Semantic Interoperability for Health Network

Deliverable 4.5: Design Guides and Roadmap

[Version 2, July 31, 2015]

Document description

Deliverable:	D4.5
Publishable summary:	
Status:	
Version:	
Public:	□No **x** Yes
Deadline:	
Contact:	Catalina Martínez-Costa catalina.martinez@medunigraz.at Stefan Schulz stefan.schulz@medunigraz.at
Editors:	Stefan Schulz stefan.schulz@medunigraz.at Dipak Kalra d.kalra@ucl.ac.uk

Table of contents

Contents

1 Introduction and objectives

1.1 Background

SemanticHealthNet (SHN) has focused on improving semantic interoperability of clinical information, based on the assumption that the co-existence between several standards, specifications and proprietary implementations for representing the content of electronic health records (EHRs) will endure. SHN has suggested a layered semantic abstraction (Fig. 1) on top of existing EHR representationsthat provides a homogeneous view, thus enabling semantic mediation across the heterogeneous underlying representations in different contexts.
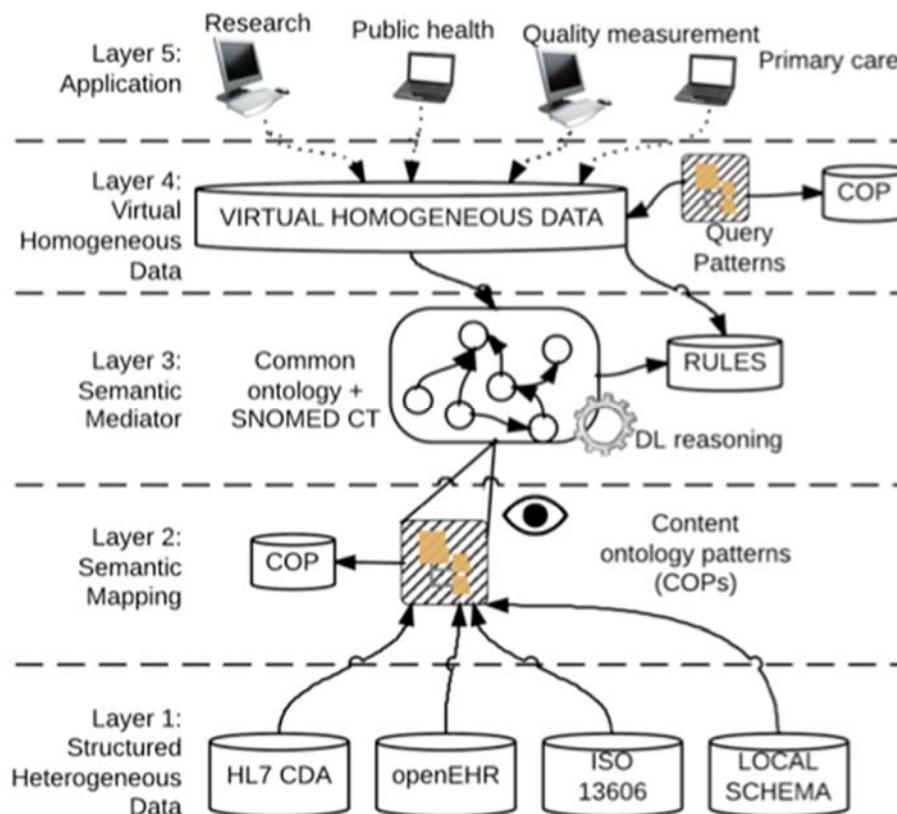
Figure 1. Layered architecture of SemanticHealthNet

The deliverables 4.1 – 4.4 have highlighted three kinds of semantic artefacts, namely **information models** (e.g. openEHR, EN ISO 13606, etc.), **ontology-based terminologies** like SNOMED CT and **clinical practice guidelines**. They demonstrated how clinical content about diagnosis and treatment represented by existing health information standards could formally be described within an ontological framework, which is mostly independent of the specific layout of clinical models, their underlying specifications, as well as on terminology pre- or postcoordination preferences.

SHN's ontological framework makes a top-level distinction between **information entities** and **clinical entities**. It has developed an ontology-based approach for both, including formal constraints on how they are related to each other. Recurring modelling tasks were identified as semantic patterns, intended to facilitate the binding of ontological descriptions to information entities in clinical models by hiding much of the complexity of the underlying description logic-based formalism.

The resulting semantic infrastructure was applied to the modelling of the Heart Failure Summary, a minimal dataset that contains essential information in order to optimize heart failure management. The semantic representation of the Heart Failure Summary aims at facilitating its exploitation by application systems such as decision support and clinical guideline systems.

In addition, the interoperation of computerized clinical practice guidelines with clinical models and ontologies was explored.

1.2   Objectives

This last deliverable of WP4 intends to present a Design Guide and a Roadmap towards semantic interoperability. By design guide we understand a practical manual for all those who want to build semantically interoperable artefacts based on the methodology developed in SemanticHealthNet and laid out, in detail, in the deliverables D4.1 to D4.4. By roadmap we understand a strategic plan that describes steps to be taken, concerning the development and implementation of semantic artefacts as well as future routes towards basic and applied research.

Whereas the design guide is mostly a summary of what has already been published in SemanticHealthNet, the roadmap is influenced by workshops organised in the last project year and by current developments in the field of basic and applied research.
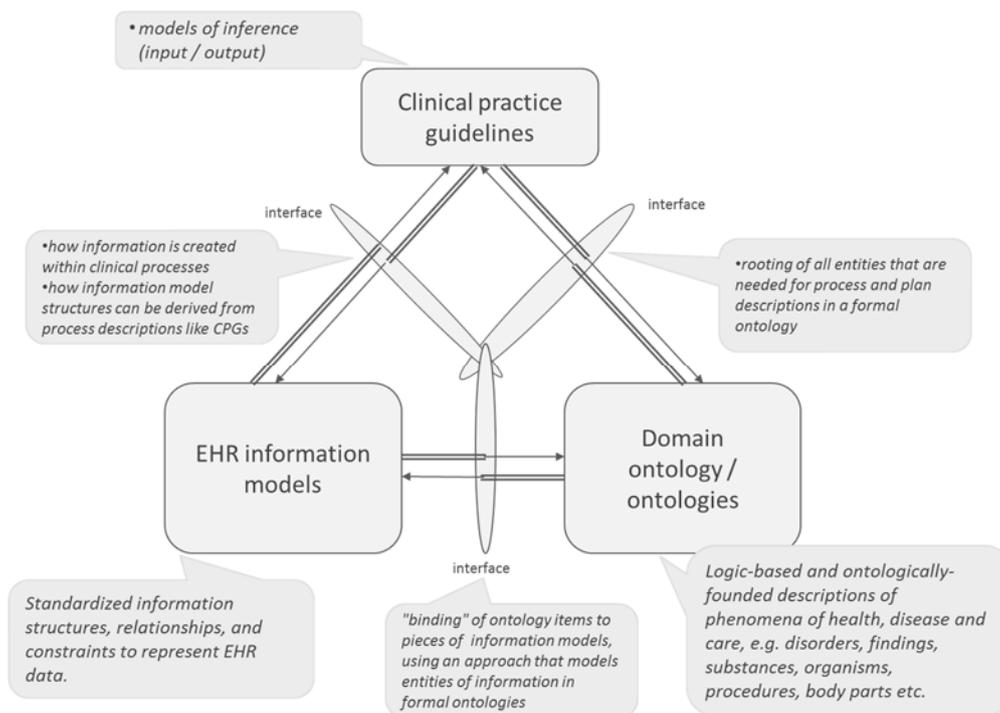
Figure 2 – Semantic Artefacts and their interfaces in SemanticHealthNet

1.3   Methodology

With the priorities highlighted in the following sections, as well as the recommendations issued, the authors attempt to outline a consensus found among most of the participants of WP4, as the result of numerous face-to-face, online and offline discussions, including joint authoring of project reports and scientific papers. These discussion and, in parallel, iterative refinement steps of the general architecture, as well as the experience gathered from the creation and testing of a number of semantic assets, led to a consolidation that is now reflected in this deliverable. The content of the design guide and the

roadmap also draw on external input, such as the CIMI workshop, the Concurrent Use workshop with CEN, and two industry workshops. External input was also gathered by peer reviews of research papers submitted to journals and conferences.

In summary, all parts of this guidance material represent ongoing, collaborative reflections, within SemanticHealthNet and well as between SemanticHealthNet and related communities, such as biomedical informaticians, clinicians, standard developers, policy makers and manufactures.

2 The SemanticHealthNet Design guide

2.1 Basic concepts

The expression "terminology systems" in a broad sense, denotes representational artefacts that describe a domain of interest, providing terms, relations between terms, and language-independent representational units (concepts, classes, relations), identified by codes. Terms can be mapped to other terminologies, and they can be bound to ontologies. Ontologies are, in this document, always understood as "formal" ones: drawing on the philosophical discipline of Ontology[1], artefacts called "formal ontologies" are implementations of theories that give precise logical/mathematical formulations of what is necessarily true "by definition" about the entities represented therein[2]. To be human understandable, ontologies include some terminological component.

There is a general tendency that modern terminology systems are based on ontologies. Especially in the basic biomedical sciences the need for representing large amounts of knowledge has fuelled the development of an increasing body of ontologies, some of them like the Gene ontology (representing cell components and processes), ChEBI (representing molecules)[3], or FMA (representing human body parts)[4] following the OBO Foundry principles[5] for modularization, naming, categorization, and interoperability. In the field of representing clinical data, the clinical terminology system SNOMED CT[6] is ontology-based.

2.2 Design guide for interoperable terminologies and ontologies

SHN posits that all semantic artefacts needed for clinical documentation and decision support share a common ontological basis. This can be summarised as follows:

Ontological upper level

Upper level ontologies provide a basic, mostly domain-independent set of categories, relations and axioms, which are easily reusable across specific domain applications. Upper-level ontologies streamline the development process of semantic artefacts. BioTopLite[7] is the domain upper-level ontology used in SHN. Its uppermost level exhibits the disjoint top categories: *Condition*, *Disposition*, *ImmaterialObject*, *MaterialObject*, *InformationObject*, *Quality*, *Role*, *ValueRegion*, *Process* and *Time*. Its upper

---

[1] Ontology: the study of what there is: Quine W. "On What There Is", Review of Metaphysics 2: 21-38.
Reprinted in From a Logical Point of View (Cambridge: Harvard University Press, 1953). pages 1-19, 1948.

[2] Hofweber T. Logic and Ontology. The Stanford Encyclopedia of Philosophy (Spring 2013 Edition), Edward N. Zalta (ed.), http://plato.stanford.edu/archives/spr2013/entries/logic-ontology/, 2011

[3] Hastings J, de Matos P, Dekker A, Ennis M, Harsha B, Kale N, Muthukrishnan V, Owen G, Turner S, Williams M, Steinbeck C. The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Res. 2013 Jan;41(Database issue):D456-63.

[4] Rosse C, Shapiro LG, Brinkley JF. The digital anatomist foundational model: principles for defining and structuring its concept domain. Proc AMIA Symp. 1998:820-4.

[5] Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ; OBI Consortium, Leontis N, Rocca-Serra P, Ruttenberg A, Sansone SA, Scheuermann RH, Shah N, Whetzel PL, Lewis S. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. 2007 Nov;25(11):1251-5.

[6] www.ihtsdo.org/snomed-ct

[7] Schulz S, Boeker M. BioTopLite: An Upper Level Ontology for the Life Sciences. Evolution, Design and Application. In GI-Jahrestagung, Koblenz. 2013: 1889-1899.

level relations are **causes**, **has condition**, **has participant**, **includes**, **precedes**, **projects onto**, and **represents.** Most of its classes and relations are aligned with the Basic Formal Ontology[8] and the Relation Ontology[9].

Ontology language

There are several languages used in ontology engineering. The most popular one is OWL (Web Ontology Language), recommended by the W3C as ontology language for the Web[10], based on Description Logics (DL)[11], which are subsets of first-order logics. OWL allows for the use of DL reasoners such as Fact++ and Hermit. The practical scalability of reasoning is crucial for the adoption of semantic technologies in real settings.

Domain level

In its examples SHN has used SNOMED CT as the terminology system to represent the clinical domain. With about 300,000 concepts, it provides a high coverage of the clinical domain. SNOMED CT uses the description logics dialect EL++. As it provides polynomial time for reasoning, it easily scales up to huge ontologies.

The current state of SNOMED CT is still characterised by its legacy and only partly follows the precepts of formal ontologies. It can be considered as a hybrid between terminology, ontology and information model. Some of its content is rather a representation of a complex clinical statement than real terminological / ontological units such as "Chronic disease - treatment changed", "Family history of radiation therapy". These have arisen precisely because SNOMED CT is often used in situations in which there is no information model. This reflects a central concern of SemanticHealthNet – the notion of the context in which a clinical statement is made or a code is used.

Clinical terminologies and ontologies have been created to address specific use cases and their further development cannot be expected to follow purist guidelines. Therefore in this design guide the goal is how to take terminology systems "as they are" and provide them with an ontology-based semantics rather than make recommendations how they should look like. SemanticHealthNet has always taken a descriptive and interpretative attitude rather than a prescriptive one.

Not the whole area of clinical care is covered by SNOMED CT; certain contextual and procedural aspects are missing. The gap is partly filled by other semantic artefacts. Of particular interest is ContSys[12], which defined concepts for clinical care and which exhibits several ontology like aspects[13].

Recommendations

- Representational units (classes, concepts) in clinical terminology systems should be rooted in an upper-level ontology. This enforces consensus about the ontological commitment[14] of each terminological unit, which determines under which upper-level category a given representational unit is placed. For instance, "diabetes mellitus" could be understood as a pathological

[8] Smith B, Kumar A, Bittner T. Basic formal ontology for bioinformatics. Journal of Information Systems. 2005: 1-16.

[9] Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, Mungall C, Neuhaus F, Rector AL, Rosse C. Relations in biomedical ontologies. Genome Biol. 2005;6(5):R46.

[10] OWL 2 Web Ontology Language Document Overview (Second Edition). http://www.w3.org/TR/owl2-overview/

[11] Baader F, Horrocks I, Sattler U. Description logics. Handbook on ontologies. Springer Berlin Heidelberg, 2009. 21-43.

[12] EN ISO FDIS 13940 Health informatics - System of Concepts to Support Continuity of Care. CEN, the European Committee for Standardization, ISO TC 215/WG 3, prEN/ISO/DIS, CONTSYS, 08/2014

[13] Martínez-Costa C, Kay S, Oughtibridge N, Schulz S. ContSys under Ontological Scrutiny. Stud Health Technol Inform. 2015;210:999

[14] Schulz S, Cornet R, Spackman K. Consolidating SNOMED CT's ontological commitment. Applied Ontology 6.1 (2011): 1-11.

process, a state, or a patient having diabetes[15] during a life phase. "Suspected diabetes mellitus" is most appropriately be understood as an information object. "Appendectomy" is a procedure; "Planned Appendectomy" is a plan, i.e. a kind of information object.

- Relations in clinical terminology systems should also be aligned with relations in an upper level ontology[16].
- Clinical terminology systems often use high-level categories like disease, disorder, condition, problem, sign, symptom, finding. These terms are fuzzy and context-dependent[17]. We recommend therefore a simplified interpretation that meets clinical intuitions. A solution proposed solution in SemanticHealthNet is to interpret all of them as "clinical life phases" or "clinical situations", i.e. phases of a patient's life in which some clinically relevant entity is present[18].

Examples

- 'Lung Cancer' as Clinical Life Phase:

$LungCancer_{CLP}$ equivalentTo *ClinicalLifePhase* and
        hasCondition some
                (*ClinicalCondition* and
                                (**finding_site** some *LungStructure*) and
                                (**associated_morphology** some *MalignantNeoplasm*))

Clinical Life Phase is a subclass of the BTL2 class *Situation*. Both SNOMED CT relations **finding_site** and **associated_morphology** are aligned with the BTL2 relation includes. The relation **hasCondition** is the interpretation of the SNOMED CT role group used in disorder / findings concepts. Apart from this, the above definition closely follows the corresponding SNOMED CT design pattern.

- 'No Lung Cancer' as a Clinical Life Phase in which there is not cancer in the lung:

$NoLungCancer_{CLP}$ equivalentTo *ClinicalLifePhase* and
            not (**hasCondition** some
                    (*ClinicalCondition* and
                              (**finding_site** some *LungStructure*) and
                              (**associated_morphology** some *MalignantNeoplasm*)))

- Suspected Lung Cancer as a diagnostic statement:

SuspectedLungCancer equivalentTo *DiagnosticStatement* and
        isAboutSituation only $LungCancer_{CLP}$ and
        hasCertainty only *isSuspected*

- Diagnostic statement is an information entity that is outcome of a Diagnosing process. It has some certainty value attached:

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

[16] Schulz S, Martínez-Costa C. Harmonizing SNOMED CT with BioTopLite. An exercise in principled ontology alignment. MEDINFO 2015

[17] Schulz S, Spackman K, James A, Cocos C, Boeker M. Scalable representations of diseases in biomedical ontologies. J Biomed Semantics. 2011 May 17;2 Suppl 2:S6. doi: 10.1186/2041-1480-2-S2-S6.

[18] Cheetham E, Gao Y, Goldberg B, Hausam R, Schulz S. Formal representation of disorder associations in SNOMED CT. International Conference on Biomedical Ontologies 2015.

> *DiagnosticStatement* equivalentTo *InformationObject* and
> **isOutcomeOf** some *Diagnosing* and
> **hasCertainty** some *Certainty*

- The clause **isAboutSituation** only *LungCancer$_{CLP}$* is an approximation of a reference to the type LungCancer[19]. It is equivalent to:

> not (**represents** some (not (*LungCancer$_{CLP}$* or (not *ClinicalLifePhase*)))

The issue here is that the referent of the information entity is a class, not an individual. Classes of classes are not allowed in description logics.

- **hasCertainty** is a subrelation of **hasPart** with *DiagnosticStatement* in the domain and *Certainty* in the range.

Limitations

- The proposed modelling style finds its limitations in the scalability. OWL-EL, the language of SNOMED CT lacks negation ("not"), disjunction ("or") and value restrictions ("only"). The practical scalability of reasoning is crucial for the adoption of semantic technologies in real settings. More and more expressive ontologies have been built in the last years but current DL reasoners cannot efficiently reason with them.
- SNOMED CT is still released as a set of tabular lists. OWL axioms can be generated out of them, but they are not the primary representation. For the construction of compositional expressions in SNOMED CT a so-called compositional grammar is used, which uses so-called role groups or relationship groups[20,21] as a syntactical element, for which there is no correlated in OWL. In the OWL generation process role groups are translated into a relation named "RoleGroup", which is not further defined. The SNOMED CT documentation is vague regarding this issue.
- The representation of grades of certainty follows SNOMED CT, which represents them as simple nominal values, without addressing probabilistic aspects of a real theory of graded statements[22].
- The current DL version of the SNOMED CT context model ("Situation in specific context" hierarchy) exhibits systematic modelling errors which would require major redesign efforts[23].

2.3   Design guide for interoperable clinical models

Domain terms, i.e. entries in domain terminologies follow the rules of human language. In the simplest case they consist of a single noun like "fracture". More commonly, terms are constituted by more

---

[19] Schulz S, Martínez-Costs C, Karlsson D, Cornet R, Brochhausen M, Rector A. An Ontological Analysis of Reference in Health Record Statements. FOIS 2014, 289-302.

[20] Cornet R, Schulz S. Relationship groups in SNOMED CT. Stud Health Technol Inform. 2009;150:223-7.

[21] Schulz S, Hanser S, Hahn U, Rogers J. The semantics of procedures and diseases in SNOMED CT. Methods Inf Med. 2006;45(4):354-8

[22] Krieger HU, Schulz S. A Modal Representation of Graded Medical Statements. Formal Grammar 2015, accepted for publication

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

complex noun phrases, consisting of additional nouns like in "femur fracture", prepositions like in "fracture of the femur" and adjectives like in "fracture of the right femur", or "suspected fracture of the right femur". However, the semantic interpretation not always parallels the syntactic makeup of such expression. Whereas "fracture of the right femur" specialises "fracture of the femur" (which specialises "fracture"), this is not the case with the expression "suspected fracture of the right femur". The latter expression denotes not a fracture but a piece of information, *viz.* the belief expressed by some health professional that a femur fracture may exist in a patient.

Clinical models have been constructed as normalised containers of clinical terms but also numeric information, like paper forms, which require a more structured and less arbitrary recording of information compared to human prose. Nevertheless, clinical models as such do not guarantee interoperability. For instance, they may split the information contained into several facets, such as

| DISORDER: | *Fracture* | LOCATION: | *Femur* |
| LATERALITY: | *Right* | CERTAINTY: | *Suspected* |

Alternatively, they may exhibit a less fragmented shape, especially if a large collection of complex terms can be used, such as in

| DISORDER: | *Fracture of the right femur* | CERTAINTY: | *Suspected* |

or

| DISORDER: | *Suspected femur fracture* | LATERALITY: | *Right* |

It is obvious that the same piece of information can be represented in many different ways even within the same information model specification such as openEHR, HL7, or EN 13606[24]. Such standards often include their own vocabulary such as to express gender, severity, laterality or diagnostic certainty, which competes with terms provided by terminologies. Furthermore, they often carry hidden assumptions. For instance, the expression "Femur fracture?" written by a general practitioner into a paper form that accompanies the patient to the hospital, implies a lower diagnostic certainty than the same term with in a structured radiology report. The information "Pack years: 10" in the admission form of a smoking cessation service is certainly more error-laden than the information "Haemoglobin: 10.0 mg/dl" in a lab report. Whereas the latter one implies that the information is the output of a well-defined diagnostic methods, the former one implies that the number of cigarettes smoked is reported by the patient, which may entail fuzziness and bias.

The problem of different distribution of contextual information (temporality, certainty, negation, plans) between clinical models and the one hand and terminologies / ontologies on the other hand is well known. E.g., the SNOMED CT concept *395099008 |Cancer confirmed (situation)|* envelops the *disease* (i.e. a clinical entity) concept Cancer into a *diagnostic statement*. Similarly, the SNOMED CT concept *160244002 |No known allergies (situation)* includes a negation. This problem had been addressed in Terminfo[25], a guide for the use of SNOMED CT in the HL7 V3 Clinical Statement pattern, which recommends which parts of a diagnostic statement should be represented by the terminology and which ones by the information model.

---

[24] Martínez-Costa C, Schulz S. Ontology content patterns as bridge for the semantic representation of clinical information. Appl Clin Inform. 2014 Jul 23;5(3):660-9

[25] Using SNOMED CT in HL7 Version 3; Implementation Guide, Release 1.5:
http://www.hl7.org/v3ballot/html/infrastructure/terminfo/terminfo.html

Similar guidance was given in the Logical Record Architecture[26] and in Detailed clinical models[27]. Principles of modelling clinical information were suggested by SIAMS[28], the Clinical Element Model[29,] and, most recently CIMI, the Clinical Information Model Initiative[30]. As shown before, SNOMED CT provides codes for contextual expressions under the *Situation with Explicit Context* concept model category and provides post-coordination patterns in its Concept model. A similar approach is pursued by the future ICD-11. The reason for this is that thus these terminologies can be used in an expressive way without depending on a separate information model. However, the ontological representation of such statements is not straightforward and still subject to discussion[31,32].

Recommendations

For rooting information structures in formal ontologies and thus improve their interoperability with other information structures as well as with pre- and postcoordinated expressions in terminologies we recommend that all fixed and variable elements, i.e. headings and value sets are annotated by means of an ontology. The requirements for such an ontology include the ones listed in 2.2.

Examples

The following example may illustrate the above requirement. It uses a very simple example and abstracts from any clinical model specification. All three information structure represent exactly the same kind of information, which is, however distributed in different ways between the information elements and the underlying ontology (Fig. 3)

Figure 3 Example for three equivalent renderings of the diagnostic statement "Suspected heart failure caused by ischemic heart disease"

Fig. 4 shows the ontological annotation of the left hand model, both in natural language and equivalent description logics representations. Each annotation is expressed in a way that it represents a facet of the whole information structure, viz. that it corresponds to a diagnostic statement.

In contrast, Fig. 5 depicts the right hand side example from Fig. 3 in the same way. A closer scrutiny suggests that the sum of the annotations for either example are equivalent. This can be proven by submitting this expression to a description logics classifier like HermIT, which computes equivalence.

All expressions that correspond to this pattern are of the type "information object". In their reference to classes of disorders (e.g. from SNOMED CT), they avoid logical implications which would be given by

---

[26] Logical Record Architecture for Health and Social Care:
http://www.connectingforhealth.nhs.uk/systemsandservices/data/lra
[27] ISO/CD TS 13972 Health informatics -- Detailed clinical models, characteristics and processes
[28] http://informatics.mayo.edu/CIMI/images/c/c6/GF_EN13606_SIAMS_for_CIMI.pdf
[29] SHARPn project: http://informatics.mayo.edu/sharp/
[30] CIMI initiative: http://informatics.mayo.edu/CIMI/
[31] Martínez-Costa C, Schulz S. Ontology-based reinterpretation of the SNOMED CT context model. ICBO 2013, Montréal, Canada: 90-95
[32] Schulz S, Martínez-Costa C, Karlsson D, Cornet R, Brochhausen M, Rector AL. An Ontological Analysis of Reference in Health Record Statements. FOIS 2014, Rio de Janeiro, Brazil. 289-302

using the existential quantifier "some". This avoids false implications, such as asserting the existence of a disease instance in a patient by mentioning the corresponding ontology class in a description logics expression.
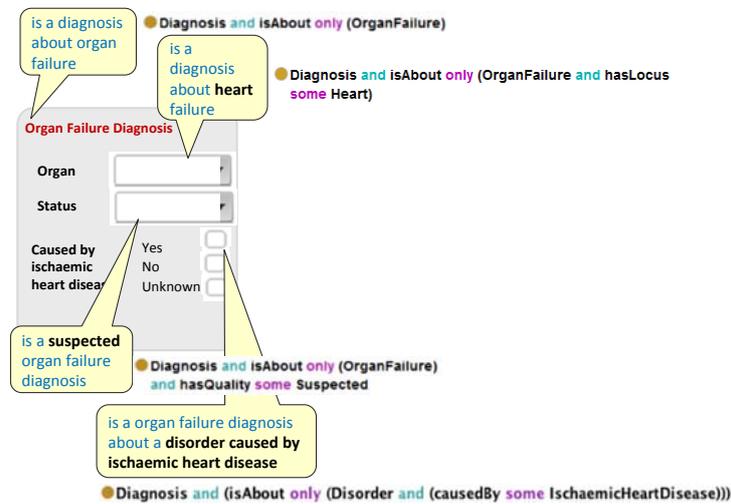

Figure 4 Semantic annotation of the left hand side example information model (Fig.3) with free text and corresponding OWL statements
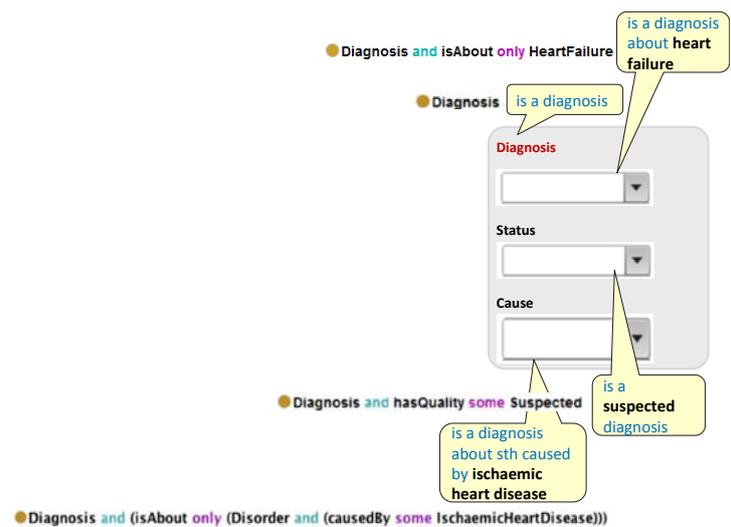
Figure 5 Semantic annotation of the right hand side example information model (Fig.3) with free text and corresponding OWL statements

Limitations

The technique to reduce the whole content of instantiated clinical models to description logics statement is, in theory, appealing, but entails several limitations:

- The performance of description logics based representation languages decreases dependent of size and language expressivity. It is still open how the performance of the technologies used for the implementation of the proposed approach (e.g. OWL, RDF, SPARQL), can be kept within reasonable values. This directly affects scalability.

- To reach a complete consensus of the model proposed is impractical. Therefore, to develop a set of models of minimal information to be interchanged could at least offer a reasonable and more than acceptable degree of semantic interoperability to be achieved.

- It cannot be expected to develop and maintain all the ontology patterns needed to cover the entire medical domain. This should be ideally supported by a community of experts supported by specific tools, the same way as actually other organisations work such as the IHTSDO. Patterns should be well-organized, ideally in a hierarchical structure where more general patterns can be specialized by more specific ones. Additionally we may want to consider also a compositional / modular structure of patterns. The approach should be harmonised with on-going activities regarding documentation patterns, e.g. promoted by CIMI.

- From a usability point of view, it requires from the implementer to be able to construct well-formed OWL expressions. Although semantic web technologies have made great progress, the command of ontology languages (like e.g. the command of the database query language SQL) cannot be expected by implementers.

The last limitation is addressed by the notion of ontology patterns. Semantic patterns (i.e. ontology content patterns) are reusable solutions to recurring modelling problems[33]. Similar to templates they have fixed and variable parts[34], but unlike databases schemas or UML models, they are based on an ontological framework, in order to bridge between the EHR modelling community and the Semantic Web and formal ontology communities, in order to guide and standardize the representation of the information meaning encoded by clinical models. Table 1 shows an OWL representation of such a semantic pattern.

---

[33] Gangemi A, Presutti V. Content Ontology Design Patterns As Practical Building Blocks for Web Ontologies. Proceedings of the 27th International Conference on Conceptual Modeling. Barcelona, Spain: Springer-Verlag; 2008: 128-41.

[34] Blomqvist E, Daga E, Gangemi A, et al. Modelling and using ontology design patterns. http://www.neon-project.org/web-content/media/book-chapters/Chapter-12.pdf

Table 1. OWL DL representation of the semantic patterns for describing the presence or absence of a symptom and its severity. They are represented in Manchester syntax and by following the naming convention as stated in the Methods Section. Symbols starting with a question mark are variables, i.e. they can be substituted by any subclass of the referred ontology class.

Symptom present record pattern

shn:SymptomPresentRecord subClassOf
     shn:SymptomRecord
          and **shn:isAboutSituation** only *?ClinicalSituation*
          and **shn:hasInformationAttribute** some *?Severity*

Symptom absent record pattern

shn:SymptomAbsentRecord subClassOf
     shn:SymptomRecord and **shn:isAboutSituation** only
       (*shn:ClinicalSituation* and not (**btl:includes** some *?ClinicalSituation*))
     and **shn:hasInformationAttribute** some *?Severity*

Table 2. OWL DL pattern instantiation with symptom related data captured by Applications A and B, for Patients A and B respectively

PATIENT A - Application A: Breathlessness and chest pain symptoms

Individual: SymptomEvaluationProcess_PatientA
Type:*sct:EvaluationSignsAndSymptoms* and *btl:hasParticipant* value PatientA
 	
Individual: SymptomA_Breathlessness_Present
Type: *shn:InformationItem*
       and **btl:isOutcomeOf** value SymptomEvaluationProcess_PatientA
       and **shn:isAboutSituation** only *sct:Breathlessness*
 	
Individual: SymptomA_ChestPain_Present
Type: *shn:InformationItem* and **btl:isOutcomeOf** value SymptomEvaluationProcess_PatientA
       and **shn:isAboutSituation** only *sct:ChestPain*

PATIENT B - Application B: Mild breathlessness on exertion but not at rest symptoms

Individual: SymptomEvaluationProcess_PatientB
Type:*sct:EvaluationSignsAndSymptoms* and *btl:hasParticipant* value PatientB
 	
Individual: SymptomB_BreathlessnessOnExertion_Present
Type:*shn:InformationItem* and **btl:isOutcomeOf** value SymptomEvaluationProcess_PatientB
       and **shn:isAboutSituation** only *sct:BreathlessnessOnExertion*
       and **shn:hasInformationAttribute** some *sct:Mild*
 	
Individual: SymptomB_BreathlessnessAtRestAbsent
Type: *shn:InformationItem* and **btl:isOutcomeOf** value SymptomEvaluationProcess_PatientA
       and **shn:isAboutSituation** only (*shn:ClinicalSituation*
       and not (**btl:includes** some *sct:BreathlessnessAtRest*))

2.4 Design guide for interoperable clinical guidelines

A clinical guideline is a set of systematically developed statements to assist the decisions made by healthcare actors about healthcare activities to be performed with regard to specific health issues[35].

Health issues can be represented by an ontology. The main relationship between them and clinical guidelines is how the applied knowledge will influence the interpretations of observed aspects of a clinical condition on the one hand, and the choices of healthcare activities to be performed, on the other hand. Despite the evolution of clinical terminology and ontology standards, there has been little use of them in clinical guidelines. E.g. in the ESC guideline we find the wording "Increased time to recover from ordinary physical activity" whereas SNOMED CT provides codes for such as breathless-ness, fatigue and tiredness.

Recommendations

In order to improve interoperability between guidelines, ontologies and clinical models, we recommend the following:

- Guideline creators should map their non-standardized vocabularies to standardized ones such as SNOMED CT.

- The semantic patterns as proposed in SHN should be used to map data elements in clinical models to a set of standardized ontologies and use them in the guideline representations.

- Information models should be based on clinical process models, whereas clinical guidelines influence the performance of clinical processes.

- A seamless knowledge management between guidelines and clinical models should be established, by using the same information model.

- EHR information models should hold the information necessary to determine whether guideline criteria are satisfied. This has also take into account that the criteria in a guideline may not be expressed at the same granularity as in the medical record.

Example:

Several clinical practice guidelines have been developed, but their lack of formalization often precludes machine processing.[36] SemanticHealthNet focused on the Guideline Definition Language (GDL)[37] and PROforma[38], their representation of constraints within the clinical guideline rules that trigger the medical decision steps, and their linkage with the electronic health record (EHR).

Fig. 6 shows the decision tree of an atrial fibrillation guideline.

Four different kinds of constraints are checked against the EHR, viz. numeric constraints (e.g. LVEF <35%); constraints including epistemic values (e.g. "severe frailty"), constraints regarding presence or absence of EHR information (e.g. LVEF not measured), and constraints that exploit the hierarchy of the underlying terminology (e.g. chronic LVSD as child of LVSD).

---

[35] Health issue is a representation of an issue related to the health of a subject of care as identified by one or more healthcare actors (e.g. a loss of weight, a heart attack, a drug addiction, etc.)

[36] Peleg M, Tu S, Bury J, Ciccarese P, Fox J, Greenes RA, Hall R, Johnson PD, Jones N, Kumar A, Miksch S, Quaglini S, Seyfang A, Shortliffe EH, Stefanelli M. Comparing computer-interpretable guideline models: a case-study approach. J Am Med Inform Assoc. 2003 Jan-Feb;10(1):52-68.

[37] https://github.com/openEHR/specifications/blob/master/proposals/GDL/GuideDefinitionLanguage_v0.9.pdf?raw=true

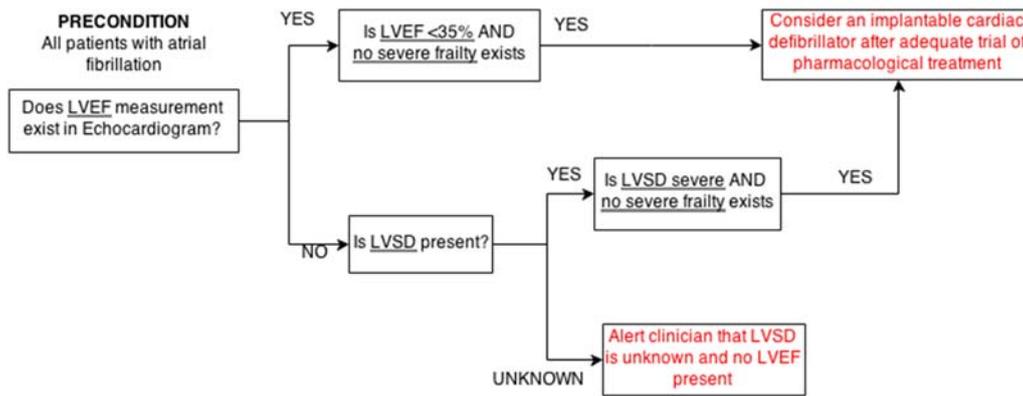[38] http://www.openclinical.org/gmm_proforma.html

Figure 6 Decision tree for rule for patients with atrial fibrillation

GDL[39] is based on the openEHR reference and archetype models. GDL rules use archetypes as input and output. The formalization of the atrial fibrillation example with GDL begins with establishing correspondences between rule variables and the archetypes[40].

Table 3 Example of GDL binding of rule variables with archetype data elements and the rule definition

```
definition = (GUIDE_DEFINITION) <
archetype_bindings = <
["gt0004"] = (ARCHETYPE_BINDING) <
        archetype_id = <"openEHR-EHR-EVALUATION.problem-diagnosis.v1">
        domain = <"EHR">
                elements = <
                        ["gt0005"] = (ELEMENT_BINDING) <
                                path = <"/data[at0001]/items[at0002.1]">
                        >
                        ["gt0006"] = (ELEMENT_BINDING) <
                                path = <"/data[at0001]/items[at0005]">
                        >
                        ["gt0014"] = (ELEMENT_BINDING) <
                                path = <"/data[at0001]/items[at0.32]">
                        >
                >
                predicates = <"/data[at0001]/items[at0002.1] is_a
        local::gt0005|LVSD|",...>
        template_id = <"diagnosis_snomedct">
>>>>
rules = <
["gt0017"] = (RULE) <
when = <"$gt0005!=null", "$gt0007==local::at0.33|unknown|">
then = <"$gt0018.create($gt0019='Alert clinician that LVSD is unknown and not LVEF present')",...>
priority = <3>
>
```

The archetype data elements in Table 3 relevant for mapping the LVSD are: (i) the diagnosis itself (e.g. LVSD); (ii) status ("unknown") and severity ("severe"). The three of them are identified by their corresponding path within the archetype.

SemanticHealthNet developed an approach how clinical guideline rules can be expressed as queries against EHR data structured by clinical models that have their data elements and values mapped to expressions in OWL DL or RDF. Using three openEHR archetypes[41] fictitious data instances of a patient with atrial fibrillation, LVEF 20%, no severe frailty, and LVSD present were used. For the complete example see D4.4

One semantic pattern used is the *Information about clinical situation* pattern. It represents some piece of information about a particular clinical situation of the patient. The pattern consists of a set of subject-predicate-object triples (with cardinality) of which the first one will presented here:

Subject: shn:*InformationItem*	Predicate: '*describes situation*'
Cardinality: 1..*	Object: shn:*ClinicalSituation*

The correspondences between the clinical model data elements and the semantic pattern triples are established as follows:

ARCHETYPE_ID / ADL Path:	openEHR-EHR-EVALUATION.problem-diagnosis.v1 /data[at0001]/items[at0002.1]
Data Element / Value:	Diagnosis / Disease
Triple representation:	shn:*InformationItem* '*describes situation*' shn:*ClinicalSituation*

The translation of the pattern representation into OWL-D is given by the following rule:

'*describes situation*'	*SUBJ*	subClassOf shn:**isAboutSituation** only *OBJ*

with the subject and object placeholders corresponding to subclasses or instances of the fillers of the Subject and Object slots in the pattern.

An individual data instance is then annotated as follows (only the first element is shown):

Individual: i1
Type: shn:*InformationItem*
            and shn:**isAboutSituation** only sct:*AtrialFibrillation*
            and …
            and …

The guideline rule can then be used as a set of queries in SPARQL (see examples in D4.4) acting on this DL representation.

Limitations

The limitations below refer to the two guideline specifications analysed (GDL and Proforma).

- Neither guidelines, nor their process descriptions have a standardised way to relate to EHR information models and/or domain ontologies, whereas there is an increasing consensus that such links would largely improve data capture, process support and selection and aggregation of data for secondary purposes.

---

[41] openEHR-EHR-OBSERVATION.imaging_exam-Echocardiography.v1
   openEHR-EHR-EVALUATION.problem-diagnosis.v1
   openEHR-EHR-EVALUATION.exclusion-problem_diagnosis.v1

- GDL allows the binding between local concepts used in the rules and concepts from reference terminologies, but does not offer support for post-coordinated expressions
- GDL does not cover process aspects, in contrast to Proforma.
- Limited support for reasoning (only taxonomic subsumption; could be also extended in the future if the use case requires it).
- PROforma does not provide a proper information model but allow the reuse and definition of what they call data definitions.
- PROforma does not provide support for managing external terminologies such as SNOMED CT.
- The need of mapping clinical model data elements and their values to semantic patterns as a previous step

3 The SemanticHealthNet Roadmap from an Industry point of view

The SemanticHealthNet roadmap can be seen as an update and refinement of the SemanticHEALTH roadmap[42], published in 2008, after seven years of research, development and discussion within and outside the SemanticHealthNet NoE, which has worked with clinical and public health domain experts to establish their semantic interoperability needs, as well as with standards and specification development experts to propose semantic interoperability assets to meet these requirements. Other multi stakeholder groups have examined the challenges with scaling up the adoption of semantic interoperability standards, including the value propositions and business models needed to grow the market and to make it more responsive to real user needs. Although industry representatives have been part of many of these activities, SemanticHealthNet wished to collate a wider consensus of industry opinions on these semantic interoperability adoption challenges, and in particular to look at the interface between industry and several other critical stakeholders who are driving the needs for, and procurement of, interoperable eHealth solutions.

A two day industry forum on semantic interoperability was held in January 2015, with around sixty participants from the health ICT industry, clinicians, health ministry and health insurer representatives, public health and clinical research representatives, standards development organisation representatives and others active in the field of health informatics.

The outcome of this event has shape this roadmap. In encompasses the perspectives of (i) clinicians and patients; (ii) healthcare purchasers; (iii) clinical research; and (iv) standards development organisations.

Main principles and recommendations arising from the panel on clinician and patient perspectives

3.1 The perspective of clinicians and patients

Clinical discourse is often vague, preliminary and imprecise. Semantic standards must cater for this. Instead of imposing a simplistic interpretation of clinical statements, which suggest unreal precision they need to cater for uncertainty and excluded findings, as well as for the distinction between plans, risks and real facts.

The interface between clinicians and documentation is the clinicians' language. In order to ensure high quality entry applications and medical terminologies must reflect the words and phrases of clinicians and the patients. The characteristics of laypersons language needs to be taken into account. Standard vocabularies should be developed by national bodies, not the vendors.

Systems should be designed in a way that they encourage the entry of structured data, ideally in harmony with the still prevailing production of free text. The proportion of information only available as free text should be reduced.

There is an agreement among all stakeholders that too much data is collected. The more data is collected, the higher the risk of decrease in data quality. Clinicians should define what data are really needed to support decision making. The starting point has to be capturing the clinical context, and for clinicians to be able to transfer their needs to clinical modellers in a standardised way. Despite the added value of clinical decision support out of good data, we cannot expect enough intrinsic motivations among clinicians to produce high quality data, whereas documentation tasks are done under time pressure and in competition with patient-related work. Therefore incentives, including monetary ones, need to be provided to ensure good documentation of original clinical data and manually curated summaries. Summaries are a good starting point for the creation of interoperable records, with the scope of structured clinical

---

[42] http://www.semantichealth.org/

eHealth interoperability needs to be simple, but a "one size fits all" strategy misses the point. Multiple standards are unavoidable, but they must be shown to interoperate. Therefore procurement contracts need to prioritise interoperability, to stimulate industry adoption of standards.

Incentives and penalties must be developed and applied to push healthcare providers, both organisations and individuals to follow evidence based practices. Health regions should provide an integration platform on which multiple vendor solutions can build, to support smaller vendors and ensure they will be interoperable.

3.2   The healthcare purchaser's perspective

The provision of the framework for interoperability is a major task of health authorities like ministries. They should specify and promote the role of eHealth to help remodel healthcare services, empower patients, shift the focus towards prevention and wellness through central data collections like registries. Although they should not be the ones who develop clinical content standards, they should make investments in these standards, define the needs for them, and promote their use by making them freely available for industry to use. It is also their task to create the market conditions that promote (multi-vendor) interoperable applications and devices. Interoperability must be mandated by financial incentives. Health ministries, insurers and commissioners should promote person centred care contracts. Such contracts will push providers to collaborate and co-ordinate care and to engage patients in the care process.

We still need evidence for the added value of semantic interoperability, to convince healthcare funders and procurers, to position semantic interoperability as an infrastructure. This requires a business model for investment in building and maintaining the necessary infrastructure, info-structure and services. Industry needs to play a stronger role in the growth of the app market SMEs bring a rapid and agile response to user needs, but they need guidance on strategies for adopting interoperability standards. Agreed standards at a European level will enable SMEs to sell their products in multiple countries, however recognising that most of them cannot easily enter the US market, while the US vendors can more easily enter Europe.

3.3   The perspective of clinical research and medical knowledge acquisition

Whereas clinical trial will continue to be golden standards in the building of biomedical knowledge, they often present a nice picture of a dirty world. The paradigm of personalised medicine requires that this needs to be complemented this with real world data. There are huge opportunities for secondary use of clinical data. This can already be initiated by reusing the data items that are already that are consistently captured in information systems. However, semantic disharmony is a major obstacle to scaling up data on a global level such as to understand chronical diseases like dementia, to combine data from multiple wearable devices and sensors.

Clinical data should feed a continuous tracking and learning mechanism to produce new knowledge capitalising on the fast evolution of "big data" technologies. This requires an open data access mentality, so that research sponsors and data sources support each other with scaling up data access and use. This requires adoption of standards. By stimulating the demand side, industry will take the risk of moving away from localised innovations toward global and interoperable ones.

3.4   The perspective of standards development organisations

SDOs want to collaborate towards interoperability. The brand wars are over, the community know each other and want to work together, data models are less and less the point of vendor competition. However, the list of currently available standards relating to EHR data is still huge. The success of standards depends on their financial and intellectual accessibility. The value of using a forthcoming standard must be defined prospectively, used to justify its adoption, and then evaluated. The benefits of standards need to be made clear to clinicians, e.g. by the avoidance of duplicate data entries and

the reduction of patient risks, which incur from terminology mappings introduce. The European perspective must place an emphasis on quality assured multi-lingual assets.

SDOs must be honest about what can be delivered, and by when. Current standards development processes necessarily take time to gain community engagement and consensus. We can work more quickly on guidance on how existing standards can be used - guidelines must be clear and help bring "interoperability of minds".

Besides the patient safety aspects they should point out the cost-effectiveness aspects of standards adoption. It is important to show that standards can be profiled to meet specific use cases, and demonstrated to work in the real world (e.g. with testing tools, support packs, a support community).

Quality Management Systems are required in health informatics standards development - top down and bottom up, including feedback loops to learn from standards adopters. Collaborative communities of standards experts in the Member States need to be formed, through coordination and iterative consolidation at the European level to strengthen the voice of Europe internationally. Certification of EHRs narrows down the diverse interpretations of the (many) available standards

4 The SemanticHealthNet Roadmap from a R & D point of view

Participants in SemanticHealthNet WP 4 have actively participated in numerous scientific events, such as workshops and conferences in the areas of biomedical informatics (MIE, MEDINFO, eHealth) , biomedical ontologies (ICBO), formal ontologies and information systems (FOIS), and SemanticWeb (OWLED). In addition, several journal publications were published or are still under review. The intensive exchange with several scientific communities and the feedback to SemanticHealthNet contributions (also in the peer review processed to which all scientific manuscripts were submitted) has shown routes to follow-up research activities.

The following list constitutes a roadmap that draws on R & D achievements during the project. It addressed activities that have been initiated but not completed, as well as new R & D to be carried out to improve semantic interoperability in health care and biomedical research. The topics vary considerably in scale, some of them would require efforts in the range of 30 – 60 PMs.

4.1 Semantic assets produced in SemanticHealthNet and further developments thereof:

- Semantic assets produced within SemanticHealthNet such as the Heart Failure Summary should be proposed to be used as test beds to exploring in follow-up standards alignment activities. For instance HL7 has tested the incorporation of Heart Failure Summary samples expressed in OpenEHR, into the Art Decor tool[43]. The Heart Failure Summary will also be used within the ASSESS-CT project to assess coverage and quality of different terminology scenarios.
- The semantic pattern approach should be harmonised with similar models for semantic interoperability, such as the ones developed by CIMI[44].
- A large collection of clinical models from different specifications should be used to identify their recurring elements, from which semantic patterns can be derived. The potential of modularisation of these patterns should be further explored, with the aim to propose a library of modular patterns, each of which annotated with OWL expressions.
- Once this pattern library is made available it is applied to a broad set of clinical models from different specifications. These enriched models are then tested against a benchmark consisting of sample record data and related competency questions.
- A similarly semantic enrichment should be applied to examples of clinical guidelines. Here, also, the feasibility of the approach has been validated against a benchmark.

4.2 External Interoperability standards

- The on-going development of SNOMED CT should incorporate principles of applied ontology. This is a major requirement for its use within the SemanticHealthNet ontological framework. The goal should be to align it with an upper level ontology and an upper domain level ontology. Use cases should be formulated how this can be used for improving quality of the terminology. In particular, the following tasks should be emphasised:
    - o Correction of the SNOMED CT concept model patterns in Situation hierarchy
    - o Complete separation between clinical entities and information entities
    - o Semantic clarification of the Role Group constructor
    - o Substitution of the compositional grammar by a semantic web conformant language, e.g. the OWL Manchester Syntax

---

[43] https://art-decor.org/
[44] http://opencimi.org/

- Other standards and specifications that are not explicitly ontologies but include major ontological content, should be submitted to ontological scrutiny and be aligned to or annotated with existing ontologies. This desideratum addresses especially Contsys, which represents processes of clinical care, but also the classical information model specifications (HL7, openEHR, EN13606) and CIMI
- As a far reaching goal, the development of all semantic artefacts (standards and specifications) should be ontology-based. This will require guidelines for ontology engineering (e.g. based on the GoodOD guidelines[45] and / or OBO Foundry principles[46]). It should also include conventions for documentation and naming, including a glossary aiming at reducing the terminological confusion in the field of biomedical semantics, terminology and ontology

4.3 Opening the scope to clinical narratives

- Further EHR standards and interoperability endeavours have to acknowledge that the acceptance of EHR and terminology standards is still poor. Most EHR information is formulated as narratives, using idiosyncratic, highly compact technical language. This practice is deeply rooted in medical culture, and the expectation that structured recording substituted free text documentation in EHRs to a larger extent seems unrealistic.
  The gap between EHR free text documentation and standardised structured documentation needs to be bridged. Using automated or semi-automated approaches to transform clinical content from a free text into a standardised and semantically explicit form as required for inter-system and inter-lingual health record interoperability, health statistics, decision support and secondary usage scenarios, is an increasingly important topic.
- This requires the development of language technologies, especially for European languages other than English. These technologies need to be leveraged by language resources such as annotated domain corpora in several languages.
- The efficient use of natural language processing for clinical information extraction requires an enlargement of the existing terminology resources. Large terminologies like SNOMED CT need to be enriched by synonyms and interface terms that represent the language usage in clinical narratives.
- The development of trustworthy disambiguation approaches constitutes is crucial due to the importance of short forms and elliptic expressions in clinical narratives.
- The role of speech recognition technology should be explored, within a processing pipeline that starts with dictation, over text production text curation, information extraction on the fly to the automated population of semantically interoperable clinical models.

4.4 Topics for basic research

- Development of design patterns for aboutness and reference, ideally within computable logic, drawing on foundations of linguistic semantics, semiotics and metaphysical investigations. These design patterns should then be applied and validated with referring expressions in clinical documentation, especially where the existence of an individual referent cannot be taken for granted
- A principled approach to graded (un)certainty statements in clinical discourse and documentation and their representation in a computable framework
- Systematic studies of scalability of ontology patterns, with regard to description logics reasoning and SPARQL querying, driven by the semantic patterns developed as intermediate representation in SemanticHealthNet

---

[45] http://www.iph.uni-rostock.de/forschung/forschungsprojekte/good-ontology-design-dfg/guideline/
[46] http://www.obofoundry.org/crit.shtml

- Co-operation with the description logics community for which experimental ontologies are created, which exhibit characteristics of the interoperability artefacts.