The Patient Summary Case: Challenges in Archetypes Terminology Binding Using SNOMED-CT Compositional Grammar

Luis Marco-Ruiz
Norwegian Centre for e-healthresearch
Tromsø, Norway
Luis.marco.ruiz@ehealthresearch.no

Rune Pedersen
Norwegian Centre for e-healthresearch
Tromsø, Norway
Rune.pedersen@ehealthresearch.no

***Abstract*— In order to cover the requirements of interoperability in the Norwegian context, we studied the adequacy of expressing the clinical semantics contained in archetypes as terminology expressions using SNOMED-CT´s compositional grammar. As a result, we identified important challenges categorized as technical, expressivity, human, and models mismatch related. Technical challenges include the binding of archetype elements to sections of the SNOMED-CT expressions that are semantically equivalent, lack of tooling for performing guided binding based on pre-defined semantic patterns, and lack of guidance about the infrastructure to use ontology-based terminologies. Expressivity challenges include variations in the precision of the semantics expressed by the archetype and the SNOMED-CT models, challenges expressing temporal semantics in SNOMED-CT, and lack of mechanisms for specifying expressions whose values are only known at runtime. Human challenges include lack of guidance to discern what to leave represented at archetype level and what to project as a terminology expression depending on the expressivity requirements, and the need for more clarity about which hierarchies and attributes to use when several options are available for expressing the same concept. Model mismatch issues were related to the issue of grounding (referencing) the sections of one model to the other and clarify the role of the context model and in which situations it makes sense to annotate archetypes using its verbose expressions. The challenges detected show a pressing need for the collaboration between the openEHR community and SNOMED International for providing better guidelines about terminology binding, better tooling for facilitating the binding process, and developing mechanisms that allow for extending SNOMED-CT with other biomedical ontologies in order to increase the coverage of archetypes semantics for scenarios with high expressivity requirements such as data reuse ones.**

Keywords-openEHR; SNOMED-CT; archetype; terminology binding; biomedical ontology.

I. INTRODUCTION

Over more than a decade, health authorities worldwide have faced a growing pressure to accomplish smooth information flow between systems of health organizations. This has led to the adoption of clinical information standards and their use in combination with terminologies. Terminologies are used both to specify the meaning of the sections inside clinical information models (semantic binding), and to specify the content of the information recorded inside these sections (content binding). However, the effort in annotating clinical information models with terminologies is often underestimated [1]. In fact, few guidelines are available with clear instructions on how terminologies should be used in combination with clinical information models [2]. For example, when one faces the terminology binding of an openEHR archetype, it is not clear what elements of the archetype should be coded and which should not; when to use a pre- or a post-coordinated expression; or how to proceed when several terminology concepts are available. In fact, terminology binding guidelines often focus on the annotation of just the main sections of clinical documents, but they do not concern aspects related to the consequences of choosing an ontology-based terminology, the granularity that should be pursued when annotating archetypes, or the requirements that advanced phenotyping queries used in clinical research may pose in the terminology binding of archetypes [1][2]. These are issues that need to be carefully addressed since they will have a major impact on the interoperability of clinical data extracts, thus determining the limitations of eHealth infrastructures. For example, the needs for harmonizing primary care and secondary care health information if only openEHR is used differ from the needs when SNOMED-CT is also used as reference terminology. One could only use SNOMED-CT for semantic tagging, i.e., indicate the meaning of each section in a clinical document (e.g., “this section contains vital signs”). Or one could decide to use the terminology to actually specify content abstracting the information shared from the syntactic representation format of one of the systems. For example, an archetype measurement of 155 mmHg systolic blood pressure could be shared just as the SNOMED-CT concept "On examination - blood pressure reading very high (finding) 163028000". Also, depending on the type of terminology selected, it may be possible to perform filtering operations based on its structure and hierarchy (e.g. which patients in treatment with anticoagulants suffered from a major bleeding event last year?).

Adopting openEHR in combination with SNOMED-CT raises questions such as: which sections of archetypes should be bound to the reference terminology? Is there any benefit in fine-grained terminology binding of archetypes items? Should we use SNOMED-CT as a vocabulary or do we need to deal with its formal ontological nature to maximize the benefits?

These research questions have been a matter of discussion among the Norwegian Centre for E-health Research, the Norwegian Directorate of eHealth (sub-ordinate institution of our Ministry of Health and Care

Services), and the national and regional comities for standardization (Nasjonal IKT). In addition, the Norwegian Agency for eHealth considers the adoption of a common reference terminology a long-term commitment.

In the Norwegian context there has been a growing interest about the implications derived from ontology-based terminologies (e.g. SNOMED-CT) and the impact on present and future health information systems. As a result, the Directorate of eHealth decided the enrollment in SNOMED International and advised the acquisition and use of SNOMED CT at national level for a three-year period.

In November 2015 several projects started exploring the mapping of SNOMED-CT towards the most commonly used Electronic Medical Record (EMR) functionalities (e.g., critical information, clinical findings, the patient pathway for chronically ill patients, and the patient summary). Among the suggestions of the directorate is the elicitation of clinical information models (archetypes) and terminology value sets for building the patient summary for continuity of care.

A. The national work with openEHR archetypes in Norway

Throughout the last four years OpenEHR has grown to gain a national anchorage in Norwegian healthcare. From 2008 to 2012, the National ICT Health Trust conducted projects using archetypes for clinical chart systems, which resulted in a recommendation to build an openEHR-based infrastructure for specialized healthcare. In 2013 the National ICT Health Trust, decided to establish the National Editorial Group for Archetypes (NRUA) for coordinating the development of the national repository of clinical models (archetypes). NRUA also established a close collaboration with the openEHR foundation and the international openEHR Clinical Knowledge Manager (CKM). Through this collaboration, NRUA has translated and reviewed existing archetypes from the international CKM into the Norwegian national CKM [27]. Currently, the Norwegian CKM includes circa 400 reviewers of which around 30% are clinicians. The review process is being conducted at a national level [8] with the collaboration of contributors from all health regions. The number of archetypes approved in Norway in October 2018 was 80, and over 100 more are in process. The most immediate goal is to have 200 nationally approved archetypes by the end of 2019 to represent the EMR core content.

B. Archetypes binding to SNOMED-CT

Archetypes terminology binding is a complex, time-consuming task. Such task is even more challenging when the terminology used is an ontology-based terminology such as SNOMED-CT. In that case, both the semantics contained in the archetype and the terminology conceptual models need to be carefully analyzed. For example, when an archetype element is bound to a terminology code, clinical modelers need to consider not only the name of the code but also the hierarchy of SNOMED-CT that contains that code to determine if it is appropriate to use it or not. For example, concepts to represent the blood pressure are available in the Observable Entity and Clinical Finding hierarchies. The different choices available when binding archetypes to terminologies need to be carefully addressed since they will impact interoperability [5]. Actually, several levels of interoperability are possible: syntactic interoperability using common data schemas, partial semantic interoperability (SIOp) where some sections of clinical statements are bound to terminologies, or full SIOp based on machine-understandable models [2]. Therefore, it is important to determine several factors to design a roadmap towards interoperability. The first factor is to determine why interoperability is desired, i.e., what are the requirements that need interoperable clinical data. The second factor is to determine what level of interoperability is needed to cover those requirements. The third factor is to determine what are the tools needed to grant the level of interoperability desired. These determinants may vary from one application context to another, and from one country to another. In the case of Norway the interoperability of clinical data pursues three main objectives:

- •Develop semantically enriched clinical information models that implementers can use to drive the development of health information systems.
- •Facilitate mappings across different information models based on the archetypes provided.
- •Facilitate the secondary use of clinical data by providing a standard format for clinical information representation and expressive queries based on standard terminologies.

In most cases the approach followed to bind archetypes to terminologies relies on assigning a terminology code to the main sections of the archetype. For example, a SNOMED-CT binding allows for specifying that the number set in the element at0078.13 of the archetype openEHR-EHR-OBSERVATION.lab_test-full_blood_count.v1 refers to the SNOMED-CT concept 250271003 /White blood cell finding. This approach endows the archetype with certain level of semantics by attaching a code that refers to a concept in the terminology. However, these semantics correspond to concepts framed within an external ontology (i.e. SNOMED-CT concept model) and may not suffice to cover the requirements previously described.

In the Norwegian case, the first requirement described (i.e. provide semantically enriched clinical information models) can be covered with that approach, however the second and third requirements cannot. When it comes to facilitating mapping tasks across different information models, the approach of binding independent terminology codes has been related to SIOp challenges. For example, Dixon et al. documented problems when establishing SIOp across organizational boundaries caused by issues interpreting the meaning of the terminology codes even when the same terminology was used by all organizations [9]. Clinical information models define a syntactic schema with rich data constraints definitions. However, the information model does not specify the semantics of concepts and relationships explicitly beyond the description associated to the code. When SNOMED-CT is used, the ontology that unambiguously defines the semantics of the concept is detached and external to the conceptual model that the archetype conveys. This also affects the third requirement since secondary use of clinical data requires running expressive queries. These queries require the machine-interpretation of semantic relations such as subsumptive or equivalence ones to manage the complexity of clinical information. Markwell documented the problems of interdependencies among the semantics of the archetype elements when they are annotated using independent terminology concepts [1]. For example, let us consider the archetype family_history whose root element is coded as 57177007 /Family history with explicit context and one of the internal nodes is coded as 73211009/Diabetes mellitus. The intended semantics are that a family member of the patient had diabetes. However, that is not unambiguously described following any formal model. Therefore, if a query returns the diabetes value one may interpret that the patient has the disease. This mismatch may cause wrong inferences and erroneous (or at least confusing) queries results if both the archetype and the terminology bindings are not analyzed carefully [10]. Another problem related to this mismatch between models is not purely technical but related to the human interpretation. Sometimes organizations using the same terminology decide to use different codes to identify the same concept, thus hampering SIOp [5][9].

Aiming to overcome these challenges and to cover the requirements set by NRUA, we determined that we needed a machine-interpretable representation of the clinical semantics carried by archetypes. We considered that the full expression of archetypes (including data constraints) as ontologies is not scalable due to computational restrictions [11], nor necessary to accomplish the requirements presented. Therefore, we opted for using SNOMED-CT compositional grammar to build semantic models that distilled the clinical semantics contained in archetypes (leaving data constraint aside). This way, each of the elements of the archetype can be represented within an ontology that provides unambiguous semantics and reasoning.

Previous studies have covered these issues for a specific set of clinical models [1], [3]-[5]. However, these studies focused mainly on other standards than openEHR and they did not consider the expressivity requirements that data reuse scenarios introduce. Many of the scenarios for data reuse and decision support requires a more complete terminological projection in order to ensure that the meaning of clinical information is preserved across systems. That is actually the case of several projects in Norway that aim to enable data reuse [6][7]. The objective of this paper is to report about the main challenges found in our evaluation for introducing SNOMED-CT in combination with the openEHR-based infrastructure for fulfilling the requirements of the Norwegian context.

II. METHOD

In order to specify the archetype clinical semantics as an unambiguous ontology model, the SNOMED-CT compositional grammar was used. Together with NRUA we selected the most representative archetypes of each of the sections contained in the Norwegian patient summary. The archetypes selected are archetypes reviewed nationally and published enabling interoperability at a national level. The sections of the patient summary correspond to those specified in epSOS [28]. Archetypes were downloaded from the Norwegian CKM and their estate was either published or in review. For each archetype, terminology binding was attempted by creating a projection of the archetype clinical semantics using the SNOMED-CT compositional grammar. We tried to maximize the coverage of the elements represented. The binding was performed from the root to the leaves. When some element/section of the archetype could not be represented using a SNOMED-CT expression for the whole archetype, we defined a new expression and tagged the reason for using several expressions to represent one archetype. Additionally, when a blocker caused by the complexity of the process, lack of tooling etc. was found, we tagged it. When we had doubts about the specific meaning of an archetype element we asked NRUA for clarification. Afterwards we reviewed all the tags and classified them into categories of challenges. Table 1 contains the set of archetypes used in the study.

TABLE I. SET OF ARCHETYPES COVERED.

Archetypes analyzed
openEHR-EHR-OBSERVATION.blood_pressure.v1
openEHR-EHR-EVALUATION.tobacco_smoking_summary.v0
openEHR-EHR-EVALUATION.immunisation_summary.v0
openEHR-EHR-EVALUATION.problem_diagnosis.v1
openEHR-EHR- openEHR-EHR-EVALUATION.adverse_reaction_risk.v1
openEHR-EHR-INSTRUCTION.medication_order.v0

III. RESULTS

After reviewing all tags four categories of challenges were identified: 1) technical, 2) SNOMED-CT expressivity related, 3) human, and 4) mismatch between models.

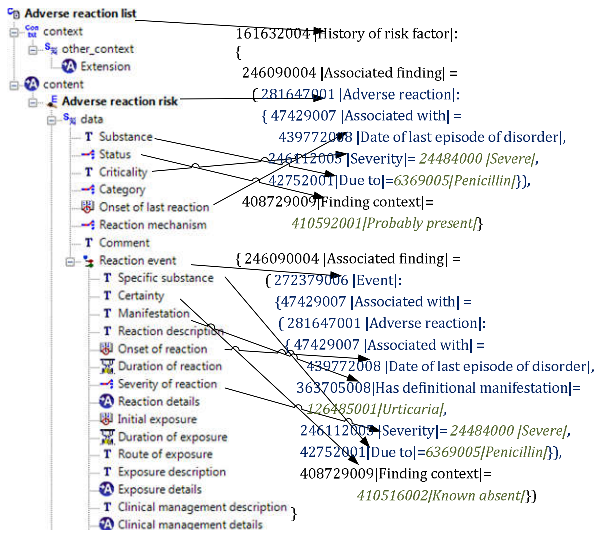

Figure 1. Adverse reaction archetypes and their projection into SNOMED-CT.

A. Technical challenges

Technical challenges are associated to limitations or barriers found when using current technologies. The first technical challenge is associated with limitations to link the archetype and the SNOMED-CT expression due to a lack of available tooling. The main technical challenge found was the lack of support of archetypes to include verbose post-coordinated expressions in their term_bindings section. This could be solved by using proper URLs pointing to an external server that contains the expression provided that URIs have been recently accepted as the best approach for annotating archetypes. However, in that case, another limitation appears related to SNOMED-CT tooling. We did not find any mechanism to reference different sections of the expression. Therefore if the entire archetype is linked to a terminology expression that represents its meaning at an ontology level, it is not possible to specify to which concept or attribute of the expression the archetype element refers to. For example, in Figure 1 black arrows shows the need for establishing references from the archetype adverse_reaction_risk elements to the parts of the expression on the right. Other studies have proposed the use of linked data principles and W3C standards for referencing different ontology sections [12][13]. However, appropriate tooling to make that process transparent to clinical modelers is necessary. These referencing mechanisms are something that, to our knowledge, have not been covered in the medical informatics literature and, in many cases, have been considered too complex to achieve. However, it is a common and necessary operation named "grounding" in semantic web programming [12]-[14]. Such operation is necessary for establishing mappings and transformation mechanisms between the syntactic and ontological layers in semantic web applications. In fact, recent projects such as the Yosemite Project [15] and HL7 FHIR RDF [16] are pursuing the serialization of information models into RDF(S) for facilitating clinical information mapping, retrieval, and management.

The second limitation is related to the lack of tooling to assist the binding process. The SNOMED-CT compositional grammar contains a huge variety of attributes and hierarchies. The correct specification of semantics depends on choosing the correct hierarchy and attribute. However, at the moment, this is an extremely difficult task since one needs to check continuously the compositional grammar and the semantics of each attribute. Tooling to support this process by offering guidance in choosing the appropriate elements/attributes is needed. Those tools should guide clinical modelers not only by restricting the set of values that can be assigned to an attribute, but also by informing about the exact meaning and provide examples during the coding workflow.

The last technical challenge identified is the information about the infrastructure to process SNOMED-CT expressions. Although some examples of reasoners are mentioned, exploiting the expressivity of SNOMED-CT in large deployments may impose high demands on the hardware infrastructure required. Guidelines for software architects should be provided to allow them planning enterprise deployments.

B. Expressivity challenges

Challenges in expressivity relate to those concepts which semantics cannot be fully expressed with the compositional grammar. Some occur because the archetype element coded has a candidate in SNOMED-CT that introduces a slight variation in the semantics originally intended in the archetype. That is the case of "reaction mechanism" (archetype element) vs. "Immune hypersensitivity reaction by mechanism (disorder)" (SNOMED-CT concept). This problem has been approached elsewhere by extending SNOMED-CT [14]; however, since nationally approved archetypes intend to be as generic as possible, divergences between the SNOMED-CT implementation used for coding and the standard release should be avoided.

Sometimes the variation in semantics occurred as a consequence of relying on a SNOMED-CT attribute that has a broader meaning. An example is displayed in Figure 1 where the element substance can be mapped to the attribute "due to" of the SNOMED-CT expression, which has a wider range than substances (any ClinicalFinding; any Procedure etc.) Another loss of meaning occurs in the example where the archetype element "onset of reaction", with an openEHR data type Date/time, acquires the meaning "date and time of the onset of reaction". However, the SNOMED-CT candidate for matching is "Date of last episode". One should note that it is unknown if it refers to the onset, end etc. Other expressivity problems occur for not being able to express contextual semantics. The SNOMED-CT context model does not allow expressing temporal orders. This was expected since archetypes typically provide much more contextual information (epistemology) than biomedical ontologies [18]. This can be seen in Figure 1 where the section Reaction event (with meaning previous adverse events) is mapped the expression section with the Event that wraps the finding adverse reaction. We did not find the way to specify this kind of semantics that often occur in patient summaries of drugs, allergies etc. The time context SNOMED-CT attribute allows expressing if the situation occurred in the present or past but not order of events. Besides, in the compositional grammar, the time context attribute cannot be used as a valid attribute of an event. Other type of expressivity problems are those related to the restrictions that the compositional grammar imposes. For some nodes there exists a valid candidate in SNOMED-CT. However, it is not possible to include it using the compositional grammar or the expression needs to be overcomplexified for including it. In some cases, we found elements that cannot be coded due to a total lack of candidates available. Examples are "Duration of exposure" or "initial exposure" element in Figure 1. This is usual in concepts whose semantics are purely contextual such as time related properties.

Other expressivity challenge concerns the need for leaving the terminology projection incomplete. Marked with green italics in Figure 1. This situation is common when the expression should contain a Qualifier Value after an attribute. In Figure 1 it is seen how the substance that causes the adverse reaction should map to the attribute “due to” that may contain a substance such as penicillin. However, this value is unknown at archetype design time and will only be determined in runtime for each instance. In order to overcome these situations a mechanism to indicate that the value is available at instance level would be needed. Although SNOMED-CT is not intended to manage instances, it would be appropriate to provide guidelines for determining how expressions that have a dependence on the archetype instance values should be managed. Finally, we detected problems related to differences between the semantics of the conceptual model carried by the archetype and the semantics that can be expressed with SNOMED-CT. This becomes evident in the archetype openEHR-EHR-EVALUATION.problem_diagnosis.v1 that uses the same clinical model for expressing both the Problem (Disorder) and the Diagnosis, whereas in SNOMED-CT these are separated concepts. A solution for this could be to use the compositional grammar to specify a concept that inherits from two similar concepts such as “Clinical finding+Disorder” (or in this case clinical finding only). However, the archetype semantics have a wider meaning that includes concepts such as Diagnosis (439401001) that may belong to a different hierarchy in SNOMED-CT. In that case it is not possible to build concepts such as “Clinical finding+Diagnosis” since the compositional grammar discourages the combination of concepts that come from different hierarchies [19].

C. Human challenges

Human challenges relate to the difficulties found by clinical modelers inherent to the complexity of the process. First, we had many doubts in determining what to represent in the SNOMED-CT expression and what to leave unrepresented. Although we had determined that we should aim for a maximum representativeness to grant SIOp and enable expressive queries for clinical research, it became clear that we had not specified use cases with the appropriate level of detail to determine what sections should be represented with the terminology expression. To avoid this challenge it is needed to have a clear set of representative use cases related to the scenarios that will exploit the terminology. We acknowledge that anticipating that use is somehow idealistic at a national level. However, a set of use cases related to each of the requirements explained would help to narrow down the binding task.

The second human challenge is related to the one just described and is the lack of guidance about what can be expressed in the terminology; i.e. guidance on what to represent. This has been already reported by studies using independent codes for the sections of the clinical model [1]. However, it is needed that both SNOMED-CT implementers and archetype editors define general directives on what will be the minimum set to code and provide guidelines with examples for that. This is extremely important when the SNOMED-CT context model is used since it opens the door to terminology projections of a high percentage of the information model elements that previously were maintained only at the syntactic level. Finally, another aspect linked to the need of guidance is how to choose the appropriate attributes. Previously it was mentioned the need of tooling, but also guidelines are needed. An example is found when deciding on expressing the blood pressure using the Observable Entity hierarchy or the Clinical Finding hierarchy. Some studies have analyzed this issue [5], but this knowledge needs to be translated into pragmatic terminology binding guidelines.

D. Model mismatch

This category identifies the challenges related to structural differences between the archetype and the expression created with the compositional grammar. A first structural challenge comes from the fact that the context model sets many of the contextual information wrapping clinical statements. However, in the archetype this is in many cases the opposite complicating to establish equivalences between the archetype and the SNOMED-CT representation of the clinical concept (i.e. grounding operation). Again SNOMED-CT and archetype editors should provide guidelines about the alignment of both models and the way of referencing one to the other. Finally, we found that the attributes of the protocol section had a very low coverage. Examples appear, for example, in the blood_pressure archetype for cuff size or systolic pressure formula. This was expected since they refer mainly to contextual information, but we found that it is necessary to clarify the use of SNOMED-CT context model with regards to clinical information models in general, and archetypes in particular.

IV. DISCUSSION

The long term needs of semantic interoperability across EHRs and secondary use of data in the Norwegian context need the representation of archetype semantics as expressions that allow for defining them unambiguously and performing expressive queries over data sets. Terminology binding of only some archetype sections does not provide that level of expressivity. Expressing them by using the ontology of the terminology becomes necessary. We have coded terminological expressions using SNOMED-CT´s compositional grammar using the archetypes that shape the Norwegian patient summary. Technical, expressivity, human and models mismatch challenges have been identified. The first challenge found is that currently SNOMED-CT expressions cannot be referenced from archetypes. On the one hand, it is not possible to include long post-coordinated expressions. Thus, the use of URLs to reference the expression should be recommended. This is a recommendation that has currently be taken into the openEHR specifications. The second challenge found is the lack of tooling available to guide the definition of expressions, not only validating post-coordinated expressions, but also assisting modelers while using the compositional grammar. In many cases, the need for continuously checking SNOMED-CT guidelines made us lose track of our own work. Nowadays, the availability of published archetypes is growing [6]. Tools that define general use cases to represent different types of archetypes would be very useful. For example, there are tools to guide clinical modelers in coding archetypes with graphical representations and powerful matching techniques [20]. This kind of tools may consider incorporating the functionality to guide users for building expressions using SNOMED-CT´s compositional grammar. Another challenge is related to the expressivity of SNOMED-CT. For simple archetypes that relate to readings such as blood pressure, it is possible to create terminology projections of archetypes´ semantics using the compositional grammar. However, for more complex archetypes such as the adverse_reaction or medication_order ones, the context model does not allow defining the relations that are needed to express them at a semantic level. This problem varied from issues related to the inability to include a concept in the compositional grammar, loss of semantics due to the variation of meaning and, in the worst case, not been able to express relations between archetype sections. This may influence the patterns that archetypes follow. We are aware that previous experiences show that the best way to approach interoperability is to annotate just some sections of the clinical information models (archetypes) [3]. However, for the national interoperability program it is important to determine until what extent the expressivity of SNOMED-CT can be used to define the semantics covered by archetypes and the implications of adopting ontology-based terminologies. This was part of the national project to study the use of formal ontologies for healthcare and the impact and possibilities regarding their adoption [2]. In that work we determined that there is not only one way to approach this challenge since determining the level of semantic enrichment of archetypes depends highly on each use case. For EHR interoperability it may suffice binding just the main sessions of the EHR, whereas for performing deep phenotyping [21], it would be necessary to define fine-grained annotations with several biomedical ontologies [2].

Both terminology and archetype experts should agree on a minimal set of recommendations for different purposes and agree on some design patterns for terminology binding that facilitate the scalability of semantic infrastructures for healthcare and research. This scalability principle should allow for starting with basic semantic annotations and progressing towards more complex semantic infrastructures on “as-needed” basis as recommended in other domains[22]-[24]. At the moment, when we reviewed the guidelines for SNOMED-CT adoption and compared them with the development of archetypes we detected an important lack of coordination between the guidelines from the archetype design point of view and the terminology point of view (e.g., SNOMED-CT guidelines). For example, the SNOMED-CT context model clearly overlaps with the function of the archetype but does not have enough expressivity to allow for projecting archetypes at a semantic level. Both openEHR and SNOMED International should agree on a minimal set of patterns as mentioned above. We do understand that the compositional grammar (and expressions like those created using the context model) may be necessary for some scenarios, but then they need to provide means for aligning SNOMED-CT expressions with archetypes. In order to enable a more complete representation of the semantics included in the archetypes at a terminology level, a more complete context model would be necessary. For example, approximately half of the concepts in the protocol section of archetypes could not be represented. However, extending the context model would lead to a mix of ontological and epistemological aspects. Even when both aspects are represented at a semantic level, it is appropriate to keep the models of meaning and information separated to enable separate scalability and reasoning [25]. An option to avoid mixing different models may be to maintain a context model ontology outside the medical terminology (SNOMED-CT). Approaches to develop parallel context models have been defined by relying on more expressive logics (OWL DL) [26]. However, these kind of logic leads to models that are not tractable [11], thus they may jeopardize large enterprise deployments. Experiences in Semantic Web development have shown that in many cases it is better to sacrifice expressiveness in order to gain scalability [12][23]. If those context models evolve into lighter models and define clear extension mechanisms to use other ontologies, they may be very useful to provide machine-interpretable representations of archetypes clinical semantics.

Finally, implementers must be aware that projecting most of the archetype as a formal ontology is expensive in terms of computational demands and models definition. Therefore, as mentioned before, semantic infrastructures should aim to follow a bottom-up approach starting by adopting a reference ontology and slowly evolving their semantic infrastructure as new requirements for higher expressivity arise.

V. Conclusion

The long-term requirements of the national interoperability frameworks in Norway need explicit representations of archetype semantics. We studied if SNOMED-CT´s compositional grammar can provide the level of expressiveness needed for covering these requirements. Technical, expressivity, human, and models mismatch challenges are present when defining archetype semantics as SNOMED-CT expressions. These challenges show a pressing need for the collaboration between archetypes and SNOMED-CT editors so both become aware of each other´s challenges. Future collaboration should: provide better guidelines for archetypes terminology binding assessing the best approach to follow depending on the interoperability needs; set the requirements for better tooling to ease the binding process; and work towards developing mechanisms that allow for improving expressivity, alignment between models, and extensibility of the SNOMED-CT concept model with other biomedical ontologies.

ACKNOWLEDGMENT

This work was funded by the Norwegian Centre for E-health Research.

REFERENCES

[1] D. Markwell, L. Sato, and E. Cheetham, “Representing Clinical Information using SNOMED Clinical Terms with Different Structural Information Models.,” in *Proceedings of the 3rd International Conference on Knowledge Representation in Medicine*, Phoenix, Arizona, 2008, vol. 410, pp. 72–79.

[2] L. Marco-Ruiz, K. Nicolaisen, R. Pedersen, A. Makhlysheva, and P. A. Bakkevoll, Ontology-based terminologies for healthcare - Impact assessment and transitional consequences for implementation, 1st ed., vol. 1. Tromsø: Norwegian Centre for E-health Research, 2017.

[3] “HL7 TermInfo Project.” [Online]. Available: http://wiki.hl7.org/index.php?title=TermInfo_Project#. [Accessed: 01-Jan-2019].

[4] T. A. Oniki, J. F. Coyle, C. G. Parker, and S. M. Huff, “Lessons learned in detailed clinical modeling at Intermountain Healthcare.,” J Am Med Inform Assoc, vol. 21, no. 6, pp. 1076–1081, Nov. 2014.

[5] A. R. Rasmussen and K. Rosenbeck, “SNOMED CT implementation: implications of choosing clinical findings or observable entities,” *Stud Health Technol Inform*, vol. 169, pp. 809–813, 2011.

[6] A. Budrionis , l. Marco-Ruiz , KY. Yigzaw, and JG. Bellika . Building a Learning Healthcare System in North Norway. Proceedings from The 14th Scandinavian Conference on Health Informatics 2016, Gothenburg, Sweden, April 6-7 2016, Linköping University Electronic Press; 2016, p. 1–5.

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

[8] H. Lærum, S. L. Bakke, R. Pedersen, and J. T. Valand, “An update on OpenEHR archetypes in Norway: Response to article Christensen B & Ellingsen G: ‘Evaluating Model-Driven Development for large-scale EHRs through the openEHR approach’ IJMI May 2016, Volume 89, pages 43-54,” *Int J Med Inform*, vol. 93, p. 1, Sep. 2016.

[9] B. E. Dixon *et al.*, “A pilot study of distributed knowledge management and clinical decision support in the cloud,” *Artificial Intelligence in Medicine*, vol. 59, no. 1, pp. 45–53, Sep. 2013.

[10] A. L. Rector, “The Interface between Information, Terminology, and Inference Models,” in *Proceedings of the 10th World Congress on Medical Informatics*, London, UK, 2001, vol. 84, pp. 246–250.

[11] D. Karlsson, M. Berzell, and S. Schulz, “Information Models and Ontologies for Representing the Electronic Health Record.,” in *International Conference on Biomedical Ontology*, Buffalo, NY, 2011, p. :153–157.

[12] L. Marco-Ruiz, C. Pedrinaci, J. A. Maldonado, L. Panziera, R. Chen, and J. G. Bellika, “Publication, discovery and interoperability of Clinical Decision Support Systems: A Linked Data approach,” *Journal of Biomedical Informatics*, vol. 62, pp. 243–264, Aug. 2016.

[13] C. Pedrinaci, J. Domingue, and A. P. Sheth, “Semantic Web Services,” in *Handbook of Semantic Web Technologies*, J. Domingue, D. Fensel, and J. A. Hendler, Eds. Springer Berlin Heidelberg, 2011, pp. 977–1035.

[14] P. D. Fensel, D. F. M. Facca, D. E. Simperl, and I. Toma, “Web Service Modeling Ontology,” in *Semantic Web Services*, Springer Berlin Heidelberg, 2011, pp. 107–129.

[15] D. Booth, “Yosemite Project | Comparing the Yosemite Project and ONC Roadmaps for Healthcare Information Interoperability,” *Yosemite Project*, 02-Apr-2015. [Online]. Available: http://yosemiteproject.org/comparing-the-yosemite-project/. [Accessed: 01-Jan-2019].

[16] “Rdf - FHIR v3.0.1.” [Online]. Available: https://www.hl7.org/fhir/rdf.html. [Accessed: 01-Jan-2019].

[17] C. Soguero-Ruiz *et al.*, “Ontology for heart rate turbulence domain from the conceptual model of SNOMED-CT,” *IEEE Trans Biomed Eng*, vol. 60, no. 7, pp. 1825–1833, Jul. 2013.

[18] C. Martínez-Costa, R. Cornet, D. Karlsson, S. Schulz, and D. Kalra, “Semantic enrichment of clinical models towards semantic interoperability. The heart failure summary use case,” *J Am Med Inform Assoc*, vol. 22, no. 3, pp. 565–576, May 2015.

[19] IHTSDO, “SNOMED CT Compositional Grammar Specification and Guide.” [Online]. Available: https://confluence.ihtsdotools.org/display/DOCSCG. [Accessed: 01-Jan-2019].

[20] E. Sundvall *et al.*, “Integration of tools for binding archetypes to SNOMED CT,” *BMC Med Inform Decis Mak*, vol. 8, no. Suppl 1, p. S7, Oct. 2008.

[21] P. N. Robinson, “Deep phenotyping for precision medicine,” *Human Mutation*, vol. 33, no. 5, pp. 777–780.

[22] “Semantic Annotations for WSDL and XML Schema.” [Online]. Available: http://www.w3.org/TR/sawsdl/. [Accessed: 01-Jan-2019].

[23] J. Kopecký and T. Vitvar, “MicroWSMO,” *WSMO Working Draft - D38v0.1 MicroWSMO*, 19-Aug-2008. [Online]. Available: http://www.wsmo.org/TR/d38/v0.1/. [Accessed: 01-Jan-2019].

[24] K. Gomadam, A. Ranabahu, and A. Sheth, “SA-REST: Semantic Annotation of Web Resources,” *SA-REST: Semantic Annotation of Web Resources*. [Online]. Available: https://www.w3.org/Submission/SA-REST/. [Accessed: 01-Jan-2019].

[25] A. L. Rector, R. Qamar, and T. Marley, “Binding ontologies and coding systems to electronic health records and messages,” *Applied Ontology*, vol. 4, no. 1, pp. 51–69, Jan. 2009.

[26] C. Martínez-Costa and S. Schulz, “Ontology-based reinterpretation of the SNOMED CT context model.,” in *Proceedings of the 4th International Conference on Biomedical Ontology*, Montreal, Canada, 2013, pp. 90–95.

[27] https://arketyper.no/ckm/ (15.02.2019)

[28] https://ec.europa.eu/digital-single-market/en/news/cross-border-health-project-epsos-what-has-it-achieved (15.02.2019)