Title:

Semantic Enrichment of Clinical Models towards Semantic Interoperability.
The Heart Failure Summary Use Case

Corresponding author:

Full name: Catalina Martínez-Costa
Postal address: Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, LKH-Eingangszentrum, 3. und 5. Stock, Auenbruggerplatz 2,8036 Graz, Austria
E-mail: catalina.martinez@medunigraz.at
Telephone: +43 - 316-385-17988

Co-authors:

Full name: Ronald Cornet
Department: Academic Medical Center
Institution: University of Amsterdam
City: Amsterdam
Country: Netherlands

Full name: Daniel Karlsson
Department: Department of Biomedical Engineering
Institution: University of Linköping
City: Linköping
Country: Sweden

Full name: Stefan Schulz
Department: Institute for Medical Informatics, Statistics and Documentation
Institution: Medical University of Graz
City: Graz
Country: Austria

Full name: Dipak Kalra
Department: Centre for Health Informatics & Multiprofessional Education
Institution: University College London
City: London
Country: United Kingdom

Keywords: Electronic health record, terminology, SNOMED CT reference terminology, ontology, semantics, knowledge representation

Abstract

Objective

To improve semantic interoperability of Electronic Health Records by ontology-based mediation across syntactically heterogeneous representations of the same or similar clinical information.

Materials and Methods

The approach is based on a semantic layer that consists of: (1) a set of ontologies supported by (2) a set of semantic patterns. (1) helps standardize the clinical information modeling task, (2) shield modelers from the complexity of ontology modelling. This approach is applied to heterogeneous representations of an excerpt of a heart failure summary.

Results

We demonstrate that semantic patterns or compositions thereof, using a set of finite top-level patterns to derive the former ones, can be used to represent information from clinical models. Homogeneous querying of the same or similar information when represented according to heterogeneous clinical models is feasible.

Discussion

This approach focuses on the meaning embedded in health records, regardless of their structure. This complex task requires a clear ontological commitment (i.e. agreement to use the shared vocabulary consistently within some context), together with formalization rules. This is supported by semantic patterns. Other potential uses such as clinical models validation require further investigation.

Conclusion

We have shown how an ontology-based representation of a clinical summary, guided by semantic patterns, allows homogeneous querying of heterogeneous information structures. Whether there are a finite number of top-level patterns is an open question.

OBJECTIVE

Semantic interoperability [1] of clinical information has been put on the agenda by many organizations and initiatives. Variegated requirements for data capture and storage have motivated the development of standards and specifications for terminologies, ontologies and clinical models. Nevertheless the semantic interoperability problem persists.

The European SemanticHealthNet network [2], follows the recommendations of its predecessor project SemanticHealth [1], in seeking for a closer integration between information models, as used in electronic health records, and terminologies and ontologies, to improve semantic interoperability. The project pursues this goal by dissecting heterogeneous representations of clinical information based on formal-ontological principles. Thus, a shared model of meaning enables precise annotations of which each information item in a clinical model signifies, using the Semantic Web language OWL DL [3]. Complicated language constructs are addressed by using semantic patterns as intermediate representations.

The project has set the focus in chronic heart failure and cardiovascular prevention, as use cases. Here we exemplify the approach proposed by focusing on the Heart Failure Summary [4], a structured clinical model driven by clinicians requirements within SemanticHealthNet, that summarizes basic aspects of heart failure diagnosis and care in order to optimize the disease management.

BACKGROUND AND SIGNIFICANCE

Semantic interoperability of Electronic Health Record (EHR) systems requires a clear sharing of roles between several layers of representational artifacts [5]: (1) generic EHR information models; (2) clinical data structure definitions, *viz.* clinical models; (3) top-level ontologies, (4) domain ontologies, and (5) terminologies. Whereas the latter three have been referred to as *models of meaning*, the former two have been referred to as *models of use* [5].

Generic EHR information models provide standardized information structures, relationships, and constraints to represent EHR data. Examples are EN ISO 13606-1 [6], openEHR Reference Model [7] or HL7 Reference Information Model (RIM) [8]. The meaning they convey relies on the intuitive and common-sense understanding of natural language labels and descriptions, not *a priori* referring to any ontological foundation. E.g., openEHR distinguishes between information structures for representing observation and evaluation results and instructions, while EN ISO 13606 provides a general "entry" information structure for the three cases. The meaning of terms such as "evaluation", "observation", "entry" are only informally elucidated in the documentation. The openEHR information model is partially ontology-based [9], but it is not rooted in any upper level ontology and thus lacks a clear ontological commitment [10].

Clinical models like EN ISO 13606/openEHR archetypes [10], HL7 CDA documents [12] or Clinical Element Models (CEMs) [13], constrain the information model structures for serving particular data capture and communication use cases. As an example, a blood pressure model constrains information structures to record the systolic, diastolic, and mean blood pressure measurement results, the patient's position, the measuring device, etc. Additionally, templates like those proposed by HL7 [14] or openEHR [15], allow using a set of clinical models together, constrained for addressing one or more particular documentation scenarios.

Ontologies formally describe properties and relations of types of entities. Domain-independent categories, relations and axioms are typically provided by *top-level ontologies* [10], whereas the types of things that make up a domain are represented by *domain ontologies*. In the former one we find categories like *Process*, *Material entity*, *Quality*, etc., whereas in a clinical domain ontology we would find classes for *Diabetes mellitus type 1*, *Left index finger*, or *Diclofenac*, ideally covering all the terms used in clinical documentation and reporting. The terms as such are organized by clinical terminologies. Ontologies have at least a minimal terminology component, consisting in a label or preferred term to make them understandable by humans. SNOMED CT [16], in addition, provides term variants and (quasi-)synonyms as possible values for data entry.

So far, there has been only partial and rather tacit consensus about the role each of the above artifacts should play and how they interface. Whereas terminology and ontology aspects are mostly covered by the same artifact, (by linking terms to ontology classes), the division between ontologies and information models ideally follows the classical distinction between ontology [17] (what exists) and epistemology (what is known) [18]. In practice, this line is often crossed both by ontologies (where they represent information entities, such as in the SNOMED CT context model) and by clinical models (where they carry their own ontology without reference to external standards).

As an example of overlapping, the SNOMED CT concept *History of vertigo* is described by following the SNOMED CT compositional syntax [19] as:

```
275543009 | H/O: vertigo | : {
 246090004 | associated finding | = 399153001 | vertigo |,
 408729009 | finding context | = 410515003 | known present |,
 408731000 | temporal context | = 410513005 | past |,
 408732007 | subject relationship context | = 410604004 | subject of record |}
```

This does not represent just the clinical type *Vertigo*, but also epistemic ("*Known present*") and temporal aspects ("*Past*") respectively. *Vertigo*, itself, is referred to by the code 399153001. This representation could also have been represented by a clinical model. Fig. 1 depicts an openEHR archetype excerpt for recording the medical history of some health issue. There, *Past* and *Known present* are embedded in the archetype information model structures, as free text labels. The health issue *Vertigo*, can be represented as free or coded text (DV_TEXT).

One could argue that ontologies just should not cross this boundary, such as by following the rules elaborated by TermInfo [20]. But even in this case the same piece of clinical information could be heterogeneously shaped by using different EHR information model structures. Within a diagnosis model, the disease and its location could be represented by a binding to a pre-coordinated SNOMED CT concept (e.g. *neoplasm of lung*); or alternatively, by structures that target disease (*neoplasm*) and location (*lung*) separately. Semantic interoperability requires means to detect that both representations are equivalent, also called iso-semantic (i.e. carrying the same meaning although heterogeneously represented).

The Clinical Information Modeling Initiative (CIMI) [21] seeks to address this problem by proposing a set of modeling patterns, defined as clinical models intended as guides for the creation of new ones. Each is associated with a set of iso-semantic models, from which one is selected as the preferred one and mappings are established across information model structures. CIMI or HL7 based models that implement the TermInfo specification might work well in isolation, but semantic interoperability issues arise when interacting with other modes, which are not necessarily compatible, whilst the anticipation of all possible iso-semantic representations would lead to an explosion of models.

MATERIALS AND METHODS

Overview

Convinced that multiple EHR standards, terminologies, ontologies as well as a large number of legacy information models will co-exist, our approach builds a semantic layer on top of them, acting as a proxy for applications that aim at accessing homogeneously content from heterogeneous repositories. This semantic layer consists of two sub-layers that include a set of semantic patterns and an ontological framework.

The ontological framework's objective is two-fold: Firstly, it provides a formal foundation that helps standardize information modeling tasks. Agreement at this level is essential to give interoperable meaning to heterogeneous EHR modeling styles. Secondly, the formal representation of clinical information conforming to an ontological framework adds value through description logic (DL) style reasoning [22] and advanced querying.

Semantic patterns, as close-to-user representations that hide the complexity of the ontology language, prevent modelers from the error-prone creation and maintenance of OWL expressions.

In order to access homogeneously clinical information of heterogeneous clinical models the following steps must be carried out: (i) identify semantic patterns to encode the information captured by clinical models and define their correspondences; (ii) instantiate the patterns with data; and (iii) create sample queries on the data.

Ontological framework

The ontological framework consists of three OWL DL ontologies: (i) SNOMED CT as domain ontology (ii) BioTopLite2 (BTL2) as top-level ontology, and (iii) an information entity ontology, which represents recurring elements of information as addressed by EHR information models.

BTL2 (prefix "btl") [23], is a reduced version of BioTop [24], a top-level domain ontology tailored for the biomedical domain. It provides upper-level types both for domain and information entities, as well as constraints on either, using a set of canonical relations, partly derived from the OBO Relation Ontology (RO). Axioms state disjointness between classes, constrain the domains and ranges of relations, define relation chains, as well as existential and value restrictions at the level of class definitions. Thus, ontology creation under BTL2 heavily constrains the freedom of the ontology engineer, which is intended, as it warrants a higher predictability of the ontologies produced [24].

SNOMED CT (prefix "sct") [16] acts as common reference point for representing the clinical content by modules based on its OWL version. Its content is largely harmonized with basic top-level classes and relations of BTL2. SNOMED CT' ontological commitment has been subject to intense debate in recent years and its consolidation is ongoing. Based on [26], we interpret SNOMED CT "Finding / Disorder") concepts as *clinical situations*. In addition, we have reinterpreted the SNOMED CT context model [27], in order to better distinguish clinical from information entities.

Finally, the EHR information entity ontology (prefix “shn”) represents pieces of clinical information as documented in the EHR. They are outcomes of actions like observations, investigations, or assessments. All of them refer to clinical entities and are further described by attributes that represent the epistemic and contextual aspects like clinical history, confirmed or suspected diagnosis. All classes of this ontology are represented as subclasses of the BTL2 top-level class *btl:InformationObject.* Information entities refer to clinical entities by means of the relations **shn:isAboutSituation** and **shn:isAboutQuality**, both defined as a specialization of the relation **btl:represents**. As subclass of *btl:InformationObject* we have placed the class *shn:InformationItem* for information entities that refer to some clinical situation; and the class *shn:ObservationResult* to refer to qualities that are directly or indirectly observed such as skin color or heart rate, and to which a value is assigned.

Additionally, the class *shn:InformationAttribute* has been created to represent parts of information entities that express their epistemic status, like *Suspected* or *Probable*.

Semantic patterns

Ontology content patterns are reusable solutions to recurring modeling problems [28]. They have fixed and variable parts [29], and are explicitly ontology-based, unlike databases schemas or UML models. In our context, named *Semantic Patterns*, they guide and standardize the meaning of the content of clinical models, and they bridge between approaches developed by distinct communities: EHR modeling [30], Semantic Web [31], formal ontology [32].

Fig. 2 depicts the semantic representation of an openEHR archetype for symptom information. The archetype consists of a set of nodes or data elements, which are constrained by value sets. White rectangles represent ontology classes, which are connected by directed arcs that represent quantified object properties. Such a representation is already a semantic pattern, with ontology classes as variable parts, since they can be specialized; and as fixed parts the classes to use and their interrelationships.

In order to obtain this representation, each node (e.g. CLUSTER, ELEMENT) of an archetype needs to be attached to the ontology according to two perspectives: (a) the information model class it corresponds to, and (b) the clinical entity it represents; the latter may require further sub-classing. It represents some quality (e.g. skin color), clinical situation (e.g. cancer), etc.

In Fig. 2, the archetype nodes (grey boxes) are placed next to their corresponding ontology top-level classes. At the core of the diagram, *shn:SymptomRecord* is the class of all information entities that represent some symptom, i.e. a clinical entity in the *shn:ClinicalSituation* class. For the rest of the archetype nodes, the same rationale is used. Some of the ontology classes are connected to *shn:SymptomRecord* and others to the *shn:ClinicalSituation* class, depending on whether they provide some epistemic information about the symptom or whether they further describe it. For instance, the archetype node labelled "Progression" is described as "The progress of the symptom relative to the past" and may take a value out of "Improving, Decreasing, Stable, Increasing, Worsening or Resolved". Since it reflects some perception about a symptom, it is epistemic information (*shn:Status*) and therefore connected with *shn:SymptomRecord*. How both classes are related within the ontology, is controlled by their internal axioms (here <**shn:hasInformationAttribute** some>). This assures the consistency of the pattern, which can be ascertained by DL reasoning.

Our hypothesis is that a limited set of generic patterns simplifies the modeling task. Patterns such as the one for the symptom archetype could be created by the specialization and composition of a set of top-level patterns, by following principles that are similar to the object-oriented paradigm [33]. Fig. 3 depicts the two top-level patterns from which the symptom pattern can be derived. Now, multiple archetype nodes correspond to one top-level ontology class. Thus, ontology classes and their corresponding relationships will be specialized in order to obtain the representation shown in Fig 2.

Use case description

Here we use an excerpt of the Heart Failure Summary, developed within SemanticHealthNet and represented as an openEHR template [34]. We demonstrate that information represented by clinical models can be expressed by semantic patterns or compositions thereof, using a set of finite top-level patterns from which the former ones are derived. Furthermore, we show how heterogeneous representations of the same or similar clinical information can be homogeneously queried.

Two test desktop applications for heart failure data recording (*Application A* and *B*) were implemented, based on different clinical models, which embed similar information into heterogeneous structures and a different level of detail. Each application consists of a set of constrained data elements bound to SNOMED CT terms. A tool developed within IHTSDO, which implements the SNOMED CT query language [35], defines reference sets from the terminology. Fig. 4 shows the entry forms that record symptom information.

Application A records the presence or absence of pre-defined symptoms by using a checklist, together with the heart failure stage by using a SNOMED CT subclass of *New York Heart Classification finding* (NYHA class). Application B provides a different set of symptoms for which their presence or absence can be indicated by selecting the SNOMED CT terms *Known present* and *Known absent* respectively. Additionally, it allows recording the severity of the symptom by selecting a severity value. The recording of the NYHA class is done in the same way as in Application A.

Comparing both applications, structural heterogeneities exist regarding the representation of the presence / absence of symptoms (check list vs. *known present / known absent* terms). Fig. 5 depicts an excerpt of their representation as ADL openEHR archetypes, where the symptoms' presence or absence is represented as an OpenEHR CLUSTER, which contains two ELEMENTs, one for the symptom itself (ELEMENT[at0001]) and one for representing its presence (ELEMENT[at0002]). There are also differences regarding the level of information detail required, in the additional recording of the severity by Application A (ELEMENT[at0005]).

RESULTSIdentification of semantic patterns and definition of correspondences

The nodes from the archetypes shown in Fig. 5 are mapped to semantic patterns. Since they record symptom information, we apply the pattern from Fig. 2, in order to represent the symptom only, its presence or absence, and its severity if indicated. Fig 6 shows the correspondences between the nodes of one of the archetypes and an excerpt of the symptom pattern.

Ontology classes (rectangles) represent the pattern's variable parts. The archetype node values are introduced as subclasses (dashed rectangles), corresponding to SNOMED CT classes. Table 1 shows the OWL DL rendering of the pattern. Negation is expressed by an additional OWL DL representation, which describes the absence of a symptom as a kind of symptom record about a patient's situation that does not include any symptom of that type. As described in Methods, this semantic pattern could have been obtained as specialization of the top-level pattern shown in Fig. 3. Both, the symptom severity and its presence or absence, are subclasses of the information attribute class (*shn:InformationAttribute*), used to represent epistemic and contextual information.

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 1. OWL DL representation of the semantic patterns for describing symptom's presence, absence, and severity. They are represented in Manchester syntax [36] and follow 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.

Instantiation of semantic patterns with patient data

When archetypes (already related to semantic patterns) are instantiated with patient data, a set of OWL DL conforming pattern instances are obtained. Table 2 shows the instances generated for the data recorded for two fictitious patients, PatientA and PatientB, by Application A and B respectively. Only instances of subclasses of *shn:InformationItem*, *btl:Process* and *shn:Patient* are created; their type is a logical expression conforming to the OWL DL symptom pattern representation (c.f. Table 1). The pattern variables have been replaced by the SNOMED CT terms, provided as patient data, and placed under the corresponding ontology classes (e.g. *sct:ChestPain* as subclass of *shn:ClinicalSituation*; *sct:Mild* as subclass of *shn:Severity*).

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*))

Table 2. OWL DL pattern instantiation with symptom related data captured by Applications A and B, for Patients A and B respectively. Other instance exemplars can be accessed here [37]

Homogeneous querying of pattern-based data instances

To query the patient data we have used SPARQL extended for the OWL Direct Semantics entailment regime (SPARQL-OWL) [38], which provides more expressive semantics than SPARQL's standard simple entailment [39]. We use this extension, implemented by the OWL-BGP API [40], which is independent of the DL reasoner used. Here we have used FaCT++ [41] and TrOWL [42] reasoners. A systematic evaluation of the reasoners that perform better was out of scope here. Table 3 depicts two SPARQL queries to retrieve the symptom-related information provided for Patients A and B. The queries follow the symptom pattern and the use of the reasoner allows their formulation at different granularity level compared to the provided at the data entry.

Query1: Information about patients with breathlessness

```
SELECT ?SymptomRecord
   WHERE {
     ?SymptomRecord a [
         a  owl:Class ;
         owl:intersectionOf (shn:InformationItem
         [a owl:Restriction ;
            owl:onProperty shn:isAboutSituation ;
            owl:allValuesFrom sct:SCT_267036007]
         [a owl:Restriction ;
            owl:onProperty btl2:isOutcomeOf ;
            owl:someValuesFrom sct:SCT_409060008]
   )]}
```

Answer: {SymptomA_Breathlessness_Present, SymptomB_BreathlessnessAtRest_Present}

Query2: Information about patients with breathlessness but not at rest

```
SELECT ?SymptomRecordDyspnea ?SymptomRecordNotRest
   WHERE {
      ?SymptomRecordDyspnea a [
         a  owl:Class ;
         owl:intersectionOf (shn:InformationItem
         [a owl:Restriction ;
            owl:onProperty shn:isAboutSituation ;
            owl:allValuesFrom sct:SCT_267036007]
         [a owl:Restriction ;
            owl:onProperty btl2:isOutcomeOf ;
            owl:hasValue ?EvaluationProcess])].
      ?SymptomRecordNotRest a [
         a  owl:Class ;
         owl:intersectionOf (
            [a owl:Class ;
            owl:intersectionOf (shn:InformationItem
                 [a owl:Restriction ;
                    owl:onProperty shn:hasInformationObjectAttribute ;
                   owl:someValuesFrom sct:SCT_410516002])]
            [a owl:Class ;
               owl:intersectionOf (shn:InformationItem
                   [a owl:Restriction ;
                      owl:onProperty shn:isAboutSituation ;
                      owl:allValuesFrom [ a owl:Class ;
                      owl:intersectionOf (shn:ClinicalSituation
                           [a owl:Class ;
                              owl:complementOf [a owl:Restriction ;
                              owl:onProperty btl2:hasPart ;
                              owl:someValuesFrom sct:SCT_161941007])]]
```

```
                    [a owl:Restriction ;
                       owl:onProperty btl2:isOutcomeOf ;
                       owl:hasValue ?EvaluationProcess])])]}
```

Answer:{SymptomB_BreathlessnessOnExertion_Present,SymptomB_BreathlessnessAtRestAbsent}

Table 3. SPARQL queries rendered using Turtle syntax [43]

Query1 retrieves instances of information about patients A and B although the patient data entered is not the same, since the SNOMED CT term *breathlessness on exertion* is a subclass of *breathlessness*.

Query 2 retrieves only data from patient B. Although the query asks for patients with breathlessness, it also explicitly asks for those which do not have breathlessness at rest, and this information has not been stated for patient A (the absence of information does not mean that the symptom is excluded).

Both queries use DL reasoning and the execution times using an Intel Core i5-3470 3.20 GHz, 8GB are: TrOWL Q1:1.183s, Q2:1.706s; FACT++ Q1:1.316s and Q2:3.053s.

DISCUSSION

We have used a Heart Failure Summary to demonstrate how semantic patterns can be applied to enable querying of heterogeneous representations of patient information.

OWL DL representations for the EHR have been proposed by several authors [44-47]. However, they are dependent of particular modeling approaches and therefore not interoperable. In contrast to our approach, most of them are limited to representing structural aspects of the clinical models and do not address their embedded meaning (e.g. "ELEMENT structure with allowed value CODED_TEXT" instead of "Diagnostic information about a disease"). We hypothesize that without any ontological commitment and formalization, the creation of ontologies adds just another complexity level to the EHR and is rather useless for interoperability.

In [5], A. Rector et al. distinguished "models of meaning" (describing our understanding of the world) from "models of use" (describing how data are displayed or captured). The latter ones are designed for specific use cases, while the former ones are largely stable and context-independent.

Semantic patterns provide certain structure but are not designed to allow several data capture possibilities at the point of care (e.g. drop-down menus vs. check lists). However, we have to admit plurality, because universal agreement on how to capturing information is not realistic. However, semantic interoperability requires that EHR systems share their "models of meaning". To this end, we have proposed semantic patterns as a bridge between heterogeneous EHR representations and a shared model of meaning. They also have the potential to ensure that more specific models such as proposed by CIMI are semantically valid derivations from higher level patterns. Further investigation is required as to which extent semantic interoperability can be achieved at the level of patterns, by using other representation language such as RDF and with limited or no DL reasoning at all. It is well known that description logics reasoning is not at zero cost and therefore may increase the query execution time beyond acceptable limits. This is the case of the OWL-BGP implementation. Optimization strategies are subject of current research [39] and some of them are implemented by this API, however the execution time might still be unacceptable for real-world implementations.

Some of the SNOMED CT concepts used in this work have been re-interpreted giving them a clearer rather than changed meaning. Issues could arise if the re-interpreted meaning differs from the one intended when the code was originally used. Variability in coding is an unavoidable problem, but clarification in naming (in)formal definitions will decrease inter-coder variability. Results from the SemanticHealthNet project have also been fed back to the SNOMED CT curators.

CONCLUSION

The semantic infrastructure proposed here addresses the complexity of the medical domain and their heterogeneous data capture and re-use needs by proposing a semantic layer on top of existing EHR representations, able to provide homogeneous access to heterogeneous datasets. They are heterogeneous not only because they use different representational languages, modeling approaches, etc., but also because they differ in context and granularity.

A set of description logics ontologies constitute the core of the proposed semantic layer. All of them adhere to formal ontology principles and exhibit reasoning capabilities. A top-level ontology enforces crisp boundaries between different entity types, which is important to keep the modelling process as standardized as possible. A balance is kept between what is ontologically correct and what is useful in practice.

The requirement of deep ontology engineering skills by those who have to model clinical information according to the proposed ontologies may be a severe obstacle. This has been addressed by simplifying semantic patterns, which help standardize the ontology-based modelling of clinical information, through their specialization and composition mechanisms. By looking at the existing content patterns available at the ontology pattern community site [48], we did not find specific patterns for the modelling of clinical information. Instead, patterns such as the agent-role or action ones could be reused. Whether there are a finite number of top-level patterns from which the others will specialize is still an open question. At this stage we can confirm that the representation of the Heart Failure Summary [4] provided a high degree of information heterogeneity and that a reduced number of top-level patterns were derived from that.

So far, we have based the patterns on an underlying OWL DL formalism. Their representation in RDF is subject of current work. The former one allows logical reasoning and therefore more advanced exploitation of information, although performance issues might limit their implementation in real systems. Here, RDF representations might be more appropriate, although less expressive and therefore more limited in terms of information exploitation,.

The use of patterns not only for interoperability purposes but also to guide the creation of clinical models and to detect semantic inconsistencies in the models is also subject of current and future research [49].

Besides, as in the case of existing EHR modeling approaches, an important success factor is to provide users with proper tools that isolate them from any technical detail. How to motivate industry partners to invest in such solutions is one of the biggest challenges.

Acknowledgments

We thank John Gutai, Accenture, for his collaboration in developing the data entry applications used.

Funding Statement

This work has been funded by the SemanticHealthNet Network of Excellence within the EU 7th Framework Program, Call: FP7-ICT-2011-7, agreement 288408, http://www.semantichealthnet.eu/.

Competing Interests Statement

The authors declare that they have no conflicts of interest in the research

Contributorship Statement

The work is the result of iterative discussions, modeling decisions, and the authoring of internal document within the SemanticHealthNet project, to which all authors equally contributed. Catalina Martínez-Costa, additionally, implemented and tested the approach with real data and led the editing of the manuscript.

References

1. Stroetmann VN, Kalra D, Lewalle P, et al. Interoperability for better health and safer healthcare. Deployment and research roadmap for Europe. 2009; ISBN-13:978-92-79-11139-6. doi:10.2759/38514
2. Semantic Interoperability for Health Network (SHN). http://www.semantichealthnet.eu/ (accessed September 2014).
3. W3C OWL working group. OWL 2 Web Ontology Language, Document Overview. W3C Recommendation 11 December 2012. http: //www.w3.org/TR/owl2-overview (accessed September 2014).
4. SemanticHealthNet Year 2, deliverable 4.2: Ontology/Information models covering the HF use case, 2013. http://www.semantichealthnet.eu/index.cfm/deliverables/ (accessed September 2014).
5. Rector A, Qamar R, Marley T. Binding Ontologies & Coding systems to Electronic Health Records and Messages, Appl Ontol 2009;4:51-69.
6. Kalra D, Lloyd D. EN13606 Electronic Health Record Communication Part 1: Reference Model. CEN TC/251, Brussels. February 2007.
7. Beale T, Lloyd D, Heard S, et al. (editors). The openEHR Reference Model version 1.0.2. http://www.openehr.org/programs/specification/releases/1.0.2 (accessed September 2014).
8. HL7 Reference Information Model (RIM). http://www.hl7.org/implement/standards/rim.cfm (accessed September 2014).
9. Beale T, Heard S. An ontology-based model of clinical information. Stud Health Technol Inform 2007; 129(Pt 1):760-764. PMID:179111819
10. Schulz S, Jansen L. Formal ontologies in biomedical knowledge representation. Yearbook of Medical Informatics 2013;8(1):132-46
11. Beale, T. Archetypes: Constraint-based domain models for future-proof information systems. Eleventh OOPSLA Workshop on Behavioral Semantics: Serving the Customer. Seattle, Washington, USA: Northeastern University; 2002:16–32.
12. Dolin RH, Alschuler L, Boyer S, et al. The HL7 Clinical Document Architecture, release 2. J Am Med Inform Assoc 2006;13:30-39
13. Coyle JF. *The Clinical Element Model Detailed Clinical Models*. Ph.D. Dissertation. 2013 University of Utah, Salt Lake City, UT, USA. Advisor(s) Stanley M. Huff. AAI3557073. http://gradworks.umi.com/35/57/3557073.html (accessed September 2014).
14. HL7 Implementation Guide for CDA® Release 2: IHE Health Story Consolidatio*n, Release 1.1 - US Realm* http://www.hl7.org/implement/standards/product_brief.cfm?product_id=258 (accessed September 2014).
15. OpenEHR templates. http://www.openehr.org/downloads/ADLworkbench/working_with_templates (accessed September 2014).
16. International Health Terminology Standards Development Organisation (IHTSDO). http://www.ihtsdo.org/ (accessed September 2014).
17. Quine WV. On what there is. Quintessence-Basic Readings from the Philosophy of W.V.Quine. Belknap Press, Cambridge 2004; Gibson, R. (ed.).
18. Bodenreider O, Smith B, Burgun A. The Ontology-Epistemology Divide: A Case Study in Medical Terminology. Third International Conference on Formal Ontology in Information Systems (FOIS 2004). IOS Press; 2004:185–95.
19. SNOMED CT Technical Implementation Guide. January 2014 International Release. http://ihtsdo.org/fileadmin/user_upload/doc/download/doc_TechnicalImplementationGuide_Current-en-US_INT_20140131.pdf (accessed September 2014).
20. HL7 TermInfo Project Wiki. Guidance on Overlap between RIM and SNOMED CT Semantics. http://wiki.hl7.org/index.php?title=TermInfo_Project# (accessed September 2014).
21. Clinical Information Modeling Initiative (CIMI). http://informatics.mayo.edu/CIMI/index.php/Main_Page (accessed September 2014)
22. Baader F, Calvanese D, McGuinness DL, et al. The Description Logic Handbook, Cambridge University Press, New York, NY; 2007
23. Schulz S, Boeker M. BioTopLite: An Upper Level Ontology for the Life SciencesEvolution, Design and Application. In GI-Jahrestagung 2013;1889-1899.
24. Schulz S, Boeker M. BioTopLite: An Upper Level Ontology for the Life Sciences. Evolution, Design and Application. *Informatik 2013*. U. Furbach, S. Staab; editors(s). IOS Press; 2013
25. Boeker M, Jansen L, Grewe N, et al. Effects of guideline-based training on the quality of formal ontologies: a randomized controlled trial. PLoS One. 2013 May 7;8(5):e61425. doi:10.1371/journal.pone.0061425. PubMed PMID: 23667440; PubMed Central PMCID: PMC3646875.
26. Schulz S, Rector A, Rodrigues JM, Spackman K. Competing interpretations of disorder codes in SNOMED CT and ICD. AMIA Annu Symp Proc. 2012;2012:819-27. Epub 2012 Nov 3. PubMed PMID: 23304356; PubMed Central PMCID: PMC3540515.

27. Martínez-Costa C; Schulz S. Ontology-based reinterpretation of the SNOMED CT context model. Proceedings of the 4th International Conference on Biomedical Ontology. CEUR Workshop Proceedings 2013; 1040:90-95.
28. 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.
29. 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 (accessed September 2014)
30. Kalra D. Electronic health record standards. 2006:136-144
31. Berners Lee T. WWW past & future. 2003; http://www.w3.org/2003/Talks/0922-rsoc-tbl/ (accessed September 2014).
32. Guarino N. Formal ontology in information systems. In Formal Ontology in Information Systems, N. Guarino (Ed.), IOS Press, Amsterdam, 1998:3-15.
33. Gangemi A. Ontology Design Patterns for Semantic Web Content. In Proceedings of the Fourth International Semantic Web Conference, 2005:262-276
34. The Heart Failure Summary template. Available at the Projects section of the OpenEHR Clinical Knowledge Manager http://www.openehr.org/ckm/. (accessed September 2014).
35. SNOMED CT Query Specification version 0. 8. http://www.ihtsdo.org/news/news-article/article/ihtsdo-public-consultation-on-an-ihtsdo-draft-standard-snomed-ct-query-specification-version-00/. (accessed September 2014).
36. Horridge M, Patel-Schneider PF. OWL 2 Web Ontology language: Manchester Syntax. 2009; http://www.w3.org/TR/owl2-manchester-syntax/. (accessed September 2014)
37. Ontology instance exemplars: https://www.dropbox.com/sh/8idp3d5fk65glul/AAAJMx3ujoOQqHDfGvpwtNewa?dl=0 (accessed September 2014)
38. Kollia I, Glimm B and Horrocks I. SPARQL Query Answering over OWL Ontologies. In Proceedings of the 8th Extended Semantic Web Conference on The Semantic Web. 2011:382-396
39. Kollia I, Glimm B. Optimizing Query Answering over OWL Ontologies. Journal of Artificial Intelligence Research 48(2013):253-303.
40. OWL-BGP SPARQL Wrapper. http://www.uni-ulm.de/en/in/ki/software/owl-bgp.html (accessed September 2014).
41. FaCT++ reasoner. http://code.google.com/p/factplusplus/ (accessed September 2014)
42. Thomas E, Pan JZ, Ren Y. TrOWL: Tractable OWL 2 Reasoning Infrastructure. In the Proc. of the Extended Semantic Web Conference (ESWC2010), 2010
43. Turtle – Terse RDF Triple Language. http://www.w3.org/TeamSubmission/turtle/ (accessed September 2014)
44. Lezcano L, Sicilia MA, Rodríguez-Solano C. Integrating reasoning and clinical archetypes using OWL ontologies and SWRL rules. J Biomed Inform 2011; Volume 44, Issue: 2, 343-353.
45. Menárguez-Tortosa M, Fernández-Breis J.T. OWL-based Reasoning Methods for Validating Archetypes. J Biomed Inform 2013; Volume 46. Issue: 2, 304-317.
46. Martínez-Costa C, Menárguez-Tortosa M, Fernández-Breis JT, et al. A model-driven approach for representing clinical archetypes for Semantic Web environments. J Biomed Inform. 2009; 42: 150–164
47. Tao C, Jiang GQ, Oniki, TA, et al. A semantic-web oriented representation of the clinical element model for secondary use of electronic health records data. J Am Med Inform Assoc 2012. Volume 20, Issue 3, 554-562
48. Ontology Design Patterns Semantic Web Portal. http://ontologydesignpatterns.org/ (accessed September 2014).
49. Martínez-Costa C, Karlsson D, Schulz S. Ontology Patterns for Clinical Information Modelling. Accepted at the Workshop on Ontology and Semantic Web Patterns (5th edition). http://ontologydesignpatterns.org/wiki/images/2/23/Paper_6.pdf (accessed September 2014).