Ontology-based Infrastructure for a Meaningful EHR Representation and Use

Martínez-Costa C., Legaz-García M.C., Schulz S., Fernández-Breis, J.T.

***Abstract*— Semantic interoperability of clinical information requires their unambiguous interpretation by clinical systems. Formal ontologies provide clinical information with a formal and principles-based representation. We describe a semantic infrastructure, which consists of an OWL DL ontological framework and a set of ontology content patterns. It aims at: (1) deal with heterogeneous representations of the same clinical information and (2) allow their advanced exploitation by using description logics reasoning.**

I. INTRODUCTION

Semantic interoperability of clinical information remains a challenge for the medical informatics community, addressed by several standards, specifications, initiatives and projects such as ISO 13606 [1], openEHR [2], HL7 [3], CIMI [4], SHN [5] or epSOS [6]. SemanticHealthNet (SHN) is an EU-funded network, which, in contrast to other proposals, does not intend to propose a new EHR standard. Instead, it provides an intermediate semantic layer able: (1) to deal with the unavoidable heterogeneity, which arises when clinical information is represented across or within the same domain and, (2), to exploit clinical information based on their richer ontological-based representation such as description logics reasoning.

SHN's semantic infrastructure is based on an ontological framework [7][8] and a set of ontology content patterns that use this framework as a reference. The framework consists of three kinds of ontologies: (1) top-level; (2) information entity and (3) medical domain, expressed in OWL 2 DL [9].

Ontology content patterns [10] can be seen as partial, frame-like views on this framework. They facilitate ontology building and maintenance, as they prevent the user from fully understanding the underlying, complex, formal expressions.

In this work we will exemplify how clinical information represented according two heterogeneous but similar drug administration models, developed in the projects SHN and epSOS, can be transformed into the ontology-based representation proposed by using ontology content patterns as "proxy". This requires establishing the correspondences between clinical model data elements and ontology content patterns. We have used the prototypical tool SWIT [11] in order to support this mapping process and to transform fictitious test data into the ontology-based representation.

Martinez-Costa C. and Schulz S. are affiliated to Medical University of Graz, 8036 Graz, Austria (corresponding author phone: +43-0316-385-17988; e-mail: catalina.martinez@medunigraz.at). Schulz S. is also affiliated to the University of Freiburg, Germany (e-mail: stefan.schulz@medunigraz.at).

Fernández-Breis J.T, and Legaz-Garcia M. C. are affiliated to University of Murcia, 30100 Murcia, Spain (e-mail: jfernand@um.es), (e-mail: mdclg3@um.es).

II. METHODS

A. SHN Semantic Infrastructure

The semantic infrastructure proposed consists of two main artefacts: an ontological framework and a set of ontology content patterns (Content OP). Fig. 1 depicts this schema and its relation with clinical models, which can be represented according to some EHR standard (e.g. openEHR), linked to a clinical domain ontology (e.g. SNOMED CT). In the following, each of these semantic artefacts is described.

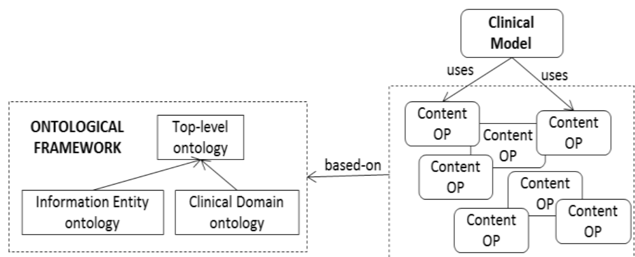

Figure 1. SemanticHealthNet (SHN) Semantic Infrastructure.

B. Ontological Framework

It consists of three ontologies:

1. A top-domain ontology, BioTopLite [12] (namespace prefix *btl:*). It provides a basic set of fundamental classes and relationships. Their use standardizes the modelling task by establishing a reference frame that restricts it by means of logical axioms grounded on formal ontological principles. BioTopLite exhibits some top classes such as *btl:Condition*, *btl:InformationObject*, *btl:Quality*, *btl:Role, btl:Process* or *btl:TemporalRegion*, by which all clinical and informational entities will be categorized.
2. A domain ontology, SNOMED CT [13] (namespace prefix sct:), with 311,000 concepts. It is partially built on formal-ontological principles, using the description logics EL++. Selected SNOMED CT content will be placed under top-level classes provided by BioTopLite.
3. An EHR information entity ontology (namespace prefix shn:) for representing pieces of information like diagnostic statements, plans, orders, etc. They are outcomes of clinical actions like observations, investigations, or evaluations. They are fundamentally different from clinical entities proper (like disorders, findings, substances), but they refer to them. Information entities

are further qualified by attributes that represent epistemic and contextual aspects of the information (e.g. past history, confirmed or suspected diagnosis, etc.). All classes of this ontology are represented as subclasses of the top-level class *btl:InformationObject*.

Information entities will refer to (types of) clinical entities by means of the relation **btl:represents** which can be further specialized by **shn:isAboutSituation** and **shn:isAboutQuality** for referring to a patient clinical situation [14] or a quality indirectly or directly observed of some material object or process.

C. Content patterns

An ontology content pattern is similar to a template used to represent and solve a software-modelling problem. They are based on a formal reference model and represent particular views on it, tailored to the needs of particular use cases [15]. Their main characteristics are [16]: (i) they can be organized in hierarchies, in which specializations are defined by following a similar paradigm to the object-oriented design, and (ii) their composition permit to cover larger modelling use cases.

We propose the use of ontology content patterns as a "proxy" which allows representing clinical information according to the ontology-based representation previously described. Although based on strict principles, they provide a higher-level representation that prevents users from a deep knowledge of ontology and description logics syntax.

Our assumption is that clinical models can be represented by the specialisation and composition of a set of ontology content patterns. In [17], we investigated the feasibility of using such patterns for representing some clinical information related to the heart failure, from which a set of ontology content patterns was created by following a bottom-up approach. We found out that they could be described by means of specialisation and composition based on a set of higher-level patterns (top-level patterns).

Here, we describe one of the patterns defined for representing medication administration related information and we show its application to two heterogeneous medication administration clinical models.

D. Medication administration pattern

The proposed medication administration pattern allows drug prescription either by referring to a pharmaceutical product and by its active ingredients, i.e. one or more active pharmacological substances contained in the product, in a certain amount, usually named strength, and in a certain form. The pattern allows encoding the following data items: the product administered; the active ingredients contained therein; the dose and form of the product; the strength and form of each active ingredient contained in the product; the starting date of administration; the ending date of administration; the duration of the treatment; the route of administration; and the frequency of the drug administration. For the latter, we only allow regular time patterns such as "twice a day" or "every two weeks", as provided by the SNOMED CT concept subclass hierarchy of 307430002 *Regular frequency*. We have not included more complex dosing instructions (e.g. double the drug intake each two weeks), as well as the drug indication (e.g. hypertension).

We have represented the pattern as a set of subject-predicate-object triples, cf. Table I. Note that the predicates are defined at the level of the pattern and are not taken from the source ontologies. In contrast to OWL DL object properties, these predicates directly connect classes.

TABLE I. MEDICATION ADMINISTRATION PATTERN TRIPLE-BASED REPRESENTATION

#	Subject	Predicate	Object
1	*btl:Plan*	isRealizedBy	*sct:Medication Administration*
2	*sct:MedicationAdministration*	hasFocusOn	*sct:Pharmaceutical Product*
3	*sct:MedicationAdministration*	hasRoute	*sct:RouteOf Administration*
4	*sct:MedicationAdministration*	hasStartTime	*btl:PointInTime*
5	*sct:MedicationAdministration*	hasEndTime	*btl:PointInTime*
6	*sct:MedicationAdministration*	hasDuration	*btl:Duration*
7	*sct:MedicationAdministration*	hasFrequency	*shn:Frequency*
8	*sct:PharmaceuticalProduct*	hasComponent	*sct:Substance*
9	*sct:PharmaceuticalProduct*	hasDose	*shn:PhysicalQuantity*
10	*sct:PharmaceuticalProduct*	hasForm	*shn:DrugDoseForm*
11	*sct:Substance*	hasStrength	*shn:PhysicalQuantity*
12	*sct:Substance*	hasForm	*shn:DrugDoseForm*
13	*shn:PhysicalQuantity*	hasValue	*xml:double*
14	*shn:PhysicalQuantity*	hasUnits	*shn:MeasurementUnits*

In the following we will describe each of the above predicates: *isRealizedBy* (1) indicates that there is a plan which is realized by a medication administration process; *hasFocusOn* (2) indicates the product and/or active ingredient to be administered; *hasRoute* (3) indicates the part of the body on which the product is to be introduced (e.g. oral, nasal, etc.); *hasStartTime* (4) and *hasEndTime* (5) describe when the drug use should start and end; *hasDuration* (6) describes the length of time for which the medicine has to be used; *hasFrequency* (7) describes the frequency of administration of the drug (e.g. twice a day); *hasComponent* (8) describes the active ingredient(s) (substance) that constitutes the product; *hasDose* (9) describes the amount of drug to be taken at any one time (e.g. 250 mg); *hasForm* (10,12) describes the physical form of a dose of drug / active ingredient (e.g. tablet, cream, etc.); *hasStrength* (11) describes the amount of the active ingredient (substance), contained in the product; *hasValue* (13) indicates the physical quantity of some material object (here product or substance) and *hasUnits* (14) the measurement units with respect to the quantity.

D. OWL DL representation

The representation of the medication administration pattern into OWL 2 DL allows: (1) the precise formalization of the ontological framework proposed and (2) the use of DL reasoning. DL reasoning is useful for the achievement of two important goals: On the one hand, it can be used for detecting equivalent clinical information from iso-semantic models (structurally different but with same meaning) [18]. It means to be able of comparing different distributions of content between information models and ontologies/terminologies, in order to test whether they are semantically equivalent. For instance, there are two possible representations to encode a breast cancer diagnosis when using SNOMED CT: (1) using one diagnosis information model element and the concept 254837009 *Breast cancer* or (2) using two information model elements for representing the disease diagnosed 363346000 *Cancer* and the disease location respectively 76752008 *Breast structure*. A DL reasoner should point out that both representations are semantically equivalent.

On the other hand, DL reasoning can be used to provide an advanced exploitation of clinical information by means of semantic query possibilities such as retrieving patients who were administered a drug that contains active ingredients of certain types, administrated via a certain route etc. [17].

Table II depicts the OWL DL representation of the medication administration pattern, according to the proposed ontological framework. By following the triple-based pattern representation shown in Table I, the subject (SUB) and object (OBJ) correspond to ontology classes and the predicate to an OWL DL expression.

TABLE II. OWL DL MEDICATION ADMINISTRATION PATTERN

Predicate	**OWL DL expression**
isRealizedBy	SUBJ subClassOf **btl:hasRealization** only OBJ
hasFocusOn	SUBJ subClassOf **btl:hasPatient** some OBJ
hasRoute	SUBJ subClassOf **btl:includes** some OBJ
hasStartTime	SUBJ subClassOf **btl:projectsOnto** some OBJ
hasEndTime	SUBJ subClassOf **btl:projectsOnto** some OBJ
hasDuration	SUBJ subClassOf **btl:projectsOnto** some OBJ
hasFrequency	SUBJ subClassOf **btl:isBearerOf** only (*shn:Frequency* and **btl:projectsOnto** only OBJ)
hasComponent	SUBJ subClassOf **btl:hasComponentPart** some OBJ
hasDose	SUBJ subClassOf **btl:isBearerOf** only (*shn:DrugDose* and **btl:isRepresentedBy** only OBJ)
hasForm	SUBJ subClassOf **btl:isBearerOf** only (*shn:DrugDoseForm* and **btl:projectsOnto** only OBJ)
hasStrength	SUBJ subClassOf **btl:isBearerOf** only (*shn:Strength* and **btl:isRepresentedBy** only OBJ)
hasForm	SUBJ subClassOf **btl:isBearerOf** only (*shn:DrugDoseForm* and **btl:projectsOnto** only OBJ)
hasValue	SUBJ **shn:hasValue** only OBJ
hasUnits	SUBJ **shn:hasInformationObjectAttribute** only OBJ

E. epSOS and SHN medication administration models

We will apply the above pattern to two models created in the context of the EU funded projects epSOS and SemanticHealthNet, to record a medication order.

The first model is part of the epSOS patient summary, which provides essential information for the general patient care. The SemanticHealthNet model is part of the heart failure summary, which provides essential information for the heart failure treatment. We have used the openEHR representation available in the project section of the Clinical Knowledge Manager (CKM) [19] for both models. Indeed, both are based on the same set of archetypes, but they are constrained differently, resulting in two slightly different templates. E.g., in the epSOS model, the focus of the medication order is the drug product, which can be refined, optionally, by active ingredients. In the SHN model the focus is on active ingredients. Another difference was related to the recording of the time the drug has to be used. It is recorded by means of the onset date and duration in epSOS and by means of the start and end date in SHN. Table III shows an excerpt of both models data elements.

TABLE III. DATA ELEMENTS EXCERPT EPSOS AND SHN MEDICATION ADMINISTRATION MODEL

epSOS	**SHN**
Product	Active ingredient
Pharmaceutical dose form	Ingredient dose form
Active ingredient	Ingredient strength
Ingredient form	Ingredient dose unit
Ingredient strength	Amount
Ingredient dose unit	Dose unit
Number of units per intake	Start date
Dose unit	Frequency
Date of onset of treatment	Stop date
Duration of treatment	
Frequency of intake	

III. RESULTS

We will use the medication administration pattern to represent, according to the ontological framework proposed, some fictitious clinical data rendered following the epSOS and SHN models. This requires establishing the correspondences between the data elements of each model and the medication administration pattern. We will apply these correspondences by using the SWIT tool in order to get the OWL DL representation of the clinical data.

A. Medication administration pattern and epSOS / SHN models correspondences

Each data element shown in Table III can be represented by means of a triple from the medication administration pattern. For instance, the active ingredient can be represented in both models with the triple number eight (*sct:PharmaceuticalProduct* hasComponent *sct:Substance*). Each row in Table IV and V provides the name of a data element in the epSOS and SHN models and the content pattern triple to which it corresponds respectively.

TABLE IV. MEDICATION ADMINISTRATION SHN / EPSOS MODELS AND PATTERN CORRESPONDENCES

SHN	**epSOS**	**Pattern Triple**
-	Product	2
-	Pharmaceutical dose form	10
Active ingredient	Active Ingredient	8
Ingredient dose form	Ingredient form	12
Ingredient strength	Ingredient strength	11+13
Ingredient dose unit	Ingredient dose unit	11+14
Amount	Number of units per intake	9+13
Dose unit	Dose unit	9+14
Start date	Date of onset of treatment	4
-	Duration of treatment	6
Frequency	Frequency of intake	7
Stop date	-	5

B. Application of the medication administration pattern to the SHN & epSOS models and OWL DL representation

We applied the correspondences shown in Table IV to both epSOS and SHN models, using the SWIT tool, which supports the mapping of archetypes to an OWL DL ontology, as well as the transformation of archetyped data into instances of that ontology (cf. Fig. 2).

Fig. 2 shows how the class *sct:Substance* from the ontology has been selected as mapping value for the active ingredient data element of the model. In the lower part of the same figure both the triple-based representation of the mapping (*sct:PharmaceuticalProduct hasComponent sct:Substance)* and how the data element mapped corresponds to the object part of the triple are shown.

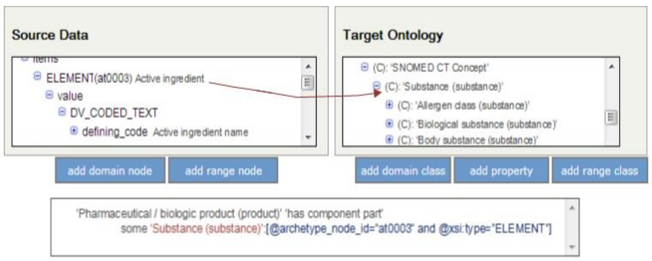

Figure 2. (Left) Excerpt of the openEHR SHN model used to record the active ingredient; (Right) Excerpt of the tree of the ontological framework

Once all the correspondences between the models and the medication administration pattern have been asserted in a process guided by the pattern, the tool allows for loading archetyped data and applying the previous correspondences, what would create instances of the target ontology. Next we show as an example the OWL DL rendering of the "2 mg tablet of bumetanide ordered" data instance:

Individual:**Order_product_Bumetanide**
Types:
btl:Plan and **btl:hasRealization** only (*sct:MedicationAdministration*
and **btl:hasPatient** some (*sct:Bumetanide*
and **btl:isBearerOf** some (*shn:DrugDoseForm*
and **btl:isRepresentedBy** only (*shn:PhysicalQuantity*
and **shn:hasValue** value "2"
and **shn:hasInformationObjectAttribute** some *sct:mg*))
and **btl:isBearerOf** some (*shn:DrugDoseForm*
and **btl:projectsOnto** some *sct:Tablet*)))

The above representation allows the advanced exploitation of the data by using DL reasoning. As an example, since bumetanide is a kind of loop diuretic drug, a query to retrieve all patients who were administered a loop diuretic drug would retrieve the above data instance.

C. Conclusions

The use of ontologies to represent clinical information in order to improve semantic interoperability was already mentioned in the SemanticHealth roadmap [20]. Thus, the main issue is now how to use them to capture the meaning of clinical information and avoid ambiguous representations. The proposed semantic infrastructure formally represents that meaning using OWL DL, thus constituting a semantic layer, which acts as a mediator across heterogeneous representations. Still in an experimental phase, SemanticHealthNet has to address scalability and performance challenges of the approach. Although it could be shown that semantic interoperability can be supported, the technological uptake of this approach will require a series of challenges (human, computational) to be met. We think that human challenges such as the ontology-based representation of present clinical information can be overcome by using semantic artefacts such as ontology content patterns, which might be implemented by tools such as SWIT. However, computational challenges in most cases require the evolution of present tools and resources, which might lead to agree on compromises between performance and functionality offered.

ACKNOWLEDGMENTS

This work has been funded by the *SemanticHealthNet Network of Excellence* within the EU 7th Framework Program, Call:FP7-ICT-2011-7, agreement no.:288408. MCLG is supported by the *Fundación Séneca* (15555/FPI/2010). JTFB is supported by the Spanish Ministry of Science and Innovation and the FEDER programme through grant TIN2010- 21388-C02-02.

REFERENCES

[1] ISO 13606. http://www.iso.org/iso/home.htm (Last accessed Jan. 2014)
[2] OpenEHR. An open domain-driven platform for developing flexible e-health systems. http://www.openehr.org (Last accessed Jan. 2014)
[3] Health Level Seven International. http://www.hl7.org/ (Last accessed Jan. 2014)
[4] Clinical Information Modeling Initiative (CIMI). http://informatics.mayo.edu/CIMI/ (Last accessed Jan. 2014).
[5] SemanticHealthNet Network of Excellence. http: //www.semantichealthnet.eu/ (Last accessed Jan. 2014)
[6] epSOS EU project: http://www.epsos.eu/home/about-epsos.html (Last accessed Jan. 2014)
[7] S. Schulz, L. Jansen, Formal ontologies in biomedical knowledge representation. Yearbook of Medical Informatics. 2013;8(1):132-46
[8] S. Schulz, C. Martinez-Costa, How Ontologies Can Improve Semantic Interoperability in Health Care. In Proc. KR4HC workshop, 2013.
[9] W3C OWL working group. OWL 2 Web Ontology Language, Document Overview. W3C Recommendation 11 December 2012. http: //www.w3.org/TR/owl2-overview (Last accessed Jan. 2014)
[10] A. Gangemi, V. Presutti, Content Ontology Design Patterns As Practical Building Blocks for Web Ontologies. In Proc. of the 27th International Conference on Conceptual Modeling, 2008, pp. 128-141
[11] SWIT tool: sele.inf.um.es/swit (Last accessed Jan. 2014)
[12] BioTopLite: http://purl.org/biotop/biotoplite.owl. (Last accessed Jan. 2014)
[13] Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT), 2008. http: //www.ihtsdo.org/snomed-ct (Last accessed Jan. 2014)
[14] S. Schulz, A. Rector, J. Rodrigues, C. Chute, B. Üstün, K. Spackman, Ontology-based convergence of medical terminologies. SNOMED CT and ICD-11. Proc. of eHealth2012. Vienna, Austria: OCG, 2012.
[15] E. Blomqvist, E. Daga, A. Gangemi, V. Presutti, Modelling and using ontology design patterns. [http://www.neon-project.org/web-content/media/book-chapters/Chapter-12.pdf]
[16] A. Gangemi , Ontology Design Patterns for Semantic Web Content. In Proceedings of the Fourth International Semantic Web Conference, 2005, pp. 262-276
[17] SemanticHealthNet deliverable 4.2: Ontology / Information models covering the heart failure use case, 2013.
[18] C. Martínez Costa, D. Bosca. M.C. Legaz-García, C. Tao, J.T. Fernández-Breis, S. Schulz, C.G. Chute, Isosemantic rendering of clinical information using formal ontologies and RDF. In Proc. MEDINFO 2013. Stud Health Technol Inform. 2013;192:1085
[19] Clinical Knowledge Manager. http://www.openehr.org/ckm/ (Last accessed Jan. 2014)
[20] V.M. Stroetmann et al., Interoperability for better health and safer healthcare. Deployment and research roadmap for Europe. ISBN-13 : 978-92-79-11139-6