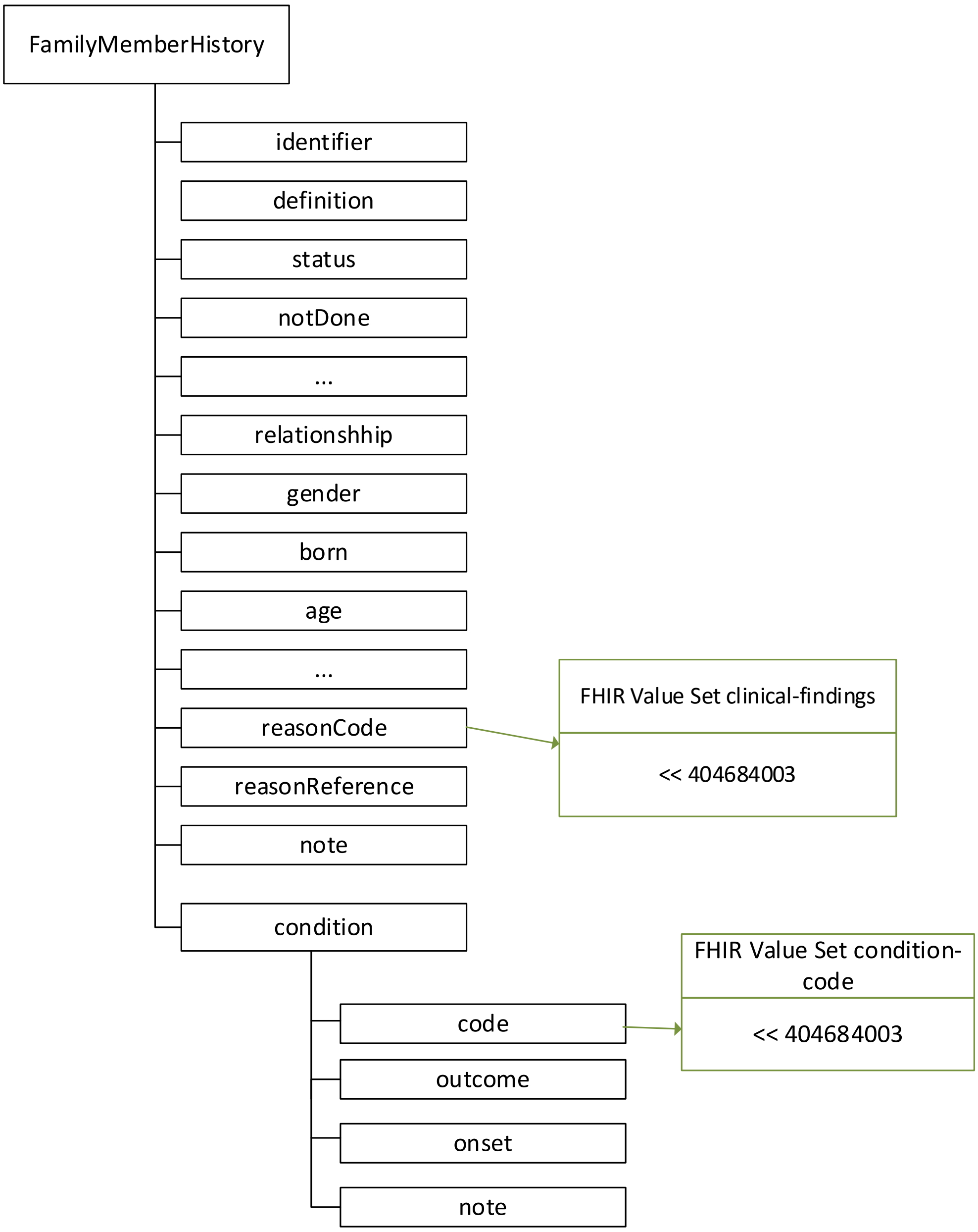


Figure 9: FamilyMemberHistory HL7 FHIR resource binding.

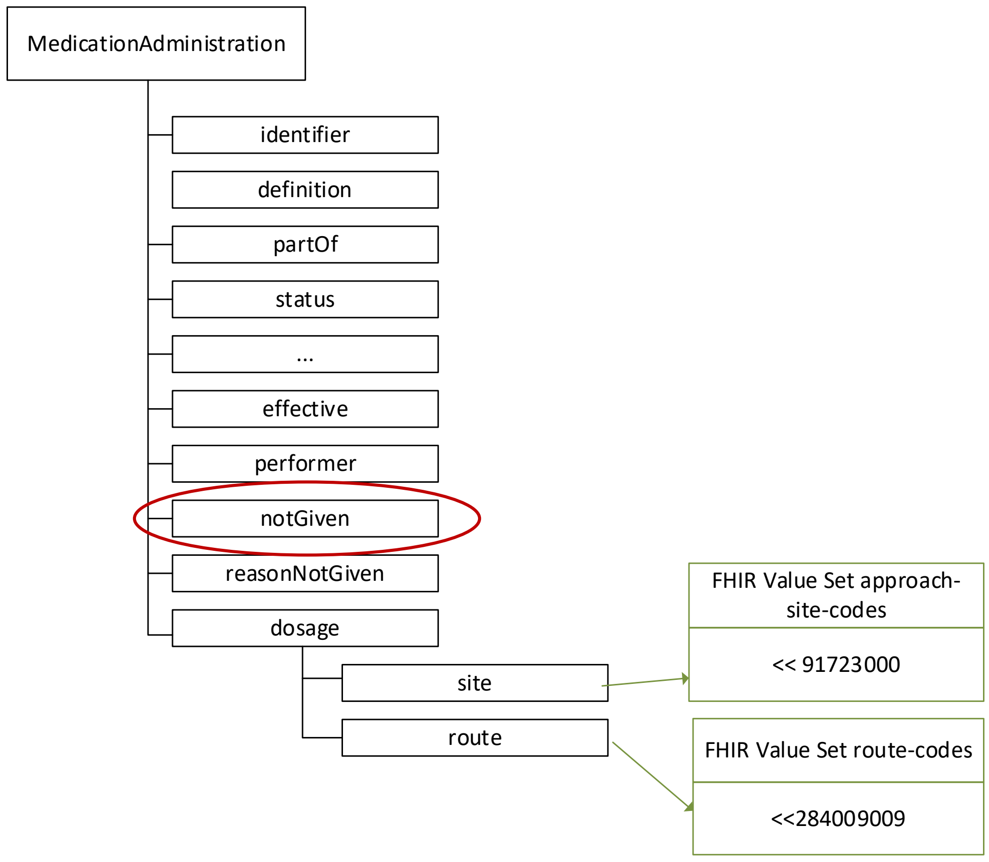


Figure 10: MedicationAdministration HL7 FHIR resource binding.

The previous examples showed the implications of making different choices in the terminology binding of CIMs when several elements of the CIM are involved in the specification of the intended meaning. Regardless of the decision made, each situation presented several challenges to deal with. If these situations are not documented and treated carefully by information architects, they may result in wrong inferences and erroneous (or at least confusing) queries, as reported by Rector (61).

A more serious problem is present when one model (the information model) introduces an element that radically varies the semantics presented by the other model (the terminology). In order to understand this situation, let us recall the case of the archetype 'symptom', previously displayed in Figure 8.

The second element starting from the top is the *Nil significant* element. The description of the element in the archetype is:

> *"The identified symptom or sign was reported as not being present to any significant degree."*

This means that, if the *Nil significant* element (of a Boolean type) is marked (i.e. set to true), then it is specifying the negation of the existence of the recorded symptom (i.e. the specified symptom is not present). The archetype of Figure 8 is annotated with the concept *19019007 | Symptom (finding)*. Together, the archetype and the annotation represent the negation of the existence of a symptom (at an information schema level). However, that negation is not present at the domain ontology level. Rather, the negation is specified in a conceptually different knowledge model (information model) that is even specified in another language (ADL, in the case of archetypes). Again, if one uses semantic web technologies to perform a query over the terminology but does not explore the semantics implicit in the data structure, this may lead to making the wrong inferences.

The same situation occurs in other standards. In FHIR the specification of negation/absence is still a matter of discussion[17], but some resources already allow it. For example in HL7 FHIR the resource MedicationAdministration shown in Figure 4 contains the element notGiven (marked by a red circle), which is of type boolean (true/false). When set to true it specifies that the medication recorded by the resource has not been administered. It is possible to see how the challenges in terminology binding arise irrespectively of the standard used, although it is possible to restrict the binding task by using value sets and explaining how to use them as in the case of HL7 FHIR.

8.1.2.2 Challenges expressing epistemics on the domain ontology side

One could argue that some of the challenges presented in the previous section could be overcome by relying more heavily on the terminology, by expressing negation and some contextual aspects through the patterns provided by SNOMED CT's context model (i.e. defining epistemic information as *Situation with explicit context).*[18] In order to understand the challenges of this approach, let us perform an experiment pushing the SNOMED CT compositional grammar as far as we can to create the most complete representation of the archetype *openEHR-EHR-EVALUATION.adverse_reaction_risk.v1* (which has a high epistemic load) possible. This experiment relies on the terminology as much as possible, including contextual aspects of the terminological expression. Although SNOMED CT was not designed for this purpose, the experiment allows us to unveil the challenges that appear when one mostly relies on the clinical domain ontology. The left side of Figure 10 presents a template for recording the list of adverse reactions; on the right, the terminological projection of it in SNOMED CT is shown. Different issues can be observed related to this approach (62):

<u>Expressivity-related challenges:</u>
Challenges in expressivity appear when the meaning of a concept cannot be fully expressed using the compositional grammar. Some occur because the archetype element coded has a candidate in

---

[17] http://wiki.hl7.org/index.php?title=Negation_Requirements

[18] The reader must be aware that this is not really a negation in the DL manner. Just a pattern to overcome the problem of tractability since negations would make SNOMED CT classification intractable.

SNOMED CT that introduces a slight variation in the semantics originally intended in the archetype. That is the case of *Reaction mechanism* (archetype element) vs. *Immune hypersensitivity reaction by mechanism (disorder)* (SNOMED CT concept). This problem has been solved elsewhere by extending SNOMED CT to local implementations (63); however, since nationally approved archetypes aim to be as generic as possible, divergences between the SNOMED CT implementation used for coding and the standard release should be avoided.

Sometimes the variation in semantics occurs as a consequence of relying on a SNOMED CT attribute that has a broader meaning. An example is given in Figure 11, where the element *substance* can be mapped to the attribute *due to* of the SNOMED CT expression, which has a wider range than only substances (any *Clinical finding*; any *Procedure* etc.). Another loss of meaning occurs in the example where the archetype element *onset of reaction,* with an openEHR data type date/time, acquires the meaning *date and time of the onset of reaction*. However, the SNOMED CT candidate for matching is *Date of last episode*. One should note that it is unclear whether this data refers to the onset or cessation.

Other expressivity problems occur due to not being able to express rich temporal semantics. For example, the context model does not allow temporal order to be expressed. This can be seen in Figure 11, where the section *Reaction event* (with the meaning *previous adverse events*) is mapped to the expression section with the *Event* that wraps the finding *adverse reaction*. It is not possible to specify the kinds of semantics that often occur in patient summaries of drugs, allergies, etc. The *time context* SNOMED CT attribute allows for expressing if the situation occurred in the present or the past, but not the order of the events. Besides, in the compositional grammar, the *time context* attribute cannot be used as a valid attribute of an *event*.

<u>Incompatibilities between the information and the domain ontology model:</u>
Other types of issues are those related to the restrictions that the compositional grammar imposes. For some nodes, there exist valid candidates in SNOMED CT, but it is not possible to include them using the compositional grammar, or the expressions need to be over-complexified to include them. This is the case of the mean arterial pressure (MAP) element in the blood pressure archetype. One could choose *163020007|O/E - blood pressure reading|:* to represent the main clinical concept and associated systolic, diastolic, etc. However, MAP cannot be set in the expression of the blood pressure reading. It could be set using *associated with|=6797001|Mean blood pressure|* but MAP is in the Observable Entity hierarchy and the property *associate with* does not mark *Observable Entity* values as permissible values in its range. In some cases, elements of the archetype cannot be coded due to a total lack of available candidates. Examples are the *Duration of exposure* or *Initial exposure* elements in Figure 11. This is usual in concepts whose semantics are purely contextual as time related properties.

Previous challenges have focused on the expression of clinical data (data section in archetypes) terminology, but the attributes of the protocol section have much lower coverage. Typically, the protocol section (according to the openEHR specification) covers different properties depending on the openEHR entity. For OBSERVATIONS it covers the method used, for EVALUATIONS it covers how the evaluation was performed and for INSTRUCTIONS it covers how they should be executed (51). All of these are contextual properties that escape the competency of a domain ontology, and therefore

SNOMED CT. For instance, examples appear in the Blood Preassure archetypes for "*cuff size*" or "*systolic pressure* formulas".

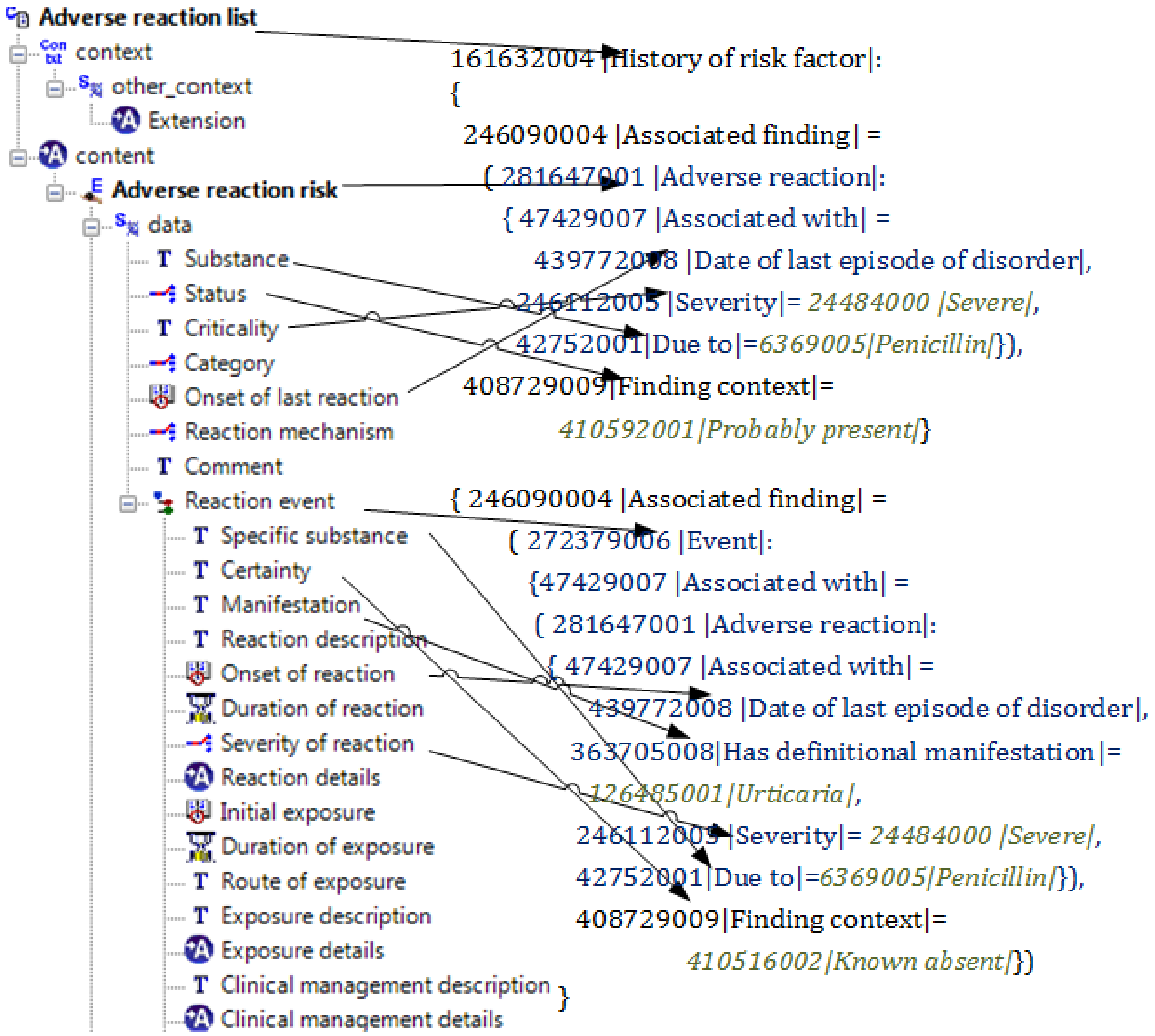


Figure 11: Attempt to project the maximum amount of information conveyed by the archetype into SNOMED CT.

The examples have focused on the most problematic situations that appear during terminology binding when CIMs attempt to accurately describe their semantics in order to maximize their chances of SIOp. Not all situations present all of these problems, and current guidelines may cover most of the scenarios for the use of terminologies (35, 54). However, in scenarios in which a high degree of expressivity is needed to maximize SIOp (such as clinical data reuse across organizational boundaries), these situations need to be tackled carefully, guaranteeing precision and a certain level of formality in the expression of both epistemic and ontological (i.e. belonging to a domain ontology) aspects. Otherwise, inconsistencies in the interpretation of the data may arise when the same dataset is interpreted by different stakeholders (31).

At this point, one could propose going beyond the use of semantic web technologies to express the information model in a language that would allow for reasoning in order to be able to connect the information model and the domain ontology in some sort of semantic layer that would allow for reasoning in both the CIM and the terminology. Some studies have explored this possibility by using languages that rely on DL to express CIMs (45, 56, 57). However, as anticipated in Chapter 6, increasing the level of expressivity using DL has a price in terms of its computational costs.

9 Language vs. type of knowledge

In the previous section, the differences between the information model (epistemology level) and the domain ontology model were explained. It was demonstrated that several options exist to bind CIMs to terminologies. It was also shown that establishing the boundaries between what to represent as CIM or terminology is not always clear, and how, depending on the design decisions made during terminology binding, different challenges appear at the implementation level.

All the examples expressed information models with languages designed to define a syntactic model of the clinical information structure (ADL, XML, JSON, etc.). However, SNOMED CT works at a formal level specifying expressions with DL. It is possible to appreciate how, besides expressing different types of knowledge (epistemology and ontology), the language used for their representation is in many cases different as well. Nevertheless, the same language could be used to specify both models. For example, OWL could be used to represent both CIMs and the domain ontology, as in Rector et al. (61), thus alleviating some of the issues presented in the example shown in Figure 11.

The decision on which language to use to represent different types of knowledge is often a source of confusion that is important to clarify. It is paramount to understand that the language used to build a model (ADL, XML, JSON, OWL, etc.) does not determine the type of knowledge expressed (information schema, domain ontology, procedural, etc.) (18). This means that an archetype or FHIR resource expressed in OWL does not become a domain ontology. In the same way, a biomedical domain ontology expressed with a syntactic language (e.g. XML) does not become a CIM. Regardless of the language used to implement the model, the domain ontology is still domain ontological knowledge (invariable truths that account for existence), and an information model continues to express an information schema with contextual properties (epistemology). Nevertheless, as in the case of the concept *Situation with explicit context*, or the CIMs defined by Rector (61), the language may be the same. A good example of this is the opencimi.org initiative that specifies CIMs in different languages, including XML, JSON and OWL, so implementers can choose the one that best fits their needs.

A question that commonly arises when the languages available for the semantic web are analysed is whether it is worth specifying information models with languages that allow for reasoning (e.g. OWL). This way, reasoning and expressive queries could be run in a straightforward manner, and challenges in binding CIMs to domain ontologies may be alleviated to some extent. In essence, CIMs can be expressed with DL using languages such as OWL (56). However, the specification of CIMs covers aspects such as cardinalities, optionality between elements, negations or restrictions on data instances that are applied at the run time. These aspects can be expressed in OWL, but ontology reasoners cannot process them in a tractable manner. For example, many of the restrictions specified with cADL (the ADL constraint language) or choices of objects (present in FHIR) cannot be expressed with OWL EL++, and they would need a more expressive profile (e.g. OWL DL) to be specified (56). In large enterprise infrastructures, this may pose a problem.

Another issue that concerns the choice of the language and the technology to manage CIMs is the dichotomy between the closed-world assumption (CWA) and the open-world assumption (OWA). On the one hand, The CWA “states that everything which is not explicitly true is considered to be false” (43). This means that if one queries an information system (e.g. implemented with a relational DB) for the list of drugs administered to a patient while she has been at the hospital, those drugs that do not appear in the list will be considered to have not been administered.[19] On the other hand, the OWA considers knowledge models to be incomplete by definition, thus it considers that if something is not in the information system, it is unknown. The inference from querying a KB of the drugs given to the patient during a period of time would not assume that drugs not present in the KB have not been administered to the patient. It is considered unknown unless the opposite is explicitly stated (i.e. penicillin NOT administered).

Most of the technologies used to implement information systems work under the CWA premise. Examples are relational database management systems. However, due to the nature of domain ontologies, a statement may be true irrespective of its presence in the model represented by the ontology. Therefore, in order to consider a statement false, it must be explicitly negated. By definition, ontologies assume their incompleteness in order to be able to accommodate new statements as the knowledge of the domain is extended, hence the OWA is more appropriate. Information management systems often operate better with the CWA since it is more natural for developers and makes the definition of data schemas more efficient, whereas knowledge management systems often prefer to work under the OWA to escalate while new knowledge is incorporate to their KB. Several initiatives have already allowed the specification of openEHR (56, 64) and FHIR (65) OWL and RD(S) respectively. In fact, at the moment RDF(S) is one of the formats to represent FHIR. Therefore, at a technical level it is possible to work with CIMs using sematic web technologies. Nevertheless, the choice of a language to use to represent one type of knowledge or another is dependent on practicalities. Different languages have different properties that affect their suitability, depending on each scenario. It is left to the information architect to leverage the different languages and technologies to define the architecture of each health information system, depending on the context and its intended use (56).

[19] This is from an information system point of view. The user (clinician) may interpret this with the Open World Assumption regardless of the assumption made by the technology that implements the information model.

10 Semantic web architecture, domain ontologies and interoperability

10.1 Semantic web architectures

Results of projects such as ASSESS-CT may provide useful advice to adopt SNOMED CT for general use. A recent review indicated that the studies reporting the implementation of SNOMED CT have increased in the last years (66). This makes it reasonable to expect that, in the medium term, SNOMED CT may finally become the de-facto reference terminology for most EHRs. However, the use of other domain ontologies and their combination into infrastructures that make use of all their expressive power is still a subject of research in medical informatics (67, 68). In order to use several ontologies to represent clinical information and knowledge, not only guidelines on their use and best practices are needed, but also a software architecture that is able to leverage several ontologies and coordinate their use with the EHR.

Semantic web architectures that exploit different domain ontologies have been used in eGovernment, eScience and eBusiness (69, 70). When it comes to large scale enterprise semantic architectures, one of the most known applications is the architecture used by the British Broadcast Corporation (BBC) to manage the complexity in automatically classifying news categories and recommending content (71). However, the complexity inherent of the healthcare domain poses a great challenge in implementing these types of infrastructures at large scale. These architectures often rely on the paradigm of Semantic Web Services (SWS) to leverage the use of syntactic technologies for information pro-cessing (XML, JSON, WSDL, etc.) as Service Oriented Architectures that are extended with a semantic layer. This semantic layer defines the meaning of the messages exchanged, the functionality, and other proper-ties of Web services in a machine-interpretable manner (68, 72). SWS often rely on the full representation of information at a semantic level when data is shared across different services. This means that data is transformed into a semantic representation in RDF(S) or OWL when it is sent to a service and, at the other endpoint, it is transformed into a syntactic level that may be different from the previous one. This provides the highest level of SIOp possible since data is machine interpretable. In addition, different end points may operate with different information models but be able to exchange information without ambiguity relying on a common ontological model.

However, the serialization of biomedical data at a semantic level would involve a complexity and amount of resources that may jeopardize enterprise developments because of the reasons explained in the previous sections. In addition, the combination of different ontologies is not a trivial task since it requires extensive mapping (matching) across their concepts which has proven to be challenging (73). This matching may introduce variations in the semantics originally intended by the ontology designers. Many of the ontologies available nowadays have risen from particular application domains. Therefore they are very disparate in structure and granularity becoming their alignment very difficult (73). The OBO Foundry has recently proposed the use of the BFO as an upper level ontology to alleviate some of the problems that arise when different ontologies must be combined (67). However, most ontologies were not designed in a generic way thinking that many other users may want to reuse them, but to provide a model for a particular domain and scenario. In addition, for the implementation of systems relying on varied ontologies, a clear commitment with W3C standards to drive semantic web implementations would be needed. ASSESS-CT also reports as one of its recommendations the need for a better alignment of SNOMED CT with semantic web standards (74).

Different uses of the ontology model may lead to different architectural requirements. Depending on the use case, biomedical ontologies can be used as mere code systems, using their concepts as codes that are common to a domain, or as formal knowledge models expressing the relationships between different concepts in machine-interpretable semantics. This leads to three possible levels in the use of formal semantics:

- **Level 1 — No use of formal semantics at all:** the biomedical ontology is used as a controlled vocabulary where its terms are used only as a commonly agreed-upon glossary without any formality. Information can be queried relying on the controlled vocabulary, but no relationships among the terms of the ontology are analysed formally.

- **Level 2 — Partial use of formal semantics:** this scenario involves the use of the formal semantics of the ontology by an external terminology server that is able to classify post-coordinated expressions and can be used to perform searches of EHR information bound to a terminology. This level uses reasoning over the reference ontology to classify terms and performs searches over the terminology hierarchy. In this scenario, some of the EHR information contains semantic or content bindings to reference terminologies.

- **Level 3 — Full use of formal semantics:** in this scenario, CIMs contain extensive bindings to reference terminologies and they are partially or fully mirrored in a model that allows for reasoning. Such a model allows for reasoning over several domain ontologies and the information model structure, thanks to interlinked relationships expressed in formal semantics (25).

Level 1 is the one used by the majority of systems that currently rely on terminologies such as ICPC 2, ICD-10, LOINC, etc. Also, many systems use SNOMED CT as a coding system to annotate CIMs and share EHR extracts without exploiting its ontology. Level 2 is, to the authors' knowledge, only implemented by certain organizations that use some of the expressivity of ontologies to implement queries that require subsumptive reasoning (75, 76). Level 3 is only elucidated in a few experimental studies (9, 25, 77). These studies are setting the basis to a better combination of CIMs with terminologies to manage aspects such as negation and disjunction (77); but, the challenge in leveraging many ontologies to describe and share EHR information exploiting ontologies properties is still relatively unexplored in the field of medical informatics. With regards to this aspect, HL7 FHIR currently allows serializing to RDF(S) (78) and openEHR can be serialized to OWL (79), thus allowing to operate the information model as a semantic model and reason over it. This view is also shared by the Yosemite project that focuses on investigating how to implement SIOp by relying on semantic web technologies (80). In the future, these initiatives may enable to perform reasoning and precision queries over health information. However, there are important challenges that will require extensive resources to make possible reaching the level 3 in the use of formal semantics since. The reason is that implementing Level 3 would require combining different domain ontologies with information models in a way that allows for reasoning. This presents challenges associated to the problem of reuse and ontology matching. In order to overcome them, the semantic infrastructure must guarantee that the different ontologies developed are compatible and scalable by committing to several design principles.

In the next section, the report presents the efforts to approach the challenges in ontology reuse and compatibility by developing best practice and top-level ontologies.

10.2 Open Biomedical Ontologies Foundry

The Open Biomedical Ontologies (OBO) foundry is an initiative run by different ontology developers from life-sciences (81). Its purpose is to allow the collaborative development of biomedical ontologies guaranteeing their interoperability and scalability (81). In this regard, the OBO Foundry has proposed a set of principles that those ontologies that wish to be part of it must commit to. These principles aim to guarantee the consistency of the ontologies developed from a logic and conceptual point of view (82).These principles include open use, collaborative development and avoid overlapping developments (82, 83). Based on these principles the OBO Foundry pursues (84):

- supporting terminologies that allow for representing research domains;
- enabling the integration of heterogeneous information repositories;
- and enabling semantic interoperability.

The former principles represent not only 'good practice' from the OBO Foundry point of view, but also from the ontologies development point of view. Therefore ontologists are advised to follow these principles, even if they do not plan to submit their developments to the OBO Foundry (85).

At the moment, the OBO Foundry contains around 60 ontologies that cover domains such as genomics, proteomics or anatomy (82). The ontologies developed have led to the availability of large semantically annotated bodies of data such as those related to gene products (86, 87). The National Institutes of Health's Roadmap and the National Center for Biomedical Ontology (NCBO) support OBO as a biomedical ontologies provider through BioPortal (88).

The biomedical domain has driven the development of ontologies for different purposes such as knowledge management, semantic search functionality, data annotation, decision support and information exchange (89). However, these developments have not always been governed in a way that could grant interoperability resulting in information silos (84). This has made ontology reuse one of the most important challenges in the adoption of different ontologies by HIS. The principles behind the OBO Foundry and the development of the BFO intend to approach the challenge of ontology reuse and scalability.

These principles are helping to improve the reuse, scalability, and interoperability of ontologies; but reuse of terms is still low across ontologies and needs to improve. In 2009, Kamdar et al. (89) looked at 337 ontologies from the BioPortal repository, and found a term overlap of 25.31% to 30.18%, and a term reuse of less than 9%, with most of the ontologies reusing less than 5% of terms. A more recent study from 2011 performed by Ghazvinian et al. (90) reported that that only between 4.0% and 5.4% of terms were reused, while the overlap was between 3.1% and 3.5%. Reused terms were identified by specific mappings between the ontologies, while overlap was found by using a lexical mapping method to discover pairs of terms with matching preferred names, but without any reference to one another.

Using the BioPortal web tool[20] to look up all the OBO Foundry member ontologies to see the degree to which they map to SNOMED CT, we found that the total number of mappings to SNOMED CT was 7,202, distributed as follows:

Ontology	Mappings to SNOMED CT	Total mappings
GO	197	273,112
PR	823	136,784
PATO	318	35,134
OBI	205	49,793
PO	18	18,785
CHEBI	5 216	407,359
GO-EXT	226	1,474,494
ZFA	112	89,194
XAO	87	38,274

Table 3: Mappings from OBO member ontologies to SNOMED CT

For an up-to-date visualization of mappings between ontologies, we recommend the Ontology Xref Service by the European Bioinformatics Institute.[21]

10.3 Basic Formal Ontology

The basic formal ontology (BFO) is an upper-level ontology for integrating different domain ontologies and ensure their scalability and interoperability (91). Therefore, the BFO does not contain domain-specific concepts such as diseases, anatomical parts or drugs that are described by domain ontologies.

As a top-level ontology, BFO does not meet the needs of experts for expressing the concepts of particular domains. Rather, it provides the scaffolding for developing ontologies for specific domains in a scalable manner in order to facilitate their posterior reuse. In accordance with the description of top-level ontologies above, the BFO provides a top-level structure in which information is compiled in separate repositories that form part of a shared framework of entities represented by lower-level (e.g. domain) ontologies (67). Committing to BFO allows ontology developers to relief the problems of ontologies mismatch that derive from parallel development of different ontologies without proper coordination. This is paramount in order to enable the use of disparate biomedical ontologies in EHRs.

[20] https://bioportal.bioontology.org/mappings

[21] http://www.ebi.ac.uk/spot/oxo/index

11 Summary of questionnaire responses

In this chapter, we will summarize the responses given to our questionnaire[22] by the respondents who were academics, developers, vendors and clinicians. Some of the responses are subject to non-disclosure agreements, and we have therefore chosen to render the answers using a collective summary method. Quotes are used with permission from the respondents.

Note that this chapter is based on the questionnaire responses, and does not contain any discussion from the authors of the report.

11.1 Transition to ontology-based terminologies

11.1.1 Use of ontology-based terminologies among respondents

The first set of questions in the terminology part of the questionnaire concerns challenges, consequences and prerequisites for transitioning to ontology-based terminologies. This section summarizes the responses given.

Among the respondents that had adopted and utilized an ontology-based terminology, SNOMED CT was by far the most common. The vendors all reported that their products were terminology-neutral, and that many proprietary terminologies could be plugged into the systems. The vendors all stated that the terminology was used in an integrated application suite for administrative, clinical and research purposes, such as billing and coding, pathology description, warning scoring, analytics and adverse event detection.

All respondents stated that they exploited the formal semantics in the ontology (SNOMED CT) to some degree, at least. One noted that the *is_a* semantic was the most generic and effective, since it is shared by most other ontology systems; moreover, their solution allowed for use-case specific semantics/relationships to be developed. The hierarchy defined by the SNOMED CT structure was also used to group related (patient) records, making analytics and reporting much simpler since records did not have to be selected individually, as described below:

> *"An administrator creating a decision support advisory for diabetic patients could use a single diabetes grouper as one of the locator criteria instead of identifying every relevant diabetes diagnosis."*

Much in the same way, the ontological structure of the terminologies can also be helpful when one organization or system pulls clinical data from another organization or system database. In addition, the expressive semantics in the ontology-based terminologies can provide the ability to express questions for querying at more complex levels of ambiguity, and also makes it possible to query both structured and unstructured data, according to one of the vendors in the group of respondents. One respondent also noted that it can minimize query time, as you do not need to formulate complex SQL to get accurate and comprehensive answers. This was especially useful when querying SNOMED terms, either individually or multiple terms grouped together, and retrieving EHR data mapped to these terms, according to one respondent.

[22] See appendix for full questionnaire

11.1.2 Benefits of transitioning to ontology-based terminologies

The perceived benefits of using ontology-based terminologies were also shared by the respondents, with the main benefits being summarized by one respondent, as follows:

- *"Interoperability*
- *Disambiguation of the clinical terms*
- *Simplification of authoring process, generic rules, easy maintenance*
- *Improved accuracy in run time provided EMR ontology-based terminology*
- *More accurate and comprehensive analytics*
- *Ability for computer-based automated analysis of unstructured data*
- *Ability to maintain rule-sets*
- *Consistency of application"*

Another respondent stated that using an international terminology such as SNOMED CT in the EHR facilitated semantic interoperability, information sharing and data comparison across both systems and organizational borders.

11.1.3 Challenges and prerequisites for transitioning to ontology-based terminologies

The most striking challenges reported by the respondents were that support for ontology-based terminologies in EHR require the terminology to be seamlessly integrated into the users' workflows. This challenge is heavily related to one of the prerequisites that multiple respondents mentioned, namely the thorough training of users involved with the terminologies. This also includes giving users the ability to influence the user interface, as well as giving them the ability to use familiar terms to select data classified by the terminology.

Other challenges mentioned include the fact that there is a multitude of standards, and that cross-ontology transformation (mapping) still is quite inaccurate. The proportion of unstructured data and legacy systems that do not easily support ontologies also pose a serious challenge in a transition process.

The challenges roughly correspond to the prerequisites mentioned by the respondents:

- Governance around the use of these terminologies
- Training of users when they are involved in using these terminologies
- Effective management of change ensuring buy-in from users
- Technology to apply these in an interoperability layer
- Expertise to enhance and evolve these terminologies and related artefacts
- Budgetary and resourcing constraints — incremental funding to enable an evolutionary approach
- Understanding of the legacy environment

In terms of implementation, one of the vendors in the respondent (*r1*) group stated that they use an internal in-system ontology implementation. This was also true for one of the other vendors (*r2*) in the group, but only when ontological reasoning was used in their CDSS decision engines. While the first respondent (*r1*) stated that internal in-system support provided better performance than specialized terminology servers, the other vendor (*r2*) stated that a combination of plug-in systems and terminology servers allowed for both the implementation of proprietary ontologies and browsing and searching ontologies while authoring clinical logic.

All respondents stated that they already had a strong underlying framework for implementing ontology-based terminologies, and that future work was focused on collaboration with content vendors and customers. One also noted that expanding the relation/attribute types supported in their system was a priority. In addition, automated context detection and machine learning for improved accuracy for tagging were mentioned in the respondents' development roadmaps.

With regard to ontologies for molecular biology and genomics, one of the respondents stated that there is an immense gap between the databases of structured and machine-readable ontologies and clinical evidence, and that this creates one of the major bottlenecks for realizing seamless knowledge lifecycles related to the development and governance of clinical practice guidelines (CPGs). However, they also explained that

> *"(...) there are national and international coordination efforts such as Gene Ontology (GO), the Open Biomedical Ontologies (OBO) Foundry or the National Center for Biomedical Ontologies. This has an immense potential to bridge the gap and harness ontology-tagged patient data within EMRs for data mining and machine learning applications, as well as to provide patient-specific scientific information from the next-gen structural journal at point of care. Yet, in developing areas of research, the information that is critical is not known a-priori.*
>
> *As a result, the use of ontologies, knowledge bases, and inference engines applied to the informational models and unstructured data can enable researchers to query populations in EHRs in ways not possible before (pheno-type queries, for example)."*

11.1.4 Ontology-based terminologies and information models

All of the respondents stated that they used terminologies to annotate sections of clinical information models. While the specific strategies used vary between vendors, they all reported that clinical knowledge models could be annotated with references to one or more terminologies or ontologies, and that expressions in the clinical logic language could use defined ontological relationships between entities. Both of the vendors stated that the interaction between information models and ontologies was governed by mappings, either internally kept or found in external databases. One elaborated:

> *"Important factors that affect how the conceptual models in the terminologies and information models interact include:*
>
> - *Requirement to map existing information models to ontologies for historical data (e.g. legacy data);*
> - *Updates and governance around change processes to accommodate evolution over time;*
> - *The best ontological models are not always user friendly, hence there is a need to provide interfaces that hide this complexity from the users."*

With regard to information models as a necessity in the future, all agreed that a flexible information model is still required, among other things, for defining unstructured data that is not yet part of the ontology-based terminology. Information models are also more efficient for storage and querying, besides the fact that the model in an ontology-based terminology is rarely used as a structure for storing data.

12 Implications of the use of biomedical ontologies

This report has presented the different types of knowledge that exist by focusing on domain ontologies and describing their characteristics and their use for specifying the semantics of the EHR and the challenges in their adoption. Over the past decade, several patterns to leverage the use of clinical information standards with information models have been defined to enable the SIOp of clinical data relying on an agreement on a reference model, CIMs and a controlled set of concepts from a biomedical terminology (8).

However, the advent of precision medicine, programmes to enable the secondary use of clinical information, the need to take a holistic view of patient health, and so on are creating new challenges for clinical information management. Specifically, more accuracy and greater precision are needed to express certain clinical observations and to organize and query data in a more expressive way. This drives the need to integrate EHR information with disparate data coming from heterogeneous sources, such as biological, environmental or sensors sources, in order to provide the best information to decision makers (92, 93). Such integration requires the use of different types of ontologies to specify and manage knowledge from the biomedical domain. The challenge therefore extends not only to one biomedical ontology (e.g. SNOMED CT), but to combine many ontologies (such as the GO, PRO, etc.) that need to be appropriately combined with the EHR information models (CIMs) in a consistent manner. SNOMED CT has been proposed as a core reference terminology that should become the hub providing the general structure to integrate different terminologies and ontologies (74). In this regard, several countries are involved in the adoption of SNOMED CT as a reference terminology. However, the best way to approach the adoption of SNOMED CT is still a matter of investigation for many countries (66, 94). Also, under the framework of the Horizon 2020 research and innovation programme, the EU has funded projects to study the organizational aspects involved in its adoption. In particular, the EU project ASSESS-CT has evaluated organizational challenges with regard to the adoption of SNOMED CT (74). These initiatives have demonstrated that the introduction of an ontology-based reference terminology involves organizational challenges and the need to establish national and regional committees to support its adoption and use. In addition to organizational aspects, the adoption of ontology-based terminologies involves making major decisions on the health information architecture to use their expressivity and combine the use of several ontologies under the same technical infrastructure.

12.1 Summary of findings

This section compiles the findings reported across the document and the literature review about the adoption and implications of biomedical ontologies and classifies them as Strengths, Weaknesses, Opportunities and Threats (SWOT). Strengths and weaknesses relate to the benefits and challenges found in the internal factors of the ontologies and semantic Web domains. The understanding of internal factors are the aspects inherent to the technological features or the characteristics of the organizations that develop the ontologies. Opportunities and threats regards the external factors that influence the adoption of ontologies and Semantic Web technologies. External factors describes the social, economic, or political factors influencing the adoption of ontology-based terminologies.

INTERNAL	**STRENGHS** *Allow efficient knowledge management of complex domains. *Avoid ambiguity in the information facilitating cross-border SIOp. *Power secondary use of clinical data by allowing expressive queries over structured and unstructured datasets. *Improve accuracy in data analysis. *Simplify authoring process, facilitate governance, and facilitate rule sets maintenance of CDSS.	**WEAKNESSES** *Governance and use of biomedical ontologies is complex. *Technologies and software architectures to leverage several ontologies are not mature in HIS. *Lack of clear strategies on how to evolve legacy HIS into systems that make use of biomedical ontologies *There is a need for more usable ontology management tools. *Computational restrictions. *Lack of clear guidelines for the alignment and co-habitation of information and domain ontology models.
EXTERNAL	**OPORTUNITIES** *Norway already has the industry and public entities with know-how on clinical information standards and HIS infrastructures. *SNOMED CT is becoming the de-facto reference medical ontology that may act as a hub to accommodate mappings to different domain ontologies/terminologies. *Extensive support for the adoption of SNOMED CT is available. *Ongoing efforts to leverage the use of SNOMED CT with other biomedical ontologies. *Ongoing development of top-level ontologies and patterns to develop more robust semantic web architectures. *Most CDSS vendors already support Level 1 in the use of formal semantics and some can accommodate Level 2 using basic reasoning. *W3C standards support the gradual incorporation of ontologies on top of legacy systems (bottom-up approach). *EU and Scandinavian countries are running projects to evaluate the adoption of SNOMED CT which conclusions are extensible to Norway.	**THREATS** *Lack of technical and organizational experts to drive biomedical ontologies adoption and governance. *Lack of experiences implementing large-scale semantic architectures capable to use different ontologies within the EHR. *Lack of nation-wide projects and objective evaluations of semantic architectures in HIS. *Budgetary and resource constraints to implement large scale pilots. *Vast amount of terminologies and ontologies developed in parallel by different bodies (complex ontology matching). *Lack of clear commitment of some biomedical ontologies with W3C standards and recommendations.

Table 4: SWOT matrix on ontology-based terminologies

Weaknesses

The weaknesses cell refers to the internal factors of biomedical ontologies that may hamper their adoption. In the literature reviewed and the set of interviews and questionnaires performed with CDS vendors and researchers this report has identified several internal weaknesses.

These weaknesses refer both to the adoption and management of ontology-based terminologies, in particular; and semantic web architectures, in general. Regarding organizational aspects, the governance and management involved in the adoption of biomedical terminologies at large scale is still a matter of discussion and research (9, 74, 95). Guidelines to implement SNOMED CT are available, but there is a lack of clear strategies on how to evolve legacy HIS into systems that make use of different biomedical ontologies (not only SNOMED CT) in combination with heterogeneous information models. In the interviews performed, CDS experts considered that transiting to ontology-based models would enhance and evolve terminologies and related CDS artefacts. However, there is a need for explicit strategies on how to use them in combination with the information models of the EHR. In this regard, it was found that technologies and software architectures for implementing formal semantics for SIOp are still not mature in the healthcare arena. Investment in pilot projects are needed in order to gain experience on how to implement those systems. An explicit strategy and allocation of resources to gain knowledge and experiences on the development of semantic web infrastructures is also necessary. Otherwise budgetary and resource constraints may affect these studies since large scale pilots are needed in order to determine how to implement semantic web architectures capable of leveraging the use of different biomedical domain ontologies. Another technical issue is how to leverage different types of knowledge such as information models, domain ontologies and procedural CDS logic when applicable (96, 97). An added burden to the combination of different types of knowledge are computational restrictions that the use of DLs impose (25, 98). Also the interviews identified that technologies to ease the management of ontologies are needed. Finally, SNOMED CT has improved a lot in the recent years drawing a clear technical path for its adoption, but it is still required development of an organizational path and development of guidelines to deal with the organizational aspects involved in the adoption of diverse biomedical domain ontologies as discussed in the semantic web architecture section.

Strengths

Strengths refer to the internal factors that may have a positive impact on the adoption of ontology-based terminologies. Besides the weaknesses identified, there are also strengths in the arena of ontologies development and semantic web research for health information systems. First, the use of formal semantics enables the efficient knowledge management of complex domains. Examples are the maintenance of SNOMED CT. Second, the adoption of ontology-based terminologies brings opportunities to solve some of the challenges that ballast the seamless SIOp among HIS since the relationships among different concepts of clinical extracts can be specified at a formal level. Third, ontologies allow performing precision searches of information that improve the accuracy of data analysis. This type of intelligent information retrieval is not limited to structured information, but also to unstructured data. This is possible because biomedical ontologies also facilitate Natural Language Processing since the ontology acts as the reference concepts and relationships that must be identified in free-text. Once free-text documents are annotated with terminological expressions, they can be retrieved searching the standard concepts of the terminology. This is an advantage recognized by the CDS researchers and vendors interviewed. Finally, when it comes to CDS systems, the interviews performed unveiled that CDS researchers and vendors identified the management of rule sets, simplification of the management and interoperability using ontologies to define a common layer that unifies disparate vocabularies as other benefits.

Threats

The threats cell in Table 4 refers to the external factors that may hamper the adoption of ontology-based terminologies. The first identified threat is the lack of nation-wide projects and objective evaluations of implementing ontology-based terminologies. At the moment, the most complete report in the medical informatics domain are the results of the ASSESS-CT project (74), but as the stakeholders recognize it is still too soon to extract solid conclusions since there are few projects that have completed a full adoption of ontology-based terminologies, such as SNOMED CT, on a national scale. Another threat comes from the lack of experts in biomedical ontologies and semantic web technologies, both at an organizational and a technical level, to assess the adoption of ontologies. Another related problem is the lack of experiences implementing large semantic web architectures capable to use different ontologies into the EHR. Although some projects have informed about the challenges and implementations of such architectures in fields such as eGoverment or eScience (69, 70), the healthcare arena has been centred on the development of vocabularies. As a consequence, the architectures to fully integrate different types of ontologies are still immature. This lack of large developments leads to a high risk when designing EHR systems that leverage the use of several ontologies since a large amount of terminologies and ontologies have been developed in parallel by different bodies, thus making ontology mapping among them extremely complex (67).

Another aspect that may cause problems is the need for a clear commitment of ontology developments with W3C standards and semantic web standards. This is a problem also identified by other studies that report the need to align SNOMED CT developments with the semantic web standards (74).

Opportunities

The items contained in the opportunities cell of the SWOT matrix refer to external factors that may contribute to facilitate the adoption of biomedical ontologies for EHRs.

First, Norway already has vendors and public entities with know-how on implementing clinical information standards and eHealth infrastructures. This will very likely ease the transition to ontology-based terminologies of HIS. However, as mentioned in the threats section, it will be necessary to educate more eHealth professionals in the management and development of systems that make extensive use of biomedical ontologies.

Second, another positive factor that may contribute to ease the adoption of biomedical ontologies in HIS, is that SNOMED CT is becoming the de-facto reference medical ontology that can act as hub to accommodate different mappings to different domain ontologies/terminologies (74).

Third, SNOMED International already provides lot of support to its members in the form of guidelines, training, conferences to share experiences, and so on (34, 37, 99).

Fourth, with regards to the challenges in integrating different ontologies, there is a high interest to study the integration of SNOMED CT with ontologies from the OBO foundry (67, 100-103). If the interest is materialized into running research projects, the results of these will help to set the basis to start integrating different biomedical ontologies into EHR systems. With regards to the architecture that should provide the infrastructure to leverage the use of different ontologies, the top-level semantic web architecture is being developed as upper-level ontologies such as the BFO (67) and BioTop (26, 27).

From the vendors point of view, there are also positive factors that will help in the transition to ontology-based terminologies and ontology-based HIS. The interviews with CDS vendors unveiled that most of them already support Level 1 and Level 2 in the use of formal semantics (i.e. they make use of subsumptive reasoning over ontologies). This can provide the basis to start building semantic web architectures capable to use different ontologies and make use of their features. As a matter of fact, W3C standards already support the gradual incorporation of a semantic layer to legacy systems (bottom-up approach) (104, 105). This is also recognized by SNOMED CT experts that recommend adopting SNOMED CT in an incremental manner (74). This should allow to transit from legacy systems to ontology-based systems gradually as new needs for formal semantics appear. Therefore it is not needed to perform dramatic changes in eHealth infrastructures that would require massive investment and replacement of running systems.

Finally, one the most important external factors that can contribute to the adoption of ontology-based terminologies is the EU context. Several initiatives in the EU are increasing the knowledge on the organizational and technical challenges involved in the adoption of biomedical ontologies such as SNOMED CT (74, 95). In addition, EU research projects in the field of semantic web architectures have set the basic architectural principles for integrating ontologies with legacy information systems (104-106). The knowledge and experts that have raised from these initiatives will help to start building semantic web architectures in healthcare that incorporate different biomedical ontologies into EHRs, CDS systems etc. However, a clear strategy to allocate resources to realize these developments into running solutions in existing EHRs is needed since, to the authors´ knowledge, no large-scale pilots including several EHRs and making extensive use of different biomedical ontologies have been performed.

12.2 Acknowledgements

We thank Daniel Karlsson for the references and clarifications provided on SNOMED CT’s concept model.

13 Literature search

13.1 Objective

This review focuses on finding the main research papers, project deliverables and technical reports related with the features and applications of ontology-based medical terminologies in general, and SNOMED CT in particular.

13.2 Method

Conference and journal papers from Pubmed, IHTSDO SNOMED CT guidelines, EU project deliverables and technical reports were reviewed. Pubmed was searched by the keywords ("snomed-ct" and "concept model"), ("snomed-ct" and "hl7"), ("snomed-ct" and "openEHR"), ("snomed-ct" and "interoperability"), ("snomed-ct" and "terminology binding"). Guidelines from IHTSDO were selected if they were related to technical implementation, the relationship with information models or treated the ontological aspects of SNOMED CT. Technical reports, papers and project deliverables known by the authors were also included if they covered any aspect of the use of SNOMED CT as ontology to express medical knowledge. Only studies or documents from 2008 were included provided that in 200X the formalism used to express SNOMED CT concept model (Ontology) was changed to OWL EL and some changes on its X model were introduced.

Eligibility criteria were based on:

a) The study or document described the use of SNOMED CT to grant interoperability exploiting some ontological feature of the terminology.
b) The SNOMED CT concept model was described or analyzed.
c) SNOMED CT in combination clinical information models were used to express clinical information or procedures using the ontology.
d) The document described general aspects about ontologies in computer science helping to understand the implications on the use of SNOMED CT no only as coding system, but as ontology to formally express clinical meaning.

Firstly, titles and abstracts were screened ruling out irrelevant papers. Secondly, the selected material was reviewed in full text selecting those compliant with the eligibility criteria.

13.3 Result

As Figure 1 shows, the result of the key word search and document inclusion from other sources produced a total of 182 results combined. The first filter by title and abstract discarded 40 documents for not complying with eligibility criteria. This left 142 papers to be reviewed in full text. The full text review excluded 48 papers leaving 74 to be used as background literature for the report.

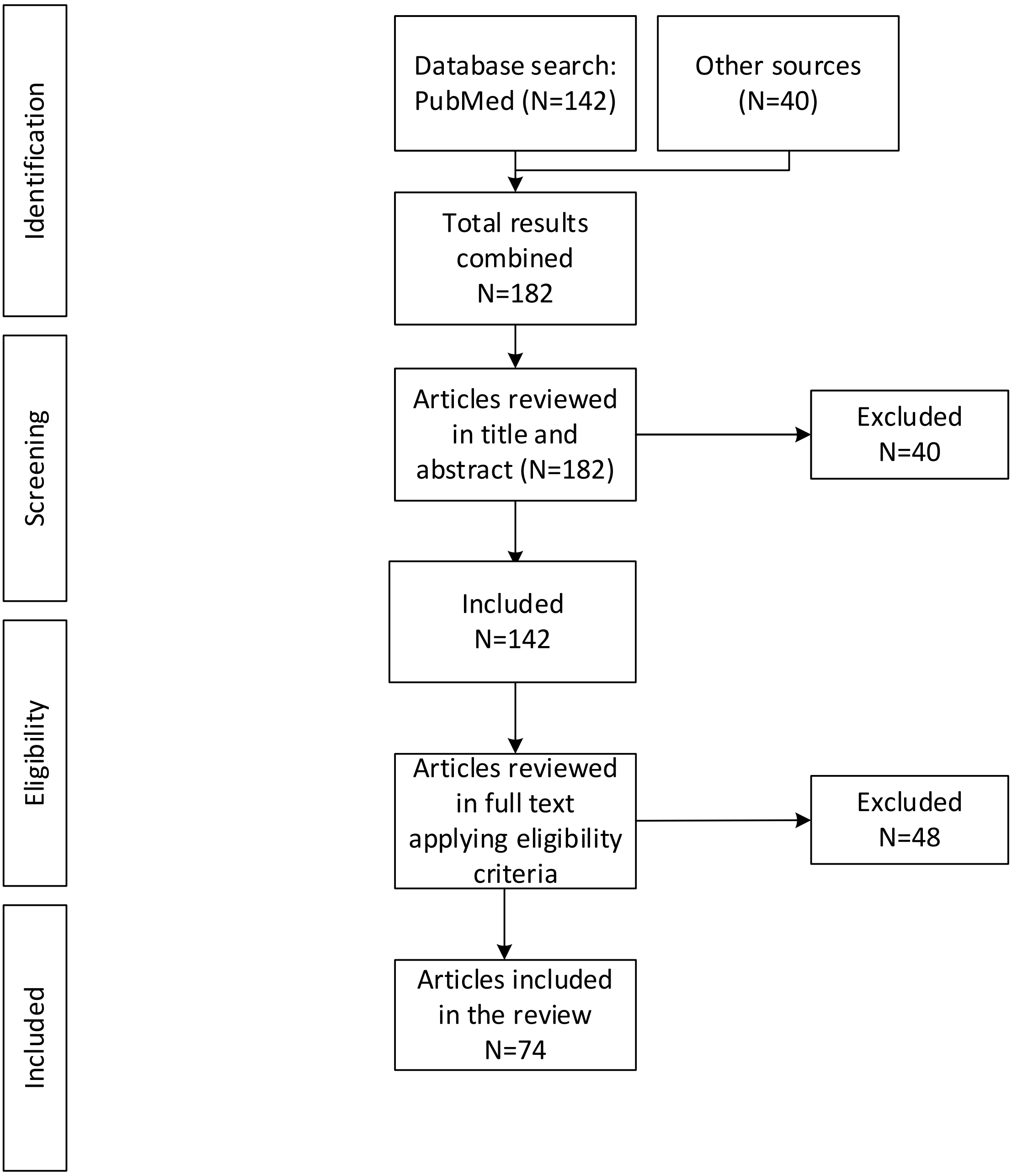


Figure 12: Workflow diagram of the review process.

Among the documents reviewed, 5 topics were identified with regards to ontologies:

- ontology design;
- theoretical background of ontologies and semantic web;
- ontology mapping;
- ontologies and information models;
- applications of ontologies.

Table 6 shows the percentage of documents that covered each particular topic. The inclusion of the papers is not exclusive. Therefore a paper may belong to two categories.

Topic	**Papers**	**Coverage**
Ontology design	[1–6,6–16]	18,9% (14/74)
Theoretical background of ontologies and Semantic Web	[12,17–22]	9.46% (7/74)
Ontology Mapping	[23–29]	9.46% (7/74)
Ontologies in combination with information models	[6,6,7,30–58]	41.89% (31/74)
Ontology applications	[6,7,10,59–74]	24.32% (18/74)

Table 5: Coverage of topics by the selected documents

13.4 References from search

[1] Beisswanger E, Schulz S, Stenzhorn H, Hahn U. BioTop: An Upper Domain Ontology for the Life Sciences: A Description of Its Current Structure, Contents and Interfaces to OBO Ontologies. Appl Ontol 2008;3:205–212.

[2] BioTop - A Top-Domain Ontology for the Life Sciences n.d. http://www2.imbi.uni-freiburg.de/ontology/biotop/ (accessed February 1, 2017).

[3] BODENREIDER O, SMITH B, BURGUN A. The Ontology-Epistemology Divide: A Case Study in Medical Terminology. Form Ontol Inf Syst 2004;2004:185–95.

[4] Fox J, Alabassi A, Patkar V, Rose T, Black E. An ontological approach to modelling tasks and goals. Computers in Biology and Medicine 2006;36:837–56. doi:10.1016/j.compbiomed.2005.04.011.

[5] Gordon CL, Pouch S, Cowell LG, Boland MR, Platt HL, Goldfain A, et al. Design and evaluation of a bacterial clinical infectious diseases ontology. AMIA Annu Symp Proc 2013;2013:502–11.

[6] Marco-Ruiz L, Pedrinaci C, Maldonado JA, Panziera L, Chen R, Bellika JG. Publication, discovery and interoperability of Clinical Decision Support Systems: A Linked Data approach. Journal of Biomedical Informatics 2016;62:243–64. doi:10.1016/j.jbi.2016.07.011.

[7] Martínez-Costa C, Cornet R, Karlsson D, Schulz S, Kalra D. Semantic enrichment of clinical models towards semantic interoperability. The heart failure summary use case. J Am Med Inform Assoc 2015;22:565–76. doi:10.1093/jamia/ocu013.

[8] Martínez-Costa C, Schulz S. Ontology-Based Reengineering of the SNOMED CT Context Model. Conference on Artificial Intelligence in Medicine in Europe, Springer Berlin Heidelberg; 2013, p. 216–220.

[9] Martínez-Costa C, Schulz S. Ontology-based reinterpretation of the SNOMED CT context model. ICBO, 2013, p. 90–95.

[10] Navas H, Lopez Osornio A, Gambarte L, Elias Leguizamon G, Wasserman S, Orrego N, et al. Implementing rules to improve the quality of concept post-coordination with SNOMED CT. Stud Health Technol Inform 2010;160:1045–9.

[11] Schulz S, Rodrigues JM, Rector AL, Spackman KA, Campbell JR, Üstün B, et al. What's in a class? Lessons learnt from the ICD-SNOMED CT harmonisation. MIE 2014:1038–1042.

[12] Studer R, Benjamins VR, Fensel D. Knowledge engineering: principles and methods. Data & Knowledge Engineering 1998;25:161–197.

[13] Uribe GA, Lopez DM, Blobel B. Architectural analysis of clinical ontologies for pHealth interoperability. Stud Health Technol Inform 2012;177:176–82.

[14] Why Do It the Hard Way? The Case for an Expressive Description Logic for SNOMED n.d. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2585532/ (accessed March 2, 2017).

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

[16] Ontological Engineering - with examples from the areas | Asunción Gómez-Pérez | Springer. n.d.

[17] Allemang D, Hendler J. Semantic Web for the Working Ontologist, Second Edition: Effective Modeling in RDFS and OWL. 2 edition. Waltham, MA: Morgan Kaufmann; 2011.

[18] Gruber TR. A translation approach to portable ontology specifications. Knowledge Acquisition 1993;5:199–220.

[19] Guariano N. IOS Press. Frontiers in Artificial Intelligence and Applications 1998;46:347.

[20] Lassila O, McGuinness D. The role of frame-based representation on the semantic web. Linköping Electronic Articles in Computer and Information Science 2001;6:2001.

[21] Rector A, Sottara D. Chapter 20 - Formal Representations and Semantic Web Technologies. In: Greenes RA, editor. Clinical Decision Support (Second Edition), Oxford: Academic Press; 2014, p. 551–98.

[22] Schreiber G, Akkermans H, Anjewierden A, Hoog RD, Shadbolt NR, Velde WV de, et al. Knowledge Engineering and Management: The CommonKADS Methodology. Cambridge, Mass: The MIT Press; 1999.

[23] Campbell JR, Brear H, Scichilone R, White S, Giannangelo K, Carlsen B, et al. Semantic interoperation and electronic health records: context sensitive mapping from SNOMED CT to ICD-10. Stud Health Technol Inform 2013;192:603–7.

[24] Dhombres F, Bodenreider O. Interoperability between phenotypes in research and healthcare terminologies--Investigating partial mappings between HPO and SNOMED CT. J Biomed Semantics 2016;7:3. doi:10.1186/s13326-016-0047-3.

[25] Etienne F. ICD-11 (JLMMS) and SCT Inter-Operation. Health Informatics Meets EHealth: Predictive Modeling in Healthcare–From Prediction to Prevention. Proceedings of the 10th EHealth2016 Conference, vol. 223, IOS Press; 2016, p. 267.

[26] Rodrigues J-M, Robinson D, Della Mea V, Campbell J, Rector A, Schulz S, et al. Semantic Alignment between ICD-11 and SNOMED CT. Stud Health Technol Inform 2015;216:790–4.

[27] Rodrigues J-M, Schulz S, Rector A, Spackman K, Millar J, Campbell J, et al. ICD-11 and SNOMED CT Common Ontology: circulatory system. Stud Health Technol Inform 2014;205:1043–7.

[28] Rodrigues JM, Schulz S, Rector A, Spackman K, Üstün B, Chute CG, et al. Sharing ontology between ICD 11 and SNOMED CT will enable seamless re-use and semantic interoperability. Studies in Health Technology and Informatics 2012;192:343–346.

[29] Rodrigues JM, Schulz S, Rector AL, Spackman KA, Üstün B, Chute CG, et al. Do we need a Common Ontology between ICD 11 and SNOMED CT to ensure Seamless Re-Use and Semantic Interoperability? MedInfo, 2013, p. 1234.

[30] Andrade AQ, Almeida MB, Schulz S. Revisiting ontological foundations of the OpenEHR Entry Model. 2012.

[31] Bird L, Brooks C, Cheong YC, Tun NN. A logical approach to semantic interoperability in healthcare. Stud Health Technol Inform 2011;168:1–9.

[32] Blumenthal D, Tavenner M. The “Meaningful Use” Regulation for Electronic Health Records. New England Journal of Medicine 2010;363:501–4. doi:10.1056/NEJMp1006114.

[33] Campbell JR, Campbell WS, Hickman H, Pedersen J, McClay J. Employing complex polyhierarchical ontologies and promoting interoperability of i2b2 data systems. AMIA Annu Symp Proc 2015;2015:359–65.

[34] Declerck G, Hussain S, Daniel C, Yuksel M, Laleci GB, Twagirumukiza M, et al. Bridging data models and terminologies to support adverse drug event reporting using EHR data. Methods Inf Med 2015;54:24–31. doi:10.3414/ME13-02-0025.

[35] Dixon BE, Simonaitis L, Goldberg HS, Paterno MD, Schaeffer M, Hongsermeier T, et al. A pilot study of distributed knowledge management and clinical decision support in the cloud. Artif Intell Med 2013;59. doi:10.1016/j.artmed.2013.03.004.

[36] Heymans S, McKennirey M, Phillips J. Semantic validation of the use of SNOMED CT in HL7 clinical documents. J Biomed Semantics 2011;2:2. doi:10.1186/2041-1480-2-2.

[37] Hwang KH, Chung K-I, Chung M-A, Choi D. Review of semantically interoperable electronic health records for ubiquitous healthcare. Healthc Inform Res 2010;16:1–5. doi:10.4258/hir.2010.16.1.1.

[38] Karlsson D, Berzell M, Schulz S. Information Models and Ontologies for Representing the Electronic Health Record. ICBO, 2011.

[39] Lozano-Rubi R, Munoz Carrero A, Serrano Balazote P, Pastor X. OntoCR: A CEN/ISO-13606 clinical repository based on ontologies. J Biomed Inform 2016;60:224–33. doi:10.1016/j.jbi.2016.02.007.

[40] Marco Ruiz L, Maldonado JA, Karlsen R, Bellika JG. Multidisciplinary Modelling of Symptoms and Signs with Archetypes and SNOMEDCT for Clinical Decision Support. Stud Health Technol Inform., Madrid: IOS press; 2015.

[41] Markwell D, Sato L, Cheetham E. Representing Clinical Information using SNOMED Clinical Terms with Different Structural Information Models. Proceedings of the 3rd International Conference on Knowledge Representation in Medicine, vol. 2008, 2008.

[42] Martinez-Costa C, Kalra D, Schulz S. Improving EHR semantic interoperability: future vision and challenges. Stud Health Technol Inform 2014;205:589–93.

[43] Martinez-Costa C, Legaz-García MC, Schulz S, Fernández-Breis JT. Ontology-based infrastructure for a meaningful EHR representation and use. Biomedical and Health Informatics (BHI), 2014 IEEE-EMBS International Conference on, IEEE; 2014, p. 535–538.

[44] Martínez-Costa C, Schulz S. Ontology content patterns as bridge for the semantic representation of clinical information. Applied Clinical Informatics 2014;5:660.

[45] Martinez-Salvador B, Marcos M, Manas A, Maldonado JA, Robles M. Using SNOMED CT Expression Constraints to Bridge the Gap Between Clinical Decision-Support Systems and Electronic Health Records. Stud Health Technol Inform 2016;228:504–8.

[46] Meizoso Garcia M, Iglesias Allones JL, Martinez Hernandez D, Taboada Iglesias MJ. Semantic similarity-based alignment between clinical archetypes and SNOMED CT: an application to observations. Int J Med Inform 2012;81:566–78. doi:10.1016/j.ijmedinf.2012.02.007.

[47] Moreno-Conde A, Moner D, Cruz WD da, Santos MR, Maldonado JA, Robles M, et al. Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis. J Am Med Inform Assoc 2015;22:925–34. doi:10.1093/jamia/ocv008.

[48] Randorff Hojen A, Rosenbeck Goeg K. Snomed CT implementation. Mapping guidelines facilitating reuse of data. Methods Inf Med 2012;51:529–38. doi:10.3414/ME11-02-0023.

[49] Rector AL, Qamar R, Marley T. Binding ontologies and coding systems to electronic health records and messages. Applied Ontology 2009;4:51–69. doi:10.3233/AO-2009-0063.

[50] Rector AL. The Interface between Information, Terminology, and Inference Models. vol. 84, London, UK: IOS press; 2001, p. 246–50. doi:10.3233/978-1-60750-928-8-246.

[51] Rector AL, Johnson PD, Tu S, Wroe C, Rogers J. Interface of Inference Models with Concept and Medical Record Models. Proc Artificial Intelligence in Medicine Europe (AIME-2001) 2001:314–23.

[52] Rodrigues JM, Kumar A, Bousquet C, Trombert B. Using the CEN/ISO standard for categorial structure to harmonise the development of WHO international terminologies. Stud Health Technol Inform 2009;150:255–9.

[53] Schulz S, Schober D, Daniel C, Jaulent M-C. Bridging the semantics gap between terminologies, ontologies, and information models. Stud Health Technol Inform 2010;160:1000–4.

[54] Stroetmann V, Kalra D, Lewalle P, Rector A, Rodrigues JM, Stroetmann KA, et al. European Commission, Semantic Interoperability for Better Health and Safer Healthcare, Deployment and Research Roadmap for Europe, 2009 2009.

[55] Semantic Interoperability for Health Network Deliverable 4.1 : Initial models covering the heart failure use case n.d.

[56] Semantic Interoperability for Health Network Deliverable 4. 3 : Ontology / Information models covering the public health use case s n.d.

[57] Semantic Interoperability for Health Network Deliverable 4.4: Report on interface specifications between semantic artefacts n.d.

[58] Semantic Interoperability for Health Network Deliverable 4.5: Design Guides and Roadmap n.d.

[59] Bodenreider O. Biomedical ontologies in action: role in knowledge management, data integration and decision support. Yearb Med Inform 2008:67–79.

[60] Campbell WS, Pedersen J, McClay JC, Rao P, Bastola D, Campbell JR. An alternative database approach for management of SNOMED CT and improved patient data queries. J Biomed Inform 2015;57:350–7. doi:10.1016/j.jbi.2015.08.016.

[61] Cornet R, de Keizer NF. Recording associated disorders using SNOMED CT. Stud Health Technol Inform 2011;169:824–8.

[62] Gimenez-Solano VM, Maldonado JA, Salas-Garcia S, Bosca D, Robles M. Implementation of an Execution Engine for SNOMED CT Expression Constraint Language. Stud Health Technol Inform 2016;228:466–70.

[63] Hojen AR, Bronnum D, Goeg KR, Elberg PB. Applying the SNOMED CT Concept Model to Represent Value Sets for Head and Neck Cancer Documentation. Stud Health Technol Inform 2016;228:436–40.

[64] Kalra D, Stroetmann V, Rosenbeck K, Randorff A, Dewenter H, Nejad RF, et al. Assessing SNOMED CT for large scale eHealth deployments in the EU 2016.

[65] Karlsson D, Nyström M, Cornet R. Does SNOMED CT post-coordination scale? Stud Health Technol Inform 2014;205:1048–52.

[66] Pathak J, Solbrig HR, Buntrock JD, Johnson TM, Chute CG. LexGrid: A Framework for Representing, Storing, and Querying Biomedical Terminologies from Simple to Sublime. Journal of the American Medical Informatics Association 2009;16:305–15. doi:10.1197/jamia.M3006.

[67] Smith B, Brochhausen M. Putting biomedical ontologies to work. Methods Inf Med 2010;49:135–40. doi:10.3414/ME9302.

[68] Soguero-Ruiz C, Lechuga-Suárez L, Mora-Jiménez I, Ramos-López J, Barquero-Pérez Ó, García-Alberola A, et al. Ontology for heart rate turbulence domain from the conceptual model of SNOMED-CT. IEEE Trans Biomed Eng 2013;60:1825–33. doi:10.1109/TBME.2013.2243147.

[69] Song T-M, Park H-A, Jin D-L. Development of health information search engine based on metadata and ontology. Healthc Inform Res 2014;20:88–98. doi:10.4258/hir.2014.20.2.88.

[70] IHTSDO. SNOMED CT Compositional Grammar Specification and Guide n.d.

[71] Data Analytics with SNOMED CT - Data Analytics with SNOMED CT - SNOMED Confluence n.d. https://confluence.ihtsdotools.org/display/DOCANLYT (accessed March 3, 2017).

[72] SNOMED CT Editorial Guide - SNOMED CT Editorial Guide - SNOMED Confluence n.d. https://confluence.ihtsdotools.org/display/DOCEG (accessed March 3, 2017).

[73] SNOMED CT Starter Guide - SNOMED CT Starter Guide - SNOMED Confluence n.d. https://confluence.ihtsdotools.org/display/DOCSTART (accessed March 3, 2017).

[74] Technical Implementation Guide - Technical Implementation Guide - SNOMED Confluence n.d. https://confluence.ihtsdotools.org/display/DOCTIG (accessed March 3, 2017).

14 References

1. Direktoratet for e-helse. Nasjonal e-helsestrategi og mål 2017-2022. 2017.
2. Meld. St. 9 (2012-2013) Én innbygger - én journal. In: omsorgsdepartementet H-o, editor. 2012.
3. omsorgsdepartementet H-o. Tildelingsbrev for e-helse for 2017. 2017.
4. Greenhalgh T, Howick J, Maskrey N. Evidence based medicine: a movement in crisis? Bmj. 2014;348:g3725.
5. Djulbegovic B, Guyatt GH, Ashcroft RE. Epistemologic inquiries in evidence-based medicine. Cancer control: journal of the Moffitt Cancer Center. 2009;16(2):158-68.
6. Olsen L, Aisner D, McGinnis JM. Institute of Medicine (US). Roundtable on Evidence-Based Medicine. The learning healthcare system: workshop summary. Washington, DC: National Academies Press; 2007.
7. Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. Science translational medicine. 2010;2(57):57cm29-57cm29.
8. Stroetmann VN, Kalra D, Lewalle P, Rector A, Rodriques JM, Stroetmann KA. Semantic Interoperability for better health and safer healthcare. European Commission, Directorate - General Information Society and Media. 2009.
9. Martínez-Costa C, Cornet R, Karlsson D, Schulz S, Kalra D. Semantic enrichment of clinical models towards semantic interoperability. The heart failure summary use case. Journal of the American Medical Informatics Association. 2015;22(3):565-76.
10. Moreno-Conde A, Moner D, Da Cruz WD, Santos MR, Maldonado JA, Robles M, et al. Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis. Journal of the American Medical Informatics Association. 2015:ocv008.
11. Nasjonal IKT. Tiltak 15.5 Folkeregistret i helsenettet 2015 [Available from: http://nasjonalikt.no/no/satsingsomrader/2_struktur_-_systemarkitektur_informasjonsgrunnlag_og_sikkerhet/tiltak_155_folkeregisteret_i_helsenettet/Tiltak+15.5+Folkeregisteret+i+helsenettet.9UFRjK5S.ips
12. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med. 2010;2010(363):501-4.
13. HL7/ASTM Implementation Guide for CDA® R2 -Continuity of Care Document (CCD®) Release 1. 2017 [Available from: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=6.
14. Lærum H, Bakke SL, Pedersen R, Valand JT. An update on OpenEHR archetypes in Norway: Response to article Christensen B & Ellingsen G:" Evaluating Model-Driven Development for large-scale EHRs through the openEHR approach" IJMI May 2016, Volume 89, pages 43-54. International journal of medical informatics. 2016;93:1.
15. Ministerio de Sanidad SSeI. Profesionales - Historia Clínica Digital del Sistema Nacional de Salud n.d. [Available from: http://www.msssi.gob.es/profesionales/hcdsns/home.htm.
16. Guariano N. Frontiers in Artificial Intelligence and Application. IOS Press. 1998;46(347).
17. Schreiber G, Akkermans H, Anjewierden A, Hoog RD, Shadbolt NR, WV V. Knowledge Engineering and Management: The CommonKADS Methodology. The MIT Press. 1999.
18. Rector A, Sottara D. Formal Representations and Semantic Web Technologies. Clinical Decision Support. Second edition ed: Oxford: Academic Press; 2014. p. 551-98.
19. Gomez-Perez A, Fernández-López M, Corcho O. Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web: Springer Science & Business Media; 2006.
20. Gruber TR. A translation approach to portable ontology specifications. Knowledge acquisition. 1993;5(2):199-220.

21. Studer R, Benjamins VR, Fensel D. Knowledge engineering: principles and methods. Data & knowledge engineering. 1998;25(1-2):161-97.
22. John F. Sowa. KR Ontology [Available from: http://www.jfsowa.com/ontology/kronto.htm.
23. Smith B, Kumar A, Bittner T. Basic formal ontology for bioinformatics. Journal of Information Systems. 2005:1-16.
24. Fox J, Alabassi A, Patkar V, Rose T, Black E. An ontological approach to modelling tasks and goals. Computers in biology and medicine. 2006;36(7):837-56.
25. Marco-Ruiz L, Pedrinaci C, Maldonado J, Panziera L, Chen R, Bellika JG. Publication, discovery and interoperability of Clinical Decision Support Systems: A Linked Data approach. Journal of Biomedical Informatics. 2016;62:243-64.
26. Beisswanger E, Schulz S, Stenzhorn H, Hahn U. BioTop: An upper domain ontology for the life sciences. Applied Ontology. 2008;3(4):205-12.
27. Schulz S, Stenzhorn H, Boeker M, Beisswanger E, Hahn U, van den Hoek L, et al. BioTop-A Top-Domain Ontology for the Life Sciences. Jan-2012[Online] Available: http://www imbi uni-freiburg de/ontology/biotop/[Accessed: 25-Oct-2012]. 2012.
28. WONCA International Classification Cometee. ICPC-2: International Classification of Primary Care: New York: Oxford Medical Publications; 1998 [Second edition:[
29. Logical Observation Identifiers Names and Codes (LOINC) [Available from: https://loinc.org.
30. Lassila O, McGuinness D. The role of frame-based representation on the semantic web. Linköping Electronic Articles in Computer and Information Science. 2001;6(5):2001.
31. Dixon BE, Simonaitis L, Goldberg HS, Paterno MD, Schaeffer M, Hongsermeier T, et al. A pilot study of distributed knowledge management and clinical decision support in the cloud. Artificial intelligence in medicine. 2013;59(1):45-53.
32. Berlin A, Sorani M, Sim I. A taxonomic description of computer-based clinical decision support systems. Journal of biomedical informatics. 2006;39(6):656-67.
33. Brachman RJ. On the epistemological status of semantic networks. Associative networks: Representation and use of knowledge by computers. 1979:3-50.
34. SNOMED CT Concept Model - SNOMED CT Starter Guide - SNOMED Confluence [Available from: https://confluence.ihtsdotools.org/display/DOCSTART/6.+SNOMED+CT+Concept+Model
35. Benson T, Grieve G. Principles of health interoperability: SNOMED CT, HL7 and FHIR: Springer; 2016.
36. IHTSDO. SNOMED CT Compositional Grammar Specification and Guide 2015 [Available from: http://doc.ihtsdo.org/download/doc_CompositionalGrammarSpecificationAndGuide_Current-en-US_INT_20150522.pdf/.pdf.
37. SNOMED CT Editorial Guide [Available from: https://confluence.ihtsdotools.org/display/DOCEG.
38. Baader F. The description logic handbook: Theory, implementation and applications: Cambridge university press; 2003.
39. Brachman RJ, Levesque HJ, editors. The tractability of subsumption in frame-based description languages. AAAI; 1984.
40. W3C. Web Ontology Language Document Overview [Second edition:[Available from: https://www.w3.org/2007/OWL/wiki/Tractable_Fragments#Classes_and_Object_Properties.
41. Baader F, Brandt S, Lutz C. Pushing the EL envelope further. 2008.
42. SNOMED CT Expression Constraint Language - SNOMED CT Languages Project Group - SNOMED Confluence n.d [Available from: https://confluence.ihtsdotools.org/display/SLPG/SNOMED+CT+Expression+Constraint+Language.
43. Hitzler P, Krotzsch M, Rudolph S. Foundations of semantic web technologies: CRC Press; 2009.
44. W3C. Tractable Fragments [Available from: https://www.w3.org/2007/OWL/wiki/Tractable_Fragments#Classes_and_Object_Properties
45. Martínez-Costa C, Schulz S, editors. Ontology-based reinterpretation of the SNOMED CT context model. ICBO; 2013.

46. Guarino N. Formal ontology, conceptual analysis and knowledge representation. International journal of human-computer studies. 1995;43(5-6):625-40.
47. Nutter J. Epistemology. Encyclopedia of Artificial Intelligence. 1990;1:280-7.
48. Bodenreider O, Smith B, Burgun A, editors. The ontology-epistemology divide: A case study in medical terminology. Formal ontology in information systems: proceedings of the International Conference FOIS (Conference); 2004: NIH Public Access.
49. Goossen W, Goossen-Baremans A, Van Der Zel M. Detailed clinical models: a review. Healthcare informatics research. 2010;16(4):201-14.
50. HL7 TermInfo Project n.d. http://wiki.hl7.org/index.php?title=TermInfo_Project# (accessed July 22, 2017).
51. Markwell D, Sato L, Cheetham E, editors. Representing clinical information using SNOMED Clinical Terms with different structural information models. KR-MED; 2008.
52. SNOMED CT Query Language - SNOMED CT Languages Project Group - SNOMED Confluence n.d. [Available from: https://confluence.ihtsdotools.org/display/SLPG/SNOMED+CT+Query+Language
53. SNOMED CT to ICD-10-CM Map n.d. [Available from: https://www.nlm.nih.gov/research/umls/mapping_projects/snomedct_to_icd10cm.html
54. Technical Implementation Guide - Technical Implementation Guide - SNOMED Confluence [Available from: https://confluence.ihtsdotools.org/display/DOCTIG.
55. Confluence S. Search and Data Entry Guide - Search and Data Entry Guide [Available from: https://confluence.ihtsdotools.org/display/DOCSEARCH/Search+and+Data+Entry+Guide.
56. Karlsson D, Berzell M, Schulz S, editors. Information Models and Ontologies for Representing the Electronic Health Record. ICBO; 2011.
57. Martínez-Costa C, Schulz S, editors. Ontology-Based Reengineering of the SNOMED CT Context Model. Conference on Artificial Intelligence in Medicine in Europe; 2013: Springer.
58. Beale T, Heard S, editors. An ontology-based model of clinical information. Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems; 2007: IOS Press.
59. CIMI. Clinical Models and SNOMED Kaiser Perspective [Available from: http://informatics.mayo.edu/CIMI/index.php/Clinical_Models_and_SNOMED_Kaiser_Perspective.
60. OpenEHR Implementers mailing list. January 2017. Available at:

http://lists.openehr.org/pipermail/openehr-implementers_lists.openehr.org/2017-January/001639.html.
61. Rector AL, Qamar R, Marley T. Binding ontologies and coding systems to electronic health records and messages. Applied Ontology. 2009;4(1):51-69.
62. Marco-Ruiz L, Pedersen R. Challenges in archetypes terminology binding using SNOMED-CT compositional grammar: the Norwegian patient summary case. Proc. 16th World Congr. Med. Health Inform. MedInfo2017, vol. in press, Hangzhou: IOS Press; 2017, p. in press.
63. Soguero-Ruiz C, Lechuga-Suarez L, Mora-Jiménez I, Ramos-Lopez J, Barquero-Perez O, Garcia-Alberola A, et al. Ontology for heart rate turbulence domain from the conceptual model of SNOMED-CT. IEEE Transactions on Biomedical Engineering. 2013;60(7):1825-33.
64. Martínez-Costa C, Menárguez-Tortosa M, Fernández-Breis JT, Maldonado JA. A model-driven approach for representing clinical archetypes for Semantic Web environments. Journal of biomedical informatics. 2009;42(1):150-64.
65. Rdf - FHIR v3.0.1 n.d. https://www.hl7.org/fhir/rdf.html (accessed August 2, 2017).
66. Lee D, de Keizer N, Lau F, Cornet R. Literature review of SNOMED CT use. Journal of the American Medical Informatics Association. 2013;21(e1):e11-e9.
67. Arp R, Smith B, Spear AD. Building ontologies with basic formal ontology: Mit Press; 2015.
68. Marco-Ruiz L, Pedrinaci C, Maldonado JA, Panziera L, Chen R, Bellika JG. Publication, discovery and interoperability of Clinical Decision Support Systems: A Linked Data approach. Journal of biomedical informatics. 2016;62:243-64.

69. Domingue J, Fensel D, Hendler JA. Handbook of semantic web technologies: Springer Science & Business Media; 2011.
70. Fensel PD, Facca DFM, Simperl DE, Toma I. Semantic Web. Semantic Web Serv., Springer Berlin Heidelberg; 2011, p. 87–104. doi:10.1007/978-3-642-19193-0_6.
71. Linked Data. 1 edition. Shelter Island, NY: Manning Publications; 2014.
72. Pedrinaci C, Domingue J, Sheth AP. Semantic web services. Handbook of semantic web technologies: Springer; 2011. p. 977-1035.
73. Shvaiko P, Euzenat J. Ontology matching: state of the art and future challenges. IEEE Transactions on knowledge and data engineering. 2013;25(1):158-76.
74. Assessing SNOMED CT for Large Scare eHealth Deployments in the EU - Assess CT Recommendations. 2016.
75. Ferguson J. Terminology Management With SNOMED CT At Kaiser Permanente 2016.
76. Panzer M. Increasing Patient Findability of Medical Research: Annotating Clinical Trials Using Standard Vocabularies. Bulletin of the Association for Information Science and Technology. 2017;43(2):40-3.
77. Campbell WS, Pedersen J, McClay JC, Rao P, Bastola D, Campbell JR. An alternative database approach for management of SNOMED CT and improved patient data queries. Journal of biomedical informatics. 2015;57:350-7.
78. RDF for Semantic Interoperability - HL7Wiki n.d. http://wiki.hl7.org/index.php?title=RDF_for_Semantic_Interoperability (accessed September 14, 2017).
79. Mission and Goals | www.opencimi.org n.d. http://www.opencimi.org/ (accessed October 5, 2015).
80. Yosemite Project | Comparing the Yosemite Project and ONC Roadmaps for Healthcare Information Interoperability n.d. http://yosemiteproject.org/recorded-webinars/2015-2/comparing-the-yosemite-project/ (accessed September 14, 2017).
81. OBO Principle: Relation: OBO Foundry; [Available from: http://obofoundry.org/principles/fp-007-relations.html.
82. The OBO Foundry 2017 [Available from: http://obofoundry.org/.
83. UBITECH. EHR Interoperability [Available from: https://www.ubitech.eu/technology/ehr-interoperability/.
84. Smith B. Tutorial: Introduction to Basic Formal Ontology (BFO 2.0). 2015.
85. OBO Principle: Overview: OBO Foundry; [Available from: http://obofoundry.org/principles/fp-000-summary.html.
86. Smith B, Brochhausen M. Putting biomedical ontologies to work. Methods of information in medicine. 2010;49(2):135.
87. Marquet G, Mosser J, Burgun A. A method exploiting syntactic patterns and the UMLS semantics for aligning biomedical ontologies: the case of OBO disease ontologies. international journal of medical informatics. 2007;76:S353-S61.
88. Rubin DL, Lewis SE, Mungall CJ, Misra S, Westerfield M, Ashburner M, et al. National center for biomedical ontology: advancing biomedicine through structured organization of scientific knowledge. Omics: a journal of integrative biology. 2006;10(2):185-98.
89. Kamdar MR, Tudorache T, Musen MA. A systematic analysis of term reuse and term overlap across biomedical ontologies. Semantic Web. 2016(Preprint):1-19.
90. Ghazvinian A, Noy NF, Musen MA. How orthogonal are the OBO Foundry ontologies? Journal of biomedical semantics. 2011;2(2):S2.
91. Ruttenberg A. Basic Formal Ontology 2017 [Available from: http://basic-formal-ontology.org/.
92. Sheth A. Internet of things to smart iot through semantic, cognitive, and perceptual computing. IEEE Intelligent Systems. 2016;31(2):108-12.
93. Sheth A, Anantharam P, Henson C. Semantic, cognitive, and perceptual computing: Paradigms that shape human experience. Computer. 2016;49(3):64-72.

94. HØJEN1a AR, Elberga PB, Andersena SK. SNOMED CT adoption in Denmark-why is it so hard? EHealth-For Continuity of Care: Proceedings of MIE2014. 2014;205:226.
95. Semantic Interoperability for Health Network Deliverable 4.5: Design Guides and Roadmap n.d.
96. Huff SM, Oniki TA, Coyle JF, Parker CG, Rocha RA. Chapter 17 - Ontologies, Vocabularies and Data Models. In: Greenes RA, editor. Clin. Decis. Support Second Ed., Oxford: Academic Press; 2014, p. 465–98.
97. Rector AL, Johnson PD, Tu S, Wroe C, Rogers J. Interface of Inference Models with Concept and Medical Record Models. Proc Artif Intell Med Eur AIME-2001 2001:314–23.
98. Karlsson D, Nyström M, Cornet R. Does SNOMED CT post-coordination scale? Stud Health Technol Inform 2014;205:1048–52.
99. Technical Implementation Guide - Technical Implementation Guide - SNOMED Confluence n.d. https://confluence.ihtsdotools.org/display/DOCTIG (accessed March 3, 2017).
100. Mailing List: obo-discuss 2017 [cited 2017 20.07]. Mailing List:[Available from: https://sourceforge.net/projects/obo/lists/obo-discuss.
101. Winnenburg R, Bodenreider O. Coverage of phenotypes in standard terminologies. Joint Bio-Ontologies and BioLINK ISMB. 2014:41-4.
102. Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, et al. The Human Phenotype Ontology in 2017. Nucleic Acids Research. 2016:gkw1039.
103. Ontology Xref Service 2017 [Available from: https://www.ebi.ac.uk/spot/oxo/index.
104. Semantic Annotations for WSDL and XML Schema n.d. http://www.w3.org/TR/sawsdl/ (accessed November 22, 2015).
105. Kopecký J, Gomadam K, Vitvar T. hRESTS: An HTML Microformat for Describing RESTful Web Services. IEEEWICACM Int. Conf. Web Intell. Intell. Agent Technol. 2008 WI-IAT 08, vol. 1, 2008, p. 619–25. doi:10.1109/WIIAT.2008.379.
106. Vitvar T, Kopecký J, Viskova J, Fensel D. WSMO-Lite Annotations for Web Services. In: Bechhofer S, Hauswirth M, Hoffmann J, Koubarakis M, editors. Semantic Web Res. Appl., Springer Berlin Heidel-berg; 2008, p. 674–89. doi:10.1007/978-3-540-68234-9_49.

[32] SOA4All | Knowledge Media Institute | The Open University n.d. http://projects.kmi.open.ac.uk/soa4all/index.html (accessed October 12, 2015).

15 Appendix

This chapter contains the following attachments:

- **Abstract for project presented at the research conference for Helseplattformen in Trondheim, 2017.**

  The project proposal was accepted for presentation as a poster.

- **Questionnaire used in this project.**

  The questionnaire was initially twofold; one part containing questions regarding ontology-based terminologies and information models, and one part regarding CDS architecture and standards. Only the parts relevant for this report is presented here, the other part can be found in the report "*Klinisk beslutningsstøtte – vurdering av standard og arkitektur*".

15.1 Abstract for research conference, Helseplattformen

Mulighetsvindu for bruk av ontologibaserte terminologier i EIEJ og Helseplattformen

Ontologibasert terminologi og standardisert informasjonsmodell for semantisk interoperabilitet og beslutningsstøtte i Helseplattformen

Nasjonalt senter for e-helseforskning

Forfattere: Kristian Nicolaisen[1], Rune Pedersen[1], Johan Gustav Bellika[1], Luis Marco Ruiz[1] og Per Atle Bakkevoll[1]

[1]Nasjonalt senter for e-helseforskning, Tromsø

Introduksjon

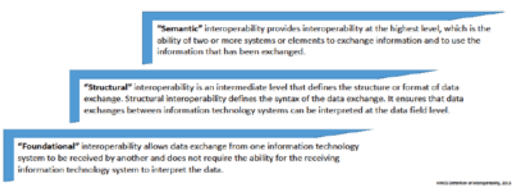
Forskningsprosjektet fokuserer på praktisk anvendelse av ontologibaserte terminologier, informasjonsmodeller og beslutningsstøtte – hver for seg, men også hvordan de fungerer sammen.

Semantisk interoperabilitet og ontologibasert terminologi

En forutsetning for semantisk interoperabilitet er bruk av ontologibaserte terminologier og informasjonsmodeller, og regionene har gjennom en årrekke arbeidet med standardisering og harmonisering av applikasjoner og innhold. I tillegg arbeides det mot en ny generasjon EPJ-systemer der strukturerte data, prosess- og beslutningsstøtte, samt økt sekundærbruk fremmes i tråd med hovedmålene i «Én innbygger – én journal».

Helse Midt-Norge oppgir i konkurransegrunnlaget at leverandøren(e)s erfaring med terminologier vil være viktig for å etablere et godt kunnskapsgrunnlag, der målet er en gradvis overgang fra ustrukturert fritekst til dokumentasjon med standardiserte terminologier. Norge er nå medlem av IHTSDO, og skal pilotere SNOMED-CT i nær framtid. Følgeforskning på utviklingen ved pilotering av SNOMED-CT vil være relevant for Helseplattformen.

Semantisk interoperabilitet



Beslutningsstøtte

Klinisk beslutningsstøttesystemer (Clinical Decision Support Systems – CDSS) er applikasjoner som bruker informasjon og data fra ulike kilder til å bistå helsepersonell i å ta kliniske beslutninger. Beslutningsstøtteverktøy er typisk designet for å integrere en medisinsk kunnskapsdatabase og spesifikk pasient- og behandlingsinformasjon med en regelmotor, for å generere case-spesifikke anbefalinger og råd til klinikere. Beslutningsstøttesystemets evne til å anvende informasjon til å gi spesifikke anbefalinger avhenger av representasjonen og formaliseringen av pasientinformasjonen og kunnskapsgrunnlaget det baserer beslutninger på. Desto mer veldefinert denne informasjonen er, desto mer pålitelig og funksjonsrik blir beslutningsstøtteverktøyet. Semantisk interoperabilitet i samspillet mellom terminologi, informasjonsmodell og beslutningsstøttesystem er derfor avgjørende for utnyttelsesgraden av teknologien.

Nærmere om prosjektet

Prosjektet vil fokusere særlig på 3 hovedfaktorer:

- Forventninger til leverandør basert på kravspesifikasjon i Helseplattformen
  - Vurdering av leverandørenes løsninger mot «Best practice» og Helseplattformens kravspesifikasjon
  - Arkitekturmessig tilnærming for integrasjon av ontologibaserte terminologier og klinisk beslutningsstøtte
  - Kunnskapsoppsummering på hvordan ontologibaserte terminologier kan brukes sammen med kliniske informasjonsmodeller og andre ontologier
- Nasjonale løsninger
  - Kunnskapsoppsummering på skybaserte løsninger for terminologitjenester og beslutningsstøtte
- Nasjonale føringer
  - Følgeforskning på pilotering av SNOMED-CT i tannhelsetjenesten

Overordnet mål

Kunnskapen fra prosjektet vil kunne løfte muligheten for semantisk interoperabilitet. Prosjektet skal også kunne bidra med et konkret kunnskapsgrunnlag i anskaffelsesprosessen Helse Midt-Norge er inne i.

Metode

Prosjektet gjennomføres som følgeforskning (multi-method assessment) for piloteringen av SNOMED-CT gjennom dialogfasen og fram mot kontraktstildeling i 2019.

NSE har bred erfaring på følgeforskning innen e-helse, inkludert semantisk interoperabilitet.

ehealthresearch.no

For mer informasjon, kontakt: Kristian Nicolaisen, tlf: 950 59 001, e-post: kristian.nicolaisen@ehealthresearch.no

15.2 Questionnaire

Clinical Decision Support Systems and Ontology-based Terminologies

Glossary	
CDSS	Clinical Decision Support System.
Reasoner	In the CDSS sense the reasoner (may be or not not be an ontology reasoner) is the software that implements an inference method that processes patient data and determines which recommendations from those in the knowledge base are applicable for that patient.
Inference method	The rules OR statistical models that produce a recommendation e.g. if BloodPressure > 145 THEN (recommend treatment, diet and sport)
Virtual Medical Record	Information model referenced by CDSS logic. It acts as an intermediate layer between the EHR information model and the CDS inference engine.
Information model	This is the data schema (called vmr in CDSS) that may be expressed as openEHR archetypes, hl7 fihr resources, hl7 vmr etc. this schema contains patient data coming from the ehr.
Concept model	Set of rules and criteria that must be followed for the appropriate use of a given terminology. It may or may not be based on formal semantics. In the former case the terminology will be an ontology-based terminology.
Ontology-based terminology	A terminology which concept model defines the relationships among concepts and validity rules by means of formal semantics. For example SNOMED CT and ICD-11.

Table 6: Questionnaire glossary

Ontology-based Terminologies

Transition to ontology-based terminology

- Have you adopted an ontology-based terminology? If so, what kind?
- In what area is the terminology in use?
- Clinical, research/academia, billing etc.
- Have you trained information architects or clinical users on how the ontology-based terminology should be used?
- Do you exploit the formal semantics provided by the ontology?
- If so, how do you exploit them?
- What are, in your view, the main benefits in using an ontology-based terminology?
- What were the main challenges faced in the transition to ontology-based terminology?
- What are the important prerequisites for a successful transition?
- What is your roadmap in regards to supporting more complex implementations of ontology-based terminologies?
- What implementation strategies are you considering in regards to transitioning to ontology-based terminologies? (specialized terminology servers vs. internal in-system support)

Ontology vs. Information model

- Do you use the terminology to annotate section of clinical information models?
- Do you use the terminology to define the possible set of values that may populate a field in the clinical information model?
- What are the important factors that affect how the conceptual models in the terminologies and information models interact?
- Do you think information models becomes redundant when implementing an ontology based terminology, and if so, **how come**?

Methods of use

- Integration with EHR
- Do you perceive a benefit in exploiting the expressive semantics of ontology-based terminologies to query clinical information from the EHR? If so, please explain.
- To get the most benefit from using SNOMED CT in patient records, one must be able not to only query the records, but also query SNOMED CT. Can you give any examples on how queries can be performed in this way?
- How crucial do you consider support for ontology-based terminologies for EHRs to be in the future in order to implementing next-generation structured journal systems also supporting molecular biology and genomics?