Semantic Interoperability for Health Network

Deliverable 4.4: Report on interface specifications between semantic artefacts

[Version 2.0, July 31, 2015]

Call: FP7-ICT-2011-7
Grant agreement for: Network of Excellence (NoE)
Project acronym: SemanticHealthNet
Project full title: Semantic Interoperability for Health Network
Grant agreement no.: 288408
Budget: 3.222.380 EURO
Funding: 2.945.364 EURO
Start: 01.12.2011 - End: 31.05.2015
Website: www.semantichealthnet.eu
Coordinators:

The SemanticHealthNet project is partially funded by the European Commission.
<table>
<thead>
<tr>
<th>Deliverable:</th>
<th>D4.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publishable summary:</td>
<td>SemanticHealthNet faces the challenge of providing means to improve semantic interoperability of Electronic Health Record (EHR) systems. Without adhering to any of the existing EHR representations and approaches, WP4 has put the focus on the application of semantic technologies to the representation of the EHR. The clinical focus has been set on chronic heart failure and cardiovascular prevention, which has driven the development of our semantic resources. The Heart Failure Summary (HFS), a summary with the essential information for the diagnosis and treatment of heart failure, is one of these resources. In previous deliverables we have concentrated on elaborating a semantic interoperability approach based on the use of an ontological framework and a set of top-level semantic patterns. We also applied such approach to heterogeneous representations of parts of the HFS in order demonstrate its benefits to achieve semantic interoperability across them. In this deliverable we have opened a bit the scope and we do not focus only on the EHR representation but we explore how it interfaces with external systems such as decision support systems (DSS) which implement certain guideline specifications. For that we have studied the interfaces or relationships between clinical practice guidelines (CPG), EHR information models and domain ontologies. In our study we have analysed the main challenges that arise at each interface and how they could be addressed by the semantic-driven approach proposed. We also describe how this interface is addressed within existing approaches (e.g. provisional IHTSDO terminology binding approach, ContSys, etc.) and how the semantic approach proposed can be used as harmonization tool, without claiming to substitute any of them. The wider scope of this deliverable makes that many of the results provided are experimental and may need of further elaboration when considered worth to be done. Within the document we also analyse the interface between clinical guidelines, EHR information models and ontologies from more practical side, by selecting two different clinical guideline representation formalisms (i.e. GDL and PROforma) and using them for modelling a simple clinical guideline rule related with the heart failure diagnosis. For both guideline formalisms we sketch how they could benefit from a more semantic-driven approach. Further elaboration on this is expected as result of the implementation of the evaluation plan carried out in conjunction with WP2.</td>
</tr>
</tbody>
</table>

| Status: |  |
| Version: | 2014-12-19 |
| Public: | □No ×Yes |
| Deadline: |  |
| Contact: | Catalina Martínez-Costa catalina.martinez@medunigraz.at Stefan Schulz stefan.schulz@medunigraz.at |
| Editors: | Catalina Martínez-Costa catalina.martinez@medunigraz.at Stefan Schulz stefan.schulz@medunigraz.at Karl-Henrik Lundell karlhenrik.lundell@gmail.com William Gossen wgoossen@results4care.nl Jean Marie Rodrigues rodrigues@univ-st-etienne.fr Dipak Kalra d.kalra@ucl.ac.uk |
# Table of contents

## Contents

Table of contents ................................................................................................................................. 3

1 Introduction and objectives ................................................................................................................. 5
  1.1 Background ................................................................................................................................... 5
  1.2 Objectives ..................................................................................................................................... 6
  1.3 SHN Semantic Architecture ........................................................................................................... 7

2 Interface between the Ontology and the EHR Information Model .................................................. 11
  2.1 Motivation ..................................................................................................................................... 11
  2.2 Current Implementations: State of the Art for terminology binding ........................................... 12
    - Lexical and semantic-based binding approaches ......................................................................... 12
    - Approaches based on guideline specifications ........................................................................... 14
    - IHTSDO approach ......................................................................................................................... 14
    - Based on a formally represented shared semantic model ......................................................... 16
  2.3 Discussion on terminology binding approaches and requirements for an interoperable EHR representation ......................................................................................................................... 17
  2.4 Relationship of the SHN approach with existing terminology binding approaches .......... 19
  2.5 The system of concepts for continuity of care – ContSys ............................................................ 21
  2.6 The role of ContSys in an example from Sweden ......................................................................... 23
  2.7 Relationship of the SHN approach with ContSys ........................................................................ 24

3 Interface between the Ontology and Clinical Practice Guidelines .................................................. 28
  3.1 Challenges ..................................................................................................................................... 29

4 Interface between EHR Information Model and Clinical Guidelines ............................................. 31
  4.1 Challenges ..................................................................................................................................... 32

5 Clinical Guidelines and their interface with Health records ............................................................ 33
  5.1 Overview ....................................................................................................................................... 34
  5.2 Use Case: “Patient with atrial fibrillation” .................................................................................. 35
  5.3 The Guideline Definition Language (GDL) .................................................................................. 37
  5.4 PROforma ...................................................................................................................................... 40
  5.5 SHN Semantic Approach – Clinical guideline rules as ontology-based queries .................... 44
    - Map clinical model data elements and their values to semantic patterns .................................. 48
    - Represent our fictitious patient data as OWL DL instances of the patterns ............................... 48
    - Formulate the rule as a set of SPARQL queries ........................................................................ 49
    - Run the queries against the test data instances ....................................................................... 51
    - Comments to the semantic approach: ....................................................................................... 51
  5.6 Benefits for GDL and PROforma of implementing the SHN semantic approach .................... 52

6 Summary and conclusions .................................................................................................................. 53

7 Annex .................................................................................................................................................. 55
  7.1 Heart Failure Summary and ESC Guidelines (Jean Marie Rodrigues) ....................................... 55
  7.2 Heart Failure Summary ................................................................................................................ 55
  7.3 Clinical Guidelines for Chronic Heart Failure ........................................................................... 56
7.4 European Society of Cardiology (ESC) Chronic Heart Failure Guideline..............................57
7.5 Examples ........................................................................................................................................58
    Heart Failure Diagnosis..................................................................................................................58
    Heart Failure Treatment options.................................................................................................60
7.6 GDL capabilities to formalize ESC/Cardiac Failure Guidelines (Jean Marie Rodrigues, Rong Chen) 61
    Capacity for Modeling the Heart Failure Summary .................................................................61
    Presentation and symptoms..........................................................................................................61
    NYHA Class..................................................................................................................................61
    Physical exam..............................................................................................................................61
    Laboratory test request...............................................................................................................62
    Rule writing using GDL...............................................................................................................62
    Heart Failure Diagnostic Rules...................................................................................................63
1 Introduction and objectives

1.1 Background

SemanticHealthNet (SHN) faces the challenge of improving semantic interoperability of clinical information. It is based on the realistic assumption that the co-existence between several standards, specifications and proprietary implementations for representing the content of electronic health records (EHRs) will endure. Thus, SHN focuses on providing an integrative semantic abstraction on top of existing EHR representations, providing a homogeneous view that enables to mediate across the heterogeneous underlying representations in different contexts.

Although the SHN approach is understood as universal (i.e. applicable for the entirety of medicine), SHN has focused on the representation of chronic heart failure, which has driven the development of semantic resources by work package four (WP4).

Previous deliverables produced by WP4 have highlighted two kinds of semantic artefacts, namely information models (e.g. openEHR, EN ISO 13606, etc.) and ontology-based terminologies like SNOMED CT. We have demonstrated how clinical content about heart failure diagnosis and treatment represented by existing health information standards, could formally be described within an ontological framework, which is mostly independent of the specific layout of clinical models, their underlying specification, as well as on pre- or postcoordination preferences.

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

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

This important aspect of semantic interoperability, namely the interoperation of computerized clinical practice guidelines (CPGs) with clinical models and ontologies will be explored in this deliverable. In addition it will integrate the description of clinical process into the interoperability framework by means of the ContSys standard. In Chapter 2 we will describe our initial experiments in the formalization of ContSys concepts as ontology ones and in exploring their relationship with the proposed SHN ontological framework.

---

1 Clinical model and clinical information model are used as synonym terms throughout the document
2 Clinical practice guideline and clinical guideline are used as synonym terms throughout the document
3 EN ISO FDIS 13940 Health informatics - System of Concepts to Support Continuity of Care. CEN, the European Committee for Standardization, ISO TC 215/WG 3, prEN/ISO/DIS, CONTSYS, 08/2014
1.2 Objectives

This deliverable aims at analysing and consolidating the interfaces between three kinds of semantic artefacts: (1) Ontologies, (2) EHR information Models, and (3) Clinical guidelines.

EHR information models provide standardized information structures, relationships, and constraints to represent EHR data. Examples are EN ISO 13606-1, openEHR Reference Model or HL7 Reference Information Model (RIM). 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.

Ontologies formally describe properties and relations of types of entities. Domain-independent categories, relations and axioms are typically provided by top-level ontologies, whereas the types of things specific to a domain of interest are represented by domain ontologies.

Clinical practice guidelines (CPGs) represent procedural knowledge on diagnosis and therapy concerning specific types of health issue, with the aim of guiding decisions concerning diagnosis, disease management, and therapy driven by current clinical information about the patient at hand, i.e. patient specific information.

Although the above mentioned three kinds of semantic artefacts are different, they do interact each other. This interaction is not so trivial since there are overlaps between them (e.g. although theoretically disjoint, information models and CPGs have practical dependencies).

Typically, CPGs will refer to specific types of clinical process expected to be carried out for good quality of care, and refer to specific clinical data required in these processes. The missing link is that neither guidelines, nor their process descriptions have a standardised way to relate to EHR information models and/or domain ontologies, whereas there is an increasing consensus that such links would largely improve data capture, process support and selection and aggregation of data for secondary purposes.

In the following conceptual model we represent three kinds of mutually related semantic artefacts. Each relationship between two semantic artefacts contributes to an interface. There are three interfaces: (1) EHR information models – Domain ontologies; (2) Clinical practice guidelines – Domain ontologies and (3) Clinical practice guidelines - EHR information models (Figure 1-1). Each interface places obligations on the two models it connects, which must be coherent, as symbolised by the overlap of the interfaces at the centre of the drawing. In each direction the interface filters and transforms information so that the receiving model sees only that portion of the originating model and associated repository that is relevant for its own purposes. Any given set of applied constraints and obligations of the models forms defined interfaces and still allows changes in any of the models independently.
In the following we provide a summary of the semantic architecture proposed within SHN to improve EHR semantic interoperability. For more details, please refer to previous deliverables.

### 1.3 SHN Semantic Architecture

SemanticHealthNet has proposed a semantic-driven architecture organized in five layers that span from heterogeneous data repositories to homogeneous and semantically explicit representations.
Structured heterogeneous data (Layer 1): This layer comprises structured clinical data, which may be physically stored within an EHR repository and accessed via an interface conforming to some standard like HL7 V3, openEHR, EN ISO 13606, or to a proprietary database schema. Advanced EHR representations structure data by clinical models (e.g. archetypes), which provide a set of data elements and value restrictions. More than one expression [Data element: Value restriction] can be used for representing the same content: e.g. [Disease: Inflammation] and [Body location: Myocardium] vs. [Disease: Myocarditis]. In the first example, the semantic relation between the disease and the body location is not provided by the clinical model. Medical data are provided at different levels of detail, depending on the context requirements. E.g. smoking-related information recorded by a GP is coarser than information recorded in a smoking cessation service. This implies co-existing clinical models, providing different level of detail of information. Contextual information is essential, i.e. information that is relevant for the understanding of the text such as temporal information (e.g. past history), the subject of the information (e.g. family member, patient), where the information was acquired (e.g. hospital), etc., as well as the way information has been acquired (e.g. by machine, physician), which will influence its degree of trust. Each data element in a clinical model should provide this information in a standardized way. Data elements must be defined and may be encoded using terminologies, increasingly grounded on an ontological basis, such as with SNOMED CT. Nevertheless, the selection of a term by a user is usually driven by its - often imprecise - textual description and not by its formal definition. Additionally, some terms blend contextual information (e.g. past history) with the clinical term itself (e.g. lung cancer) (boundary problem⁴).

Semantic mapping (Layer 2): Semantic content patterns are introduced to bridge between structured data and their semantic representation. They describe recurring information structures and provide a particular view on the underlying model of meaning, tailored to the needs of particular use cases⁵.

---

preventing users from a deep knowledge of the underlying ontology formalisms. They can be seen as templates with fix and variable parts. Patterns can be organized in hierarchies, in which top-level patterns can be specialized following a paradigm similar to object-oriented design, and in which their composition permits to cover more extensive modelling use cases. Patterns may be encoded as Subject-Predicate-Object (SPO) triples, to which data elements and value restrictions of clinical models are mapped. Opposed to the unrelated data elements within clinical models (i.e. only structurally related), predicates constitute links between classes as dictated by the pattern such as given by the following examples:

\[
<\text{InformationItem} \; '\text{describes situation}' \; \text{Inflammation}\>,
<\text{InformationItem} \; '\text{describes situation}' \; \text{Disease}\>
<\text{Disease} \; '\text{has location}' \; \text{Myocardium}>.
\]

**Semantic mediator (Layer 3):** The core is constituted by ontologies which formalize clinical data meaning: information entity ontologies and medical domain ontologies, constrained by means of canonical categorisations and relations provided by a top-level ontology. We consider SNOMED CT a good candidate as medical domain ontology, due to its high coverage and its grounding in formal-ontological principles. However, SNOMED CT is not complete, and many instances of clinical domain expressions require augmentation either through using additional ontologies, or expanding SNOMED CT. A strict separation between *Information* and the entities represented by this information is fundamental. E.g. *Diagnosis* is a piece of information created by some clinician, at some place, at some point in time by following a certain method. A *Diagnosis* represents a possible clinical situation of the patient such as *inflammatory disorder*. A cause of the *Inflammatory disorder* can be a *Viral infection*. *Severity* information modifies the information object, as severity is an outcome of the clinician’s interpretation. Severity is therefore epistemic information that reflects what is known about certain clinical situation. This separation is essential for a consistent interplay of information models with clinical terminologies. This is also essential for obtaining data from systems deploying technologies depicted in layer 1. Ontologies lend themselves to be expressed in a logic-based formalism, e.g. description logics (DL), which enables inferencing, pursuing three main goals: (i) detect semantic equivalences across different distributions of content between information models and terminologies; (ii) advanced exploitation of clinical information by means of semantic queries, and (iii) providing clear-cut meanings for human understanding to avoid ambiguities.

According to the first goal, detecting the equivalence of different combinations of constituents to represent the same data, can be detected by machine reasoning:

\[
\text{InflammatoryDisorder \ and \ hasLocation \ some \ Myocardium \ equivalentTo \ Myocarditis}.
\]

The second goal refers to the ability to query data at different levels of detail (e.g. patients with inflammatory diseases, with cardiac disorders, etc.).

**Virtual homogeneous data (Layer 4):** It provides a homogeneous view on data extracted from heterogeneous systems. Data are expressed at different levels of detail, but can be accessed homogeneously thanks to the underlying ontology-based annotations. In contrast to traditional databases, where complete information is assumed, ontologies assume incomplete information, which affects how data can be queried. Query patterns based on the ontology content patterns used for the semantic mapping should facilitate queries in the same way that content patterns facilitate representations in layer 2.
Application (Layer 5): Here we find clinical systems and services with different information needs. A public health system might be interested in aggregating information results to extract population-level information while a primary care system might be interested in the results of a lab test. A decision support component might need to retrieve one very specific data value. Each use case requires querying data at different detail and precision. Besides, all the data retrieved might not deserve the same trust level and it must be explicitly shown (e.g. seniority of the author, supporting evidence, indication of whether a diagnosis is tentative or confirmed etc.).
2 Interface between the Ontology and the EHR Information Model

2.1 Motivation

Electronic health records offer a range of potential benefits and many of them depend on being able to consistently process meaningful clinical information within them. In the last decades, two different threads have pursued this goal, a structural thread and a terminology thread. Different approaches to the use of structure and terminology have been developed in proprietary clinical systems and efforts to develop standards have tended to separate both aspects. As a consequence the boundaries between the models used for terminology construction and health record construction are blurred. The interface between information models and terminologies is known as **terminology binding**. The problem of delineating the border between both is known as **the boundary problem**.

EHR information models should represent the entities in which information is recorded (e.g. documents, headings, entries, etc.). They should refer to clinical concepts such as heart rate, myocardial infarction, etc. by means of their binding to terminologies, such as SNOMED CT and/or their relation to generic definitions in concept systems such as ISO 13940 (ContSys).

Multiple EHR information models have been proposed by different EHR standards and specifications (e.g. HL7 RIM, openEHR RM, EN ISO 13606 RM). Besides, current medical information systems are mostly based on proprietary information models. All of them have in common their use of terminology codes to represent some key parts of the EHR (e.g. diagnosis or drug related information).

Since standardized and proprietary information models have been independently developed, they are neither syntactically nor semantically interoperable with each other. Syntactic interoperability can be more easily achieved than semantic interoperability by defining a set of correspondence or mapping rules across their information model structures, as shown in Table 2-1 between OpenEHR and EN ISO 13606 specifications.

<table>
<thead>
<tr>
<th>OpenEHR</th>
<th>ISO 13606</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLUSTER</td>
<td>CLUSTER</td>
</tr>
<tr>
<td>PARTY PROXY</td>
<td>RELATED PARTY</td>
</tr>
<tr>
<td>CARE ENTRY</td>
<td>ENTRY</td>
</tr>
<tr>
<td>ADMIN ENTRY</td>
<td>ENTRY</td>
</tr>
</tbody>
</table>

*Table 2-1 Example of OpenEHR-ISO13606 information model structure correspondences*

However, semantic interoperability requires interpreting the meaning of each of the information model structures. In dual model-based EHR health information architectures (e.g. openEHR, EN ISO 13606, HL7 CDA templates, etc.), information model structures are constrained to create clinical information models (viz. clinical models). These provide structured and constrained representations of clinical recording scenarios. More than one clinical model can be created for addressing the same use case. They will possibly contain structural differences (e.g. providing a value set of disorders for a diagnosis versus providing a set of disease-specific yes/no indicators) and also differences in the level of

---

7 https://openehr.atlassian.net/wiki/display/ stds/openEHR+to+ISO+13606-1,+ISO+21090+mapping
information detail required (e.g. whether or not to record the severity of the disease diagnosed). Templates such as the ones proposed by HL7 CCDA or openEHR are usually used to constrain a set of more generically specified clinical models for specific recording scenarios.

The flexibility offered by EHR standards and specifications is understood in a context in which multiple ways of recording clinical information and its context are required. However, semantic interoperability requires recording the meaning of the information collected by each of the possible clinical models produced, independently of how this information was structured or rendered.

The ontological framework proposed within SHN aims at providing a representation in which information model structures and terminologies are uniformly represented and constrained by clearly delimiting the border between both. It enables to deal with the existing plurality of representations. The framework has been mostly built from existing ontologies, avoiding the creation of new ones. To standardize the way these ontologies relate each other, the biomedical top-level ontology BioTopLite has been used.

SHN’s framework makes the meaning explicit that is carried by information structures within clinical models. However, this requires a deep understanding of the underlying ontologies and their representational formalisms, in our case description logics (DLs). In order to mitigate this we have proposed the use of semantic patterns.

As a result of practice in providing such expressions to clinical models, we noticed that they very often followed some recurrent modelling patterns. We name these semantic patterns. Such semantic patterns avoid that modelers need of a deep knowledge of ontologies and DLs. Here, they are used to guide the ontology-based representation of clinical information. In order to do so correspondences between clinical model nodes and semantic patterns are established. Semantic patterns contain fixed and variable parts, the latter being filled by the data instances from the clinical models.

### 2.2 Current Implementations: State of the Art for terminology binding

Finding the right target identifiers or expressions to bind information model structures from clinical models to the terminology is not trivial, mainly due to the high flexibility provided by clinical models and the lack of explicit relationships between their elements. In the last years, several strategies for doing the binding have been proposed. In the following we describe some of them:

**Lexical and semantic-based binding approaches**

Sundvall\(^8\) and Qamar et al.\(^9\) propose a semi-automatic SNOMED CT terminology binding strategy based on lexical and semantic procedures. Their goal is to reduce the effort required for doing the binding, by helping modelers to arrive at a final decision by providing a set of candidate terms.

---


Sheng Yu et al.\textsuperscript{10} propose the creation of “Terminological Shadows” from archetypes. A shadow consists of a tree of terminological concepts, identified by using automatic and configurable search algorithms, performed on archetypes. This shadow establishes links between archetypes nodes and the terminology system (e.g. blood pressure observable entity SNOMED CT term and the archetype node that represents a clinical event of measuring blood pressure).

Figure 2-1 depicts an example of candidate terms retrieved by using Sundvall and Qamar et al. approach.

![Example of candidate terms retrieved](image)

Figure 2-1 14 results were returned for 'platelet count'. The user has reduced the number of results by selecting only the 'finding' and 'procedure' categories of SNOMED CT.\textsuperscript{8}

Such approaches help inform the binding decision, but they do not address the complexity derived from the use of pre-coordinated and post-coordinated terminological expressions. A pre-coordinated expression\textsuperscript{11} identifies a composite meaning, which can be broken into atomic representational units. Pre-coordinated expressions have a single identifier (e.g. colon cancer is non-atomic and has a single identifier). Elkin et al.\textsuperscript{12} state that the use of post-coordination (i.e. compositional terminologies) is one potential solution to the problem of terminology content completeness, since anticipating and pre-coordinating all possible expressions within a terminology is not feasible. However, it requires the means to determine if expressions are equivalent. Besides, there is the risk of generating expressions whose equivalency cannot be algorithmically determined—i.e. there maybe two different ways of composition which mean the same thing but no way to determine automatically that they are equivalent\textsuperscript{13} (e.g. Colon cancer can be represented either as: \textit{Malignant Neoplasm - Has finding site – Colon}; or as: \textit{Colon – HasSpecimen - Malignant Neoplasm}). The complexity we refer to is how to determine such equivalences, since binding based on lexical and semantic techniques allows the co-existence of

\textsuperscript{10} Sheng Yu, Damon Berry, Jesús Bisbal, An Investigation of Semantic Links to Archetypes in an External Clinical Terminology through the Construction of Terminological\textsuperscript{\textregistered} Shadows”

\textsuperscript{11} http://www.cap.org/apps/docs/snomed/documents/snomed_ct_glossary.pdf


both pre-/post-coordinated expressions. Elkin et al. state that to be able to compare both kinds of expressions they must share the same data schema.

Approaches based on guideline specifications

Markwell et al. summarize two implementation specifications for the use of SNOMED CT in HL7 and openEHR respectively. These specifications address the most common overlaps and provide modeling guidelines to resolve the ambiguities. The HL7 specification is the outcome of the TermInfo project\textsuperscript{14}. The openEHR specification is part of work carried out within the UK NHS, which showed the complexity of binding archetypes to the terminology due to their high flexibility and the lack of explicit relationships between their elements. An example of guideline rule specification in the context of the TermInfo project is as follows\textsuperscript{15}:

“An Observation class instance in which the Observation.value is a SNOMED CT expression representing a [<<404684003 | clinical finding] or a [<<413350009 | finding with explicit context] SHALL NOT contain an Act.code which when interpreted with the Observation.value yields a meaning that is substantially different from the meaning implied if the Act.code was "ASSERTION".

- For example, an Act.code meaning "Past history" or "Family history" may substantially alter the interpretation of a [<<404684003 | clinical finding] and should not be used in this way. Instead the SNOMED CT context model should be used to capture these significant differences in meaning.”

The specification of modeling guidelines to bind a terminology to an information model is not trivial. While it might work for a concrete implementation involving a specific information model and terminology, it does not guarantee semantic interoperability across representations based on other guidelines.

Markwell et al.\textsuperscript{6} state that for the process of integrating terminology with the information model certain normalization of the information model is essential to ensure that the record can be used to answer questions such as, “does the patient have a past clinical history of respiratory problems?.

IHTSDO approach

Within IHTSDO, a terminology binding approach\textsuperscript{16} is recently being investigated in the context of SNOMED CT, in which four types of bindings are proposed. Two of them, named “Value set binding” and “Dependency constraint binding”, focus on the definition of value sets. The first one allows an information model artefact to reference one or more values, represented as SNOMED CT codes. The second one constrains the value of one data element based on the value of another, where at least one of these data elements is coded. The third and fourth bindings named “Model meaning binding” and “Expression template binding”, are focused on making explicit the meaning represented by information artefacts.

The value set binding diagram (see Figure 2-2) shows how information model structures are bound to SNOMED CT value sets described by using expression constraints such as < 71388002 | procedure>, which means all subclasses (but not the class itself) of the procedure concept.

\textsuperscript{14} http://www.hl7.org/Special/committees/terminfo/
Figure 2-2: Value set binding diagram (left: clinical model tree; right: terminology artefacts (i.e. concepts, attributes, expressions and value set); middle: the terminology binding is represented as a line from the information model to the terminology which can have metadata associated.)

The following diagram (see Figure 2-3) shows the dependency constraint binding by using a discharge summary model in which the value set used to populate the 'Procedure' data element is dependent on the sex of the patient. In this case, if the Patient is Male, then the value of Procedure is selected from the 'Male procedure reference set', which includes some procedures that are only applicable to Males. If the Patient is Female, then the value of Procedure is selected from the 'Female procedure reference set'. If, however, it is not known whether the Patient is Male or Female, or they are neither, then the value of Procedure is selected from the 'All procedures reference set'. In this example, the condition which determines which value set is used is considered as part of the binding itself.

Figure 2-3: Dependency constraint binding diagram (left: clinical model tree; right: terminology artefacts (i.e. concepts, attributes, expressions and value set); middle: the terminology binding is represented as a line from the information model to the terminology which can have metadata associated.)

Figure 2-4 shows the model meaning binding in which each information model data element is bound to a concept from the terminology. That concept does not necessarily represent the actual values but the conceptual category of concepts represented by the data element (e.g. observable entity, clinical finding, etc.).
Finally, the expression template binding (see Figure 2-5) takes the model meaning binding one step further. An expression template binding indicates how values recorded in two or more fields may be combined to represent a composite meaning. Expression template bindings take a set of data elements in the information model, and link them to an expression template. This kind of terminology binding is the one closest to the approach provided within SHN WP4, in which expression templates would correspond to the proposed semantic patterns.

Based on a formally represented shared semantic model

Rector et al. address the terminology binding from a formal perspective, proposing a methodology that uses OWL to formally define the interface between the information model and the terminology, called “Code Binding Interface”.

They define information models and codes as data structures and symbols which both constitute the model of use. They are analogous to grammars rather than knowledge bases (e.g. notions such as “missing” and “null” make sense in terms of data structures, but not in terms of patients and their
diseases). Instances of these models are data structures to be validated. Codes and data structures do not “mean” anything on their own; they are merely valid or invalid.

In contrast, for the authors, the ontology (i.e. model of meaning) is intended to represent our conceptualisation of the world. Individuals in the ontology represent particular things in the world – e.g. a particular case of diabetes, an individual patient, a particular administration of a dose of insulin.

The authors propose to decouple the coding system from the model of meaning so that the reasoning processes about the model of meaning and the coding system model are always separate. For that, they propose to do the binding at the level of the information, stating that the latter can be linked to ontologies only indirectly via meta-models or codes and they provide a formalization example in OWL. They recognize that the representation proposed is tedious, and OWL could be seen as an assembly language. They state the need for a high level language adapted to the task of representing information systems and their binding to coding systems, along with ‘compilers’ to transform it to OWL in a standard way.

The methodology proposed by Rector et al. provides mechanisms for formalizing the binding but not for establishing which bindings are valid and which are not (e.g. a diagnosis data structure can have only clinical findings as values).

### 2.3 Discussion on terminology binding approaches and requirements for an interoperable EHR representation

Each of the described approaches provides highly valuable output and addresses particular challenges. In reality, terminology models and information models often overlap. The reason for this is that terminologies or information models are normally designed and created in isolation. Some clinical terminologies support features such as negation and qualification, others do not. In addition, not all information models provide comparable sets of features. Besides, compositional terminologies are required in order to avoid their content explosion, becoming unmanageable (as in the case of SNOMED CT and its post-coordination mechanism). This compositionality implies a certain order and requires certain structure. For instance, the phrase “numbness of right arm and left leg” could be represented using the following combination of codes from SNOMED CT: *Numbness* (44077006), *Right* (24028007), *Arm* (40983000), *Left* (7771000) and *Leg* (30021000). By changing the order of the codes as shown next, the meaning can be changed to “numbness of left arm and right leg”: *Numbness* (44077006), *Left* (7771000), *Arm* (40983000), *Right* (24028007) and *Leg* (30021000). In SNOMED CT, the following definitions could be provided, by following the concept model structure:

```
Numbness of right arm and left leg ===
44077006 | Numbness |
363698007 | Finding site | = 6921000 | Right upper extremity structure |
363698007 | Finding site | = 32153003 | Left lower extremity structure |
```

Or also by post coordinating the laterality:

```
Numbness of right arm and left leg ===
44077006 | Numbness |
```

---

HL7 versions 2 and 3 as well as in the HL7 Clinical Document Architecture (CDA) use a name value pair, or an entity-attribute-value approach to provide a solution to this problem. For example, in HL7 Version 3, a RIM class named **Observation** represents clinical observations. Among many attributes, it has a **code**, a **value** and a **targetSiteCode**. Next, their use to represent the “numbness of right arm and left leg” data instance:

**Observation.code**: Diagnosis.primary (LOINC 18630-4)

**Observation.value**: Numbness (SNOMED CT 44077006)

**Observation.targetSiteCode**: Arm (SNOMED CT 40983000)

  - Qualifier.name: Laterality; Qualifier.value: Left (SNOMED CT 7771000)

**Observation.targetSiteCode**: Leg (SNOMED CT 30021000)

  - Qualifier.name: Laterality; Qualifier.value: Right (SNOMED CT 24028007)

In the RIM specification, the three above attributes are specified as follows:

- SET<CE> **methodCode**: Attribute methodCode of type SET and cardinality 0..*
- SET<CD> **targetSiteCode**: Attribute targetSiteCode of type SET and cardinality 0..*
- ANY **value**: Attribute value of type ANY and cardinality 0..*

The above attribute specification provides constraints for the syntax to be used. E.g. the **targetSiteCode** can be specified as a set of CD (concept descriptors). However it does not constrain the set of values allowed for certain attributes\(^\text{18}\). This has to be additionally specified and there is no formal way to guarantee the semantic consistency across the value sets specified. This means that the above representation allows encoding “numbness of left head”, without raising an error.

**Observation.code**: Diagnosis.primary (LOINC code 18630-4)

**Observation.value**: Numbness (SNOMED CT 44077006)

**Observation.targetSiteCode**: Head (SNOMED CT 302548004)

  - Qualifier.name: Laterality; Qualifier.value: Left (7771000 SNOMED CT)

One possibility to solve this problem would be to provide guidelines for the use of the terminology within the information model, as TermInfo does. However, guidelines for each information model and terminology model would have to be provided and the resulting representations would not be each other interoperable.

Besides, the specification of value sets for each of the attributes requires keeping the semantic consistency across them. Without the underpinning by a shared model of meaning inconsistent representations might be produced. SNOMED CT is built on a model of meaning, named Concept Model. It is more and more based on ontological principles and, while not perfect, offers higher degree of semantics than any other terminology. However, SNOMED CT as a terminology is not intended to support a

---

\(^{18}\) The new datatypes in HL7 CDA and v3 allow for full post-coordinated expressions
complete representation of EHR content. In practice SNOMED CT provides a valuable component within the EHR as a common representation of a large range of clinical concepts. This representation includes not only the identifiers for concepts but also defining relationships between these concepts which assist effective retrieval.

Given the above, we enumerate a set of requirements desired for an EHR representation, to be semantically interoperable:

1. Provide a clear delineation between information model and terminology model; In order to do so, clear definitions of information model and terminology model must be provided.
2. Support post-coordinated expressions and means to detect equivalences across pre-coordinated concepts and post-coordinated expressions (also when concepts from different terminologies are involved);
3. Assure that the level of expressivity provided fits the intended purpose
4. Regard to implementability and maintainability: realistic to be implemented today and with the potential to be improved tomorrow
5. Assure scalability, to allow for representing new information models and ontologies

2.4 Relationship of the SHN approach with existing terminology binding approaches

The use of methods based on lexical and/or semantic analysis\textsuperscript{8,9,10} could benefit from the approach proposed within SHN and vice-versa.

On the one hand, terminology binding could use semantic patterns to suggest the appropriate concepts. As an example, the following diagnosis pattern has some fixed and variable parts. The variables, which are preceded by a question mark, are constrained to specific terminology concepts such as SNOMED CT clinical findings in the case of ?Situation. This could be used to restrict the analysis.

\begin{verbatim}
shn:InformationItem 'describes situation' ?Situation
shn:InformationItem 'results from process' sct:DiagnosticProcess
shn:InformationItem 'has attribute' ?InformationAttribute
\end{verbatim}

On the other hand, the suggestion of semantic patterns to be used could benefit from the lexical and semantic analysis performed by these approaches. For instance, the term suggested as binding of the root element of a clinical model could give a hint of the kind of pattern to use (e.g. observation result if it is bound to an observable entity in SNOMED CT). The annotation of semantic patterns with a list of keywords (e.g. diagnosis, lab result, examination, etc.) could also help to improve the pattern suggestion process, and the degree of precision obtained.

Semantic patterns also have the potential to ensure that more specific models such as the one proposed by CIMI are semantically valid derivations from higher level (top-level) patterns.

Top-level patterns are based on a set of generic ontology classes and predicates that can be specialized and composed by following the ontology constraints. These constraints can be used to determine which elements to include in a clinical model or in a terminology binding.
As a difference between clinical models and patterns, where the elements of the first one are only structurally related (e.g. list, tree, etc.), within patterns elements are connected by semantic relationships (e.g. `shn:isAboutSituation`, `btl:isOutcomeOf`, etc.). These relationships can be used to guide the decision of which elements to include in a model, reducing the arbitrariness. Currently, this decision is made by each modeller for themselves and it is not constrained, which can lead to the creation of non-interoperable models even for the same use case.

If semantic patterns are not applied at clinical models’ design time, they still can be used to detect semantic inconsistencies across them. As an example,

Figure 2-6 shows an excerpt from a CIMI model that records observation results. It records: (i) what is observed, `ELEMENT[id5.0.2]` (e.g. colour of the eye); (ii) the reason to perform the observation, `ELEMENT[id5.0.3]` (e.g. problem wearing contact lens); (iii) the method used to observe, `ELEMENT[id5.0.4]` (e.g. eye examination); (iv) the status of the observation, `ELEMENT[id5.0.5]` (e.g. performed, planned); and (v) the priority to perform the observation, `ELEMENT[id5.0.6]` (e.g. high, normal).

```
ITEM_GROUP[id1.1.1] matches { -- Observation set
    item matches {
        ELEMENT[id5.0.2] occurrences matches {1} matches { -- Code
            value matches { CODED_TEXT matches (id0.0.102)}}
        ELEMENT[id5.0.3] matches { -- Reason
            value matches { TEXT matches (id0.0.103)}}
        ELEMENT[id5.0.4] occurrences matches { -- Method
            value matches { TEXT matches (id0.0.104)}}
        ELEMENT[id5.0.5] occurrences matches {0..1} matches { -- Status
            value matches { TEXT matches (id0.0.105)}}
        ELEMENT[id5.0.6] occurrences matches {0..1} matches { -- Priority
            value matches { TEXT matches (id0.0.106)}}
    }
}
```

Figure 2-6 Excerpt of the CIMI model (CIMI-CORE-ITEM_GROUP.observation.v1.0.0) to record observation results

Figure 2-7 shows another CIMI model that records observation requests and references the above model by composition (keyword “link”). Besides, it also references a model to record observation actions (in the most updated version of the model appear as commented). Within this last model we have found a content overlapping with the observation result model, since it also provides elements for recording the reason, method, status and priority of the observation.

```
ITEM_GROUP[id1.1.1.1.1] matches { -- Observation Request
    item matches {
        use_archetype ITEM_GROUP [id0.0.0.1.1, CIMI-CORE-ITEM_GROUP.request_action.v1] -- Request action
        use_archetype ITEM_GROUP [id0.0.0.2.1, CIMI-CORE-ITEM_GROUP.observation.v1] -- Request observation
    }
    link matches { LINK[id0.0.0.3.1] -- Requested observation link }
}
```

Figure 2-7 Excerpt of the CIMI model (CIMI-CORE-ITEM_GROUP.observation_request.v2.0.0) to record an observation request

Semantic patterns could avoid such an overlapping situation, by providing formal modelling guidelines, based on the ontological framework, to distinguish across what is observed, the observation procedure

---

19 https://github.com/opencimi/archetypes/tree/master/miniCIMI
and the result of the observation. Additionally, they can help to guide binding or to detect inconsistencies in terminology bindings. For instance, the pattern logic axiom (shn:InformationItem and shn:isAboutSituation only shn:ClinicalSituation), relates an information entity (i.e. shn:InformationItem) to a clinical entity (shn:ClinicalSituation) and the latter is equivalent to SNOMED CT clinical findings. Therefore, if a information model structure is mapped to that axiom, its value is only valid if it is of the type clinical finding.

Approaches based on the guidelines specifications could also use the ontological framework proposed (model of meaning) to help the definition of value sets (e.g. specify which codes can be entered in the family history section of the record).

2.5 The system of concepts for continuity of care – ContSys

ContSys (system of concepts for continuity of care) is an international standard (ISO/FDIS 13940) that defines a conceptual framework for continuity of care. As Robson notes\(^{20}\), “developing a conceptual framework forces you to be explicit about what you think you are doing. It also helps you to be selective; to decide which are the important features; which relationships are likely to be of importance or meaning; and hence, what data you are going to collect and analyse”. ContSys has introduced concepts for clinical content related to a person’s health state as well as the provision of healthcare services related to this health state. ContSys also describes the clinical context characterized by processes in which organizations and persons participate, with the health state as the input and a consequent transformed health state as outcome.

ContSys enables a common clinical foundation of health care/clinical processes and define clinical concepts related to these processes. Thereby Contsys provides a conceptual basis for information management and information modelling in health care.

In practice this standard aims to be used whenever requirements for information in healthcare are specified. This will cover all levels of specifications in the development of:

- Enterprise/clinical models as a common basis for interoperability on international, national or local levels;
- Information systems;
- Structured information/clinical information models for specified types of clinical processes.

Clinical relevance, clinicians’ involvement and acceptance, as well as information management as an integrated part of clinical practice and clinical quality management are examples of requirements for effective information systems in health care. The business/enterprise/clinical views as the foundation for information is needed to fulfil such requirements.

In the development of health informatics, the clinical perspective probably often has been presumed to be obvious and unambiguously understood and common to all providers. This presumption may limit the true interoperability of existing information systems. A general problem is also that no generally accepted explicit enterprise/clinical models have been available to use as a starting point.

To support good clinical quality, semantic interoperability needs to be grounded in everyday clinical work. The real clinical situations and the constant reference to health and healthcare related concepts and terms have to be identified and systematically defined. This is the task of ontologies, which characterise, define, and hierarchically arrange concepts as well as their interrelations in a formal way.\(^\text{21}\)

Processes in real life apply information concerning concepts of the domain ontology. For example, clinical processes apply the content of a clinical ontology to a certain health problem in its organizational and institutional context. Clinical processes are fundamental in order to contextualize information in clinical documentation. Whether, e.g. the datum “past history of ischemic heart disease” is the interpretation of history taking, whether it is the result of the lookup in an old medical record, or whether it results from the interpretation of a current ECG, is crucial for the reliability of this finding.

SNOMED CT is increasingly referred to as an ontology. Yet, it should not be forgotten that SNOMED CT is also a terminology in the sense that it links clinical terms in several languages to SNOMED CT concepts. ISO/FDIS 13940 – Contsys as a system of concepts related to a clinical process model. This international standard also can claim to be ontological and be applied in an ontology-based approach. A convergence between SNOMED CT concept model and Contsys would be desirable.

The figure below shows the ContSys-based strategy for how the clinical reality can constitute the basis for healthcare information management. Every arrow in the figure represents successive structured and systematic modelling with defined interrelations and specialization with traceability. In this way the needed content and context is defined, its interrelations are specified and specific information is traceable to the models for clinical content and context.

![ContSys-based ontological approach](image)

**Figure 2.8** The ContSys-based ontological approach for semantic interoperability for patient care, knowledge management (clinical guidelines, protocols and clinical decision support) and follow up by secondary use of quality indicators, and other secondary data uses

The identification of the information needed in each step in a specific clinical process is governed and controlled by the definitions of clinical concepts and the allowed clinical characteristics in clinical ref-

\(^{21}\) Thomas Hofweber, Logic and Ontology, 2011, Stanford Encyclopedia of Philosophy
ference models (see the following section). In this way the risk of leaving out generally needed epistemic context for interpretations can be avoided. For example in the clinical process for chronic heart failure the degree of impaired circulatory function need to be specified. A body function identified as the results of certain healthcare investigations is conceptually an observed condition. Observed conditions are defined as observed aspects of the health state of a subject of care. Observed condition has a clinical reference model defining possible characteristics for such conditions – regardless type of clinical process. The observed condition can consequently be traceable to the healthcare investigation (a healthcare activity element) and the method used when performing the observation. The healthcare investigation activity element is also traceable to the step of the clinical process where it was performed. Was it performed before or after a specific treatment?, etc. All healthcare activities and all health conditions can be specified on the basis of the same basic clinical reference models.

The results of clinical process analyses systematically performed based on these clinical reference models can identify relevant clinical information for various clinical processes concerning:

- knowledge management/clinical decision support
- clinical data for information/documentation needed for co-operation in continuity of care and
- quality indicators for follow up

### 2.6 The role of ContSys in an example from Sweden

ContSys has not been released as final standard yet, and there is only limited experience in its use. Some Swedish experiences of usability of ContSys in developmental stages of eHealth are presented here.

In this work a set of clinical reference models, harmonised with openEHR and EN ISO 13606 information models and the SNOMED CT concept model. The clinical reference models are designed based on conformance to ContSys.

These clinical reference models cover all concepts in the clinical process model and all perspectives of clinical practice (i.e. documentation and communication in provision of care to an individual, follow up and clinical process management).

The method to create these models is to add clinical characteristics to each single clinical concept in ContSys. Those characteristics have been fetched from earlier and tested applications – openEHR reference model and SNOMED CT concept model attributes respectively. Neither openEHR reference model nor the SNOMED CT concept model relate to an explicit clinical process model. In order to achieve conformance to ContSys clinical process model a clinical reference model for clinical context have been added.

The clinical reference models for single concepts have then been combined in two steps to cover clinical situations and complete steps in the clinical process respectively.

Figure 2-9 shows an example of a compound clinical reference model (for healthcare needs assessment as a step in the clinical process) represented in UML.

Clinical reference models are intended to constitute the main link/interrelation/concretization between the clinical concept- and clinical information models. They provide means for traceability and
consistency of all clinical data to the core concepts in the clinical process and concept models. In this way clinical reference models have capacity to play a role in ontology-based approaches by clarifying the consistency of relations between internal layers/aspects in the domain ontology.

Figure 2-9 An example of a clinical reference information model for the process step of healthcare needs. The figure depicts how information classes and attributes are used to represent a set of concepts relevant for a certain clinical situation

2.7 Relationship of the SHN approach with ContSys

(partly extracted from submission to MIE 2015, still under review)

The ContSys framework is provided as a set of textual definitions supported by informative UML models. Our hypothesis is that the formalization of ContSys concepts as ontology concepts will help to implement the standard by facilitating its integration with existing information models and ontology-based medical terminologies. This formalization will possibly add precision to ContSys’ concept definitions and may detect hidden ambiguities and contradictions.

The concepts in ContSys are described based on their relations with the healthcare/clinical process model (Figure 2-10), which provides the clinical context. In an initial analysis of the standard we have focused on the concepts under Health issue and Health state.
In ContSys healthcare activities are organized in two main categories: healthcare investigations (outer circle) and treatments (inner circle). The initial input in the process model is the Health state of a Subject of care, a holistic concept for a person’s health. The final outcome is also a Health state, after the person being exposed to Healthcare activities (a set of interrelated activities transforming the inputs into outputs constitutes a process). Healthcare activity elements are direct or indirect. The direct ones are performed in direct interaction with the subject of care and have one or both of the purposes investigation or treatment. Examples of indirect Healthcare activities are to observe, assess, plan and evaluate.

Healthcare processes transform the input Health state of the Subject of care into the output Health state of the Subject of care whilst relating to different types of Health issues, and the associated activities performed. Examples of health issues are a loss of weight, a heart attack, an injury, etc. Health conditions are Health issues that represent existing or potential observations of a subject of care’s Health state. During a Healthcare process, Health conditions are observed by the Subject of care and by Healthcare professionals, which analyse, assess and draw conclusions, which modify these conditions. ContSys provides two main types of Health conditions (see Figure 2-11): (1) Observed conditions (by the subject of care, or by professionals) and (2) Potential health conditions (possible future or current health condition that is not yet observed, but is considered possible to become observable in the future as an aspect of a current or future health state). An Observed condition can be specialised as a Professionally assessed condition (observed condition assessed by a professional) and as a Resultant condition (actual outcomes). A Potential health condition can be specialised as a Target condition (desired possible outcomes), a Risk condition (probable consequences of adverse events), a Prognostic condition (potential health conditions representing the expected course of a Health state), a Considered condition (suspected but not yet observed) and a Excluded condition (not observed after relevant investigations).
We have initially tested our hypothesis (i.e. that the formalization of ContSys concepts as ontology concepts will help to implement the standard by facilitating its integration with existing information models and ontology-based medical terminologies) by placing selected ContSys content under the biomedical top-level ontology BioTopLite2 (BTL2), which includes enough classes, relations and axioms, in order to address the needs of most life sciences ontologies and, nonetheless, to provide a sound framework and guidance for developers. Top-level ontologies provide a basic set of fundamental classes and relationships. Their use standardizes the modelling task by establishing a reference framework that restricts it by means of logical axioms grounded on formal-ontological principles. The building of domain ontologies under BTL2 heavily constrains the choices of the ontology engineer, which is intended, as this results in a higher predictability of the ontologies produced.

The ContSys standard describes a system of concepts for continuity of care but also claims that it does not provide an information model, record model or document structure. ContSys should be complemented by further concretizations and used together with other standards like EN ISO 13606 and EN ISO 12967. Thus, it focuses on the referent of the information, not the information itself (i.e. information about it). According to this it is clearly of ontological / terminological nature.

However, our initial tests, according to our understanding of ContSys, have shown that ContSys concepts under Health issue, like Health condition correspond to information entities and not to referents thereof. This interpretation is based on the definition of Health issued provided by the standard. Whereas this is often characterized as epistemic, in opposition to ontological, we have developed a strategy in SemanticHealthNet that seamlessly integrates both kinds of terms / concepts under a common upper-level framework, which, however, enforces a strict division into “domain” entities and referring entities. Our approach alleviated the boundary problem, i.e. lack of a clear delimitation between information model and terminology, which is a recognized barrier to interoperability affecting SNOMED CT and EHR representations.
ContSys concepts need to be defined with utmost clarity if clinical models from electronic health record (EHR) standards should implement it, in order to avoid the boundary problem. The ContSys definition of Health state is close to the meaning of CONDITION in BTL2 (although in both cases it might not be necessarily defective), where it is defined as the “union of material entity, process, function, and disposition. The rationale of this class is to represent the ambiguous nature of what is commonly referred to by health-related condition (not necessarily pathologic)”. In this definition (from BTL2), a health or medical condition is not defined as an observation or perception (i.e. epistemic entity) as it is in ContSys. The wording “Considered condition” suggests that a condition has been considered, which matches the standard reading of “condition” as Health state. But what is meant here is the outcome of this process of considering, i.e. an information entity, which is however a clearly different entity.

Our initial pilot study suggests that the rooting of a standard like ContSys in an ontological framework makes sense not only for adjusting terms and definitions, but also because it has the potential to detect logical contradictions.

Future work will involve additional parts of ContSys, together with an investigation of future alignment with the SNOMED CT concept model and stricter naming conventions targeting self-explaining concept labels.
3 Interface between the Ontology and Clinical Practice Guidelines

According to the definitions provided by the system of concepts for continuity of care (ContSys) a clinical guideline is a set of systematically developed statements to assist the decisions made by healthcare actors about healthcare activities to be performed with regard to specific health issues\(^{22}\).

Healthcare processes and their constituent activities need medical knowledge to be carried out. Clinical guidelines and protocols\(^ {23}\) are one of the main means to support the application of such knowledge in clinical work. Of course we do need clinical training and methods etcetera, however, for the purpose of this deliverable, we focus on the guidelines and protocols.

Health issues and health conditions, methods and workflow, as defined within ContSys, are among the main concepts related with the representation of clinical processes and healthcare activities (performed for clarifying/investigating health conditions and for treatment of the identified health problems). Health conditions and health issues, as described in Section 2.7 from Chapter 2, can be represented by an ontology. The main relationship between them and clinical guidelines is how the applied knowledge will influence:

- the interpretations of observed aspects of a health state as health conditions, and
- the choices of healthcare activities (actions) to be performed.

Knowledge management (like all types of management processes) can be divided into strategic and operational levels. The strategic level is how the knowledge is built up from evidence based science for the specific type of health problem and clinical process. This knowledge base should then be operationalized in the practical support and recommendations for assessments and performance in the local context – usually by developing clinical protocols based on the clinical guidelines that contain the knowledge base.

Protocols should include support for the two main components of knowledge management:

- criteria for interpretations and assessments of observations in identifying health conditions
- recommendations for performance of healthcare activities (informing care plans for individual patients)

In the figure below clinical practice is represented as the horizontal bar *Clinical process*. Clinical guidelines and the knowledge base for these are included in the horizontal process management bars. Red arrows show how knowledge management interact with the clinical process. This is a way to show one important relationship between the ontology and clinical guidelines.

\(^{22}\) Health issue is a representation of an issue related to the health of a subject of care as identified by one or more healthcare actors (e.g. a loss of weight, a heart attack, a drug addiction, etc.)

\(^{23}\) According to ContSys a protocol is a customized clinical guideline. It is more precise.
3.1 Challenges

Within the challenges that exist for implementing the interface between both ontologies and clinical guidelines we highlight the following ones:

- **Use of one or more common and standardized ontologies**: EHR systems use heterogeneous terminologies and ontologies. Most of clinical guideline experts use at least partly their own local ontology. The challenge will be to handle various standardized ontologies together in a consistent architecture. E.g. in the European Society of Cardiology heart failure guidelines (see details in Annex) they do not always align their terminology with a common reference terminology/ontology as SNOMED CT. We may find in the EHR SNOMED CT concepts such as breathlessness, fatigue and tiredness and in the clinical guideline statements such as “Increased time to recover from ordinary physical activity”.

- **Ontologies shall take into account the information needs of clinical guidelines**: They must provide the concepts required by clinical guidelines as well as concepts at different levels of granularity. E.g. Value sets as the classes of the NYHA assessment scale. SNOMED CT observable entity concepts related to Echocardiography (Left Ventricular Ejection Fraction) or Echocardiogram.

Besides, it is important to follow ongoing developments in the space of clinical terminologies. For instance, for lab values there is an agreement between LOINC and SNOMED CT. The observable entity goes to LOINC and the value set with the clinical findings goes to SNOMED CT, as part of recent agreements. Given that LOINC has fully coded the observable entities and the clinical findings of several assessment instruments, this rule would not always apply, but further developments in this space can be expected.
In order to deal with the above challenges, two decisions can be taken:

- To prescribe that guideline experts map their non-standardized vocabularies to standardized ones such as SNOMED CT;
- To use semantic patterns as proposed within SHN to map data elements in clinical models to a set of standardized ontologies and use this in the guideline representations;

More about the interface between ontologies and clinical guidelines can be read in Chapter 3.
4 Interface between EHR Information Model and Clinical Guidelines

We describe the relationship between information models and clinical guidelines, based on the overarching assumption that they are part of a larger whole: the architecture of health information technology. RM ODP and concurrent use are part of such approaches.

Inspired by the EN ISO 12967 (HISA) series, the work by Blobel (2010)\(^{24}\) on the Generic Component Model (GCM) and the overall parts of TOGAF\(^{25}\), clinical guidelines, described by their constituent processes, are placed on layer two of the four levels of such an architecture:

- Level one is the general business description of health care, or a specific domain in health care.
- Level two describes the processes.
- Level three is the information level
- Level four is the technology specification.

Although the above architectural approach offers overview and guidance, the exact relationships across the different layers are often a bit vague, unless a specific methodology is followed in detail. In particular, looking at other representations such as the Generic Component Model (Figure 4-1) the order of these layers is represented in different dimensions. It is beyond the scope of SHN workpackage 4 to go into the specifics of all these methodologies, but the actual goal of the work is to be very precise on the conceptual level of making relations explicit between the levels 2 and 3: clinical guidelines and information models.

Processes and guidelines sit on layer 2. They are described based on stakeholders, actors and their business goals. The overall guidance for their description is derived from the ContSys standard (ISO 13940). Clinical processes are in ContSys defined as comprehensive healthcare processes where subjects of care and healthcare professionals interact to handle specified health problems. Processes serve a purpose for one or more stakeholders and are carried out by various actors. None of them will be made explicit in this document.

The different purposes for use of clinical data each require a careful analysis on detailed level of the requirements and processes applied\(^{26}\). In particular, questions as determination of validity, relevance and reliability of the data apply here. Differences like the selection process of unit of study, such as level of a single problem, patient, or sample, or population levels will apply. Each process will determine which rules for aggregation need to be taken into account.

Hence, each purpose will come with a specific set of attributes and constraints for the data processing, such as entry, storage, management, presentation, communication, aggregation, etc. Conflict resolution can be necessary in case of different requirements. However, it is mostly dealt with by expressing a core set of data elements, and apply appropriate coding for data element and where applicable to the value sets. It is rarely possible to organize this for large data sets at once. Hence in the future there

---


will be a need for additional conflict resolution, were the ontology approach might help. Figure 4-2 illustrates overlap and differences in representing EHR data for different purposes (See figure 4-2 for identified purposes for data use). This can be seen as the playground for application of clinical models in various health care processes.

![Diagram](image)

Figure 4-2 Different purposes for data use requiring specific processes to manage them

Information models should be based on specific and defined clinical concept and clinical process models, that is, they should be clear for clinicians, be applicable to clinical practice, and fit in the common clinical processes. Clinical guidelines influence the performance of clinical processes by showing information concerning the medical knowledge. The knowledge management and the documentation of the performance of clinical processes should be as integrated as possible.

Such integration requires that the information in both clinical guidelines and the clinical documentation (i.e. EHR) use the same information model. The consequences of this are that on the one hand, clinical guidelines and protocols should be built up as preformed structures for clinical documentation applying one and the same information model and on the other hand, that the clinical documentation records the information used by clinical guidelines and protocols.

This can be done both from a top down perspective (an overarching structure for a disease is further specialised), or bottom up (a collection of clinical models (e.g. Detailed Clinical Models, templates or archetypes) is composed into a larger whole, such as a process description for disease management). A combination of these approaches is also possible.

**4.1 Challenges**

Within the challenges that exist for implementing the interface between information models and clinical guidelines we highlight the following:

- The information model must hold the patient information necessary to determine whether certain clinical guidelines criteria are satisfied

E.g. It is possible that “persistent highly elevated cholesterol” is never explicitly recorded but it is recorded as raw measurements and therefore has to be calculated from a range of values

Another example is that the information in the criteria used in a guideline may not be expressed at the same granularity as in the medical record (e.g. diabetes mellitus vs. insulin dependent diabetes mellitus).
Clinical models must be designed by sharing the same clinical process and concept models. Sharing the same concept model means that clinical models and clinical guidelines share the same ontology for representing their concepts.

E.g. In order to share the same process model, clinical guideline rules can be transformed into pre-formed documentation templates to be directly used in clinical practice (one example is presented in Annex for OpenEHR and the European Society of Cardiology heart failure guideline).

- EHR clinical information must be quality assured. It must faithfully represent data. For instance, as part of such quality, the information model must allow specifying how and when (both in time and in relation to the steps in the clinical process) the information was recorded as well as which codes were used and which clinical model was applicable during time of recording.

More about the interface between information models and clinical guidelines can be read in the following chapter, where it is illustrated with some guideline representations for decision support against data in the EHR. In addition to the relationships between clinical guidelines and ontologies, and clinical guidelines and information models, it is important to look at an important implementation aspect. That is the way information from the EHR can be retrieved for the various purposes as decision support, quality indicator, epidemiology and so on. The next chapter will explain how the rules usually present in a clinical guideline can be expressed such that both the requirements for the ontology – i.e. the SNOMED CT codes – and for the information model – i.e. the precise set of data elements and their relationships – can be identified and expressed in logical rules for both decision support and querying data.

5 Clinical Guidelines and their interface with Health records
5.1 Overview

In this chapter we analyse the interface between clinical guidelines, EHR information models and ontologies from a practical side, by selecting two different clinical guideline representation formalisms and using them for modelling a simple clinical guideline rule related with the heart failure diagnosis.

Numerous clinical practice guidelines (CPG) have been developed by governments and professional organizations are distributed in narrative form. However, their lack of formalization does not allow them to be interpreted automatically by computers, and their interpretation by human is sometimes ambiguous.

A number of guideline-based decision support systems are based on guideline models which are capable of formalising medical knowledge as electronic applications that can be executed (or enacted) to generate patient-specific recommendations for clinical decisions and actions. Such methods employ different representation formalisms and computational techniques, for example:

- Rule-based: Arden Syntax, GDL
- Logic-based: PROforma
- Network-based: PRODIGY
- Workflow (Petri Nets): GUIDE.

More details about computer-interpretable guideline models can be found in a study carried out by Mor Peleg et al. Here we will describe the representation of clinical guidelines by using two different formalisms: the Guideline Definition Language (GDL) and PROforma. We will focus on describing the capability of these formalisms for representing constraints within the clinical guideline rules that trigger the medical decision steps, and their linkage with the electronic health record (EHR). This linkage which we also refer to as interface is the focus of this chapter. According to Rector et al., the issues within the interface of clinical guidelines with the EHR can be classified as:

<table>
<thead>
<tr>
<th>Issue name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistency of meaning</td>
<td>All concepts within the clinical guideline rule and the medical record must be mapped to a common ontology (e.g. SNOMED CT), in</td>
</tr>
</tbody>
</table>

---

27 [http://www.openclinical.org/gmmintro.html](http://www.openclinical.org/gmmintro.html)
29 [http://www.openclinical.org/gmm_proforma.html](http://www.openclinical.org/gmm_proforma.html)
30 [http://www.openclinical.org/gmm_proforma.html](http://www.openclinical.org/gmm_proforma.html)
order to share their meaning (e.g. hypertension means the same in every system)

<table>
<thead>
<tr>
<th>Adequacy of structure</th>
<th>Inferential abstractions:</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is possible that it is never explicitly recorded “persistent highly elevated cholesterol” in the medical record, but it is recorded as raw measurements and therefore has to be inferred from a range of values</td>
<td></td>
</tr>
</tbody>
</table>

| Differences in granularity and classification axes: |
| The information in the rule criteria may not be expressed at the same granularity as in the medical record (e.g. diabetes mellitus vs. insulin dependent diabetes mellitus). |

| Unambiguous encapsulation of information between information and concept models | Pre and Post coordinated expressions (e.g. confirmed + hypertension vs. “confirmed hypertension”). Formal means to determine when two items of information are equivalent |

We will also describe the representation of clinical guideline rules as ontology-based queries, by following the formal-ontological approach proposed within our workpackage.

Finally we will comment on the benefits and drawbacks found for each of the above approaches and we will also sketch how GDL and PROforma could benefit from the semantic approach proposed.

### 5.2 Use Case: “Patient with atrial fibrillation”

As running example we will use one rule extracted from Deliverable 1.2 for patients with atrial fibrillation (see Table 4-1). Figure 4-3 shows its graphical decision tree.

**Table 4-1 Decision rule for patients with atrial fibrillation. Words underlined correspond to the constraints to be checked against the medical record. In red, the decision steps triggered by the rule when some constraint is satisfied.**

<table>
<thead>
<tr>
<th>Precondition: Patients with atrial fibrillation (AF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. If (Not LVEF measurement in Echocardiogram)</td>
</tr>
<tr>
<td>a. If (LVSD (Left ventricular systolic dysfunction) present)</td>
</tr>
<tr>
<td>i. Message: Treat according to LVSD guidelines</td>
</tr>
<tr>
<td>ii. If (severe LVSD &amp;&amp; and patient not unduly frail)</td>
</tr>
<tr>
<td>1. Message: Consider an implantable cardiac defibrillator after adequate trial of pharmacological treatment</td>
</tr>
<tr>
<td>b. if (Unknown LVSD)</td>
</tr>
<tr>
<td>i. Message: Alert clinician that LVSD is unknown and not LVEF present</td>
</tr>
<tr>
<td>2. If (LVEF measured &amp;&amp; LVEF &lt;35% &amp;&amp; patient not unduly frail)</td>
</tr>
<tr>
<td>a. Message: Consider an implantable cardiac defibrillator after adequate trial of pharmacological treatment</td>
</tr>
</tbody>
</table>
The above rule allows checking four different kinds of constraints against the EHR:

- Numeric comparison constraints (e.g. LVEF <35%);
- Constraints that include epistemic values (e.g. severe frailty)
- Constraints that include the presence or absence of information in the EHR (e.g. LVEF not measured)
- Constraints that exploit the taxonomic hierarchy of the terminology, here SNOMED CT (e.g. chronic LVSD as child of LVSD).

In order to consistently establish correspondences between the clinical concepts in the above rule and the corresponding ones in the medical record, we will use as common ontology SNOMED CT. Having such a common ontology will allow the consistent sharing of meaning of the concepts used (e.g. LVSD means the same in every system). Table 4-2 shows the mapping of the rule clinical concept names to SNOMED CT terms.

<table>
<thead>
<tr>
<th>Concept name</th>
<th>SNOMED CT term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atrial fibrillation</td>
<td>49436004 AF - Atrial fibrillation (Type:= clinical finding)</td>
</tr>
<tr>
<td>LVEF</td>
<td>250908004 LVEF - Left ventricular ejection fraction (Type:= observable entity)</td>
</tr>
<tr>
<td>severe</td>
<td>24484000 severe (Type:= qualifier value)</td>
</tr>
<tr>
<td>LVSD</td>
<td>134401001 left ventricular systolic dysfunction (Type:= clinical finding)</td>
</tr>
<tr>
<td>Frailty</td>
<td>248279007 frailty (Type:= clinical finding)</td>
</tr>
</tbody>
</table>

In addition to the four different kinds of constraints, we can observe in the above table that for representing “severe frailty” there is not an only one SNOMED CT term but two, what means that post-coordination of terms is required.

Next, we will apply GDL and PROforma formalisms to represent this rule.
5.3 The Guideline Definition Language (GDL)

GDL is a formal language to represent clinical knowledge for computerized decision support. It is agnostic regarding natural language and reference terminology, since it is based on the openEHR reference and archetype models.

GDL allows formalizing guideline rules but does not cover the process aspects of clinical guidelines. It is technology independent and therefore can be implemented by using any rule engine (e.g. JBoss Drools).

GDL rules are described together with metadata (e.g. use, purpose, misuse, version, etc.) in the same file together with rule definitions by following openEHR archetype like formalization. The rules use archetypes as input and output. The file has a definition section where GDL rule variables (i.e. concepts from the guideline rule which must be matched with archetype data elements) are mapped to archetype data elements defined by an archetype ID and path. In the same section it is possible also to define preconditions that have to be met before the rule can be executed. Rules are defined in a separate rule section and are composed by the clauses when, then and priority. Each term locally defined in a GDL guideline file has a local identifier starting with “gt” (guideline term).

In the following we will describe the formalization of our running example rule with GDL. First, we have established the correspondences between the rule variables and the archetypes used. We have used the following openEHR four archetypes:

<table>
<thead>
<tr>
<th>Archetype</th>
</tr>
</thead>
<tbody>
<tr>
<td>openEHR-EHR-OBSERVATION.imaging_exam-Echocardiography.v1</td>
</tr>
<tr>
<td>openEHR-EHR-EVALUATION.problem-diagnosis.v1</td>
</tr>
<tr>
<td>openEHR-EHR-EVALUATION.exclusion_problem_diagnosis.v1</td>
</tr>
<tr>
<td>openEHR-EHR-EVALUATION.recommendation.v1</td>
</tr>
</tbody>
</table>

The correspondences are established in the binding section of the rules file as shown for the binding of the LVSD rule variable (see Error! Not a valid bookmark self-reference.). For the LVSD we will use the openEHR-EHR-EVALUATION.problem-diagnosis.v1 archetype. In that case the three archetype data elements relevant for mapping the LVSD are: (1) the diagnosis itself (e.g. LVSD); (2) the status (e.g. unknown) and the severity (e.g. severe). The three of them are identified by their corresponding path within the archetype. We have used the GDL reserved word predicate for this binding in order to establish the mapping of the local term LVSD to its corresponding code in an external terminology (in our case SNOMED CT). We have also indicated the template where the corresponding SNOMED CT values are defined (template_id = "diagnosis_snomedct").

---

Once the bindings for all rule variables have been provided, the next step is to define the guideline rule. Following, we show the formalization of the rule that triggers the message “Alert clinician that LVSD is unknown and not LVEF present” (see Table 4-4). Since GDL rules are language independent, the alerts can be produced in several languages.

Table 4-3 Example of GDL binding of rule variables with archetype data elements

```plaintext
definition = (GUIDE_DEFINITION) <
archetype_bindings = <
  ["gt0004"] = (ARCHETYPE_BINDING) <
    archetype_id = "openEHR-EHR-EVALUATION.problem-diagnosis.v1"
    domain = "EHR"
    elements = <
      ["gt0005"] = (ELEMENT_BINDING) <
        path = "/data[at0001]/items[at0002.1]"
      >
      ["gt0006"] = (ELEMENT_BINDING) <
        path = "/data[at0001]/items[at0005]"
      >
      ["gt0014"] = (ELEMENT_BINDING) <
        path = "/data[at0001]/items[at0.32]"
      >
    >
predicates = "/data[at0001]/items[at0002.1] is_a local::gt0005|LVSD|,...>
    template_id = "diagnosis_snomedct"
>>>>
```

The when clause checks if the rule variable that correspond to LVSD diagnosis has “unknown” as status value and then triggers the recommendation to the clinician. For each rule a priority value can be indicated in order to establish the rules execution order.

For defining the GDL rules and the bindings we have used the GDL editor. The editor also allows you to test the guideline rule defined with fictitious patient data. In the following screenshot we see how the tool triggers the message “Consider an implantable cardiac defibrillator after adequate trial of pharmacological treatment” when the following patient data are recorded:

- Atrial fibrillation
- LVEF measurement 20 %
- Mild Frailty
As main benefits of the GDL formalism we highlight:

- Its independency of reference terminologies and natural languages due to the use of archetypes as input and output of rules
- It allows the binding between local concepts used in the rules and concepts from reference terminologies
- It can be implemented by any rule engine, since it is independent of a concrete technology

The use of archetypes allows that the rules can be expressed by using different information models and not just the openEHR one. Actually the GDL editor is being adapted for supporting CIMI models, which are basically based on a simplified version of the openEHR reference model.

It also allows binding local rule terms to lists of concepts or a refset in different target reference terminologies.

As main drawbacks or limitations we have found the following:

- Their dependency to the openEHR reference and archetype model

The use of archetypes as input and output of GDL rules offers a lot of flexibility but at the same time inherits all their legacy problems, such as the lack of semantic relationships across their data elements and the famous boundary problem (no clear border definition between what is represented by the information model and the ontology/terminology) (see Chapter 2). Bindings of rule variables to archetype data elements are dependent on the structure of the archetype. Ideally this binding should be driven by finding the concepts independently of the structure used to represent them.

Regarding the inherited boundary problem, how to express negation is one of the most used examples. In the above rule, in order to express it a different archetype as the problem/diagnosis one has to be used (i.e. exclusion problem diagnosis). This may cause confusion when checking in a rule if certain diagnosis exists or not:
Frailty exists means that there is no frailty. Besides Frailty severity has to be interpreted together with the excluded diagnosis statement (i.e. no severe frailty exists) although no explicit relationship exists between them within the archetype (only structural relationship).

- No support for postcoordinated expressions (there is a plan to include it in the future)
- Limited support for reasoning (only taxonomic subsumption; could be also extended in the future if the use case requires it)

The GDL specifications allow term bindings between local gt codes to externally defined pre-coordinated or post-ordinated concepts. The current GDL reference implementation (GDL Editor) does not however use a full-blown terminology server for reasoning. There is development plan to integrate the GDL editor/rules engine with CTS2 terminology server in the near future.

Although taxonomic subsumption reasoning can be included in the tool, it does not support more advanced reasoning mechanism such as equivalence detection or dealing with pre/post coordinated expressions.

### 5.4 PROforma

PROforma is a formal knowledge representation language designed to capture the content and structure of a clinical guideline in a form that can be interpreted by a computer. It is a task-based guideline representation format, which means that it expresses clinical processes in terms of tasks and not rules (as GDL does). Task-based systems attempt to contextualise rules within explicit and intuitive models of the clinical process with the aim of combining the programmatic richness of procedural representations with the logical clarity of declaratively expressed knowledge.

The PROforma language structure is based on a clinical process model referred to as the domino model (see Figure 4-6). The domino model describes a relationship between actions, decisions, beliefs, plans, problem goals and candidate solutions and the inference and processes linking them. This model can be reified into a minimal set of executable generic tasks (i.e. enquiries, decisions, plans and actions). It is from these tasks and the logical constructs associated with each task, that the PROforma language is derived.

---

A PROforma guideline consists of a set of tasks and a set of data items whose values are accessible to those tasks. There are four kinds of tasks:

- Decisions represent the choice points in a guideline
- Actions represent some action performed by a human or software agent
- Enquiries represent the acquisition of data from some external source
- Plans allow other tasks to be composed in order to perform some complex operation

Each task has a set of properties (e.g. precondition, postcondition, description, trigger, etc.) for which values may be defined. For creating and editing tasks in PROforma we have used the Tallis Composer. Figure 4-7 depicts an example of task and how data are collected by this task.

Several data sources can be connected with a task. Figure 4-8 shows how the task collects data about the severe frailty or not of the patient. PROforma does not have a proper information model and allows specifying the binding of clinical guideline rule variables and the corresponding concepts in the record by using existing data definitions (e.g. booleanType in Figure 4-8), or defining them from scratch.
Figure 4-8 Example of data source definition for representing the existence or not of severe frailty

PROforma allows creating process descriptions in which several tasks are executed successively. The link between tasks is known as scheduling constraint and will determine the order of their execution at runtime. Figure 4-9 shows an example of scheduling constraints defined between the tasks used to represent the atrial fibrillation guideline rule example. Arrows represent the execution order (LVSD enquiry will only run after the LVEF enquiry is completed).

![Figure 4-9 Scheduling constraints defined between the tasks used to represent the atrial fibrillation guideline rule. Green rhombus represents enquiry tasks; Pink circles represent decision tasks; Blue squares represent action tasks.](image)

Decision tasks make a choice between several different options, known as candidates (e.g. two different recommendations). Each candidate requires the specification of a set of arguments for and/or against them. An argument consists of a logical expression (e.g. LVSD_degree = “Severe”) and a support mode (e.g. for) specifying the conditions under which a candidate must be selected or discarded (e.g. raise certain recommendation).
Finally, as in the case of GDL, it allows you to test the guideline rules definition with fictitious patient data as shown in the following figure for the patient with the following recorded data:

- NO LVEF measurement
- LVSD present and severe
- No severe frailty

As main benefits of PROforma we highlight:

- It expresses clinical processes in terms of tasks and not rules what makes it easier to use and more expressive (e.g. allows creating sequenced, iterative and concurrent tasks)
- It provides an easy-to-use graphical interface that allows clinicians to see graphically how the different tasks within a guideline rule are executed
• It provides high expressivity for describing tasks and conditions (e.g. it allows dealing with temporal criteria)
• It allows linking the tasks with external EHR data repositories
• It has been used in a variety of research projects and a smaller number of clinical applications.

As main drawbacks of the PROforma formalism we highlight:

• It does not provide a proper information model but allow the reuse and definition of what they call data definitions.

Marcos et al.\(^{34}\) propose the use of archetypes for representing the concepts of the guideline rules as it is done by the GDL formalism. Health records are mapped to a set of archetypes, and the latter ones to the guideline rules variables. The first mapping from health records to a set of archetypes allows performing a series of operations or abstractions on the EHR in order to avoid the “impedance mismatch”\(^{35}\) between the EHR and the clinical decision support system (CDSS), among others.

• No support for managing external terminologies such as SNOMED CT (no support for any kind of reasoning; it does not allow binding guideline rule concepts to external terminology terms).

The mentioned work proposed by Marcos et al\(^{34}\) allows the use of external terminologies such as SNOMED CT and certain reasoning capabilities (e.g. taxonomic subsumption). However no more advanced reasoning support is provided (e.g. dealing with pre/post-coordinated terms).

5.5 SHN Semantic Approach – Clinical guideline rules as ontology-based queries

In this section we describe how clinical guideline rules could be expressed as queries and run against structured EHR data by following the semantic approach developed within our workpackage. We do not provide any guideline formalism or the implementation of a task network or process model. Our goal is to show the benefits of working with an ontology-based representation in order to exploit clinical data (here by means of queries).

In the SHN approach clinical models’ data elements and their values are mapped to ontology-based expressions represented in OWL DL or RDF, by following a set of predefined ontology design patterns (viz. semantic patterns) (cf. Deliverable 4.3). When clinical models are instantiated with data, patterns are also instantiated by following the correspondences previously defined. This transformation steps correspond to layers 1, 2 and 3 from the proposed semantic-driven architecture (see Figure 1-2).

---


In the following we will show how our atrial fibrillation use case is modelled by following the SHN semantic driven approach. By following the architecture described in Section 1.3, we have to describe the steps carried out within the next five layers:

- Layer 1: Structured Heterogeneous Data
- Layer 2: Semantic Mapping
- Layer 3: Semantic Mediator
- Layer 4: Virtual Homogeneous Data
- Layer 5: Application

We depart from a set of heterogeneously structured clinical data (Layer 1). Here, we have used the three following OpenEHR archetypes for structuring our fictitious data instances:

<table>
<thead>
<tr>
<th>Archetype</th>
</tr>
</thead>
<tbody>
<tr>
<td>openEHR-EHR-OBSERVATION.imaging_exam-Echocardiography.v1</td>
</tr>
<tr>
<td>openEHR-EHR-EVALUATION.problem-diagnosis.v1</td>
</tr>
<tr>
<td>openEHR-EHR-EVALUATION.exclusion-problem_diagnosis.v1</td>
</tr>
</tbody>
</table>

As fictitious data instances we will represent a patient with: Atrial Fibrillation and LVEF 20% and No severe frailty and LVSD present.

Then, we will carry out the following four steps (in brackets its corresponding layer in the semantic architecture):

1. Map clinical model data elements and their values to semantic patterns (Layer 2)
2. Represent our fictitious patient data as OWL DL instances of the patterns (Layer 3)
3. Formulate the rule for the atrial fibrillation use case as a set of SPARQL queries (Layer 4)
4. Run the queries against the test data instances (Layer 5)

We have used the following semantic patterns described as Subject-Predicate-Object (SPO) triples (cf. Deliverable 4.2 and 4.3):

**Top-level pattern I_CS_PT:**

The *information about clinical situation* pattern (I_CS_PT) can be used to represent some piece of information about a particular clinical situation of the patient. Clinical situations, as described in\(^{36}\), correspond to SNOMED CT clinical findings. The pattern consists of the following SPO (Subject-Predicate-Object) triples, described as enumerated in the left column of the table:

- S1: Type of clinical situation in focus (e.g. *heart failure, smoking*, etc.)
- S2: Process performed to acquire the information. (e.g. *diagnostic measures, physical examination, history taking*, etc.)

---

- S3: Any epistemic and contextual aspect that qualifies or modifies the information related with the clinical situation in focus (e.g. severity, temporal context, finding context (absence, certainty, etc.), etc.)

<table>
<thead>
<tr>
<th>#N</th>
<th>Subject</th>
<th>Predicate</th>
<th>Cardinality</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>shn:informationitem</td>
<td>'describes situation'</td>
<td>1..*</td>
<td>shn:ClinicalSituation (CS_PT)</td>
</tr>
<tr>
<td>S2</td>
<td>shn:informationitem</td>
<td>'results from process'</td>
<td>1..*</td>
<td>btl:Process (CP_PT)</td>
</tr>
<tr>
<td>S3</td>
<td>shn:informationitem</td>
<td>'has attribute'</td>
<td>0..*</td>
<td>shn:InformationAttribute</td>
</tr>
</tbody>
</table>

Table 4-5 Semantic pattern I_CS_PT

In the above table, the Object part of the triple may contain the name of another pattern in brackets, which means that the main pattern can be composed with that one.

Examples of the application of this top-level pattern are Diagnoses, Co-morbidities, Symptoms, Allergies, etc. The following tables show its application for representing “Atrial Fibrillation” and “Severe LVSD”. The Object part of the triples has been constrained by following the cardinality and value constraints indicated by the pattern.

<table>
<thead>
<tr>
<th>#N</th>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>shn:informationitem</td>
<td>'describes situation'</td>
<td>sct:AtrialFibrillation</td>
</tr>
<tr>
<td>S2</td>
<td>shn:informationitem</td>
<td>'results from process'</td>
<td>sct:DiagnosticProcess</td>
</tr>
</tbody>
</table>

Table 4-6 Atrial fibrillation example based on I_CS_PT pattern

<table>
<thead>
<tr>
<th>#N</th>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>shn:informationitem</td>
<td>'describes situation'</td>
<td>sct:LVSD</td>
</tr>
<tr>
<td>S2</td>
<td>shn:informationitem</td>
<td>'results from process'</td>
<td>sct:DiagnosticProcess</td>
</tr>
<tr>
<td>S3</td>
<td>shn:informationitem</td>
<td>'has attribute'</td>
<td>sct:Severe</td>
</tr>
</tbody>
</table>

Table 4-7 Severe LVSD example based on I_CS_PT pattern

**Top-level pattern I_NCS_PT:**

The information about no clinical situation pattern (I_NCS_PT) is defined as a specialisation of I_CS_PT and can be used to represent the absence of a particular patient clinical situation. It consists of the following SPO triples:

- S1: Clinical situation in focus (e.g. heart failure, smoking, etc.)
- S2: Process performed to acquire the information. (e.g. diagnostic, physical examination, history taking, etc.)
- S3: Any epistemic and contextual aspect that qualifies or modifies the information related with the clinical situation in focus (e.g. severity, certainty, temporal context, etc.).
- S4: The information attribute that represents the absence of the clinical situation in focus. This triple is coloured in grey since it further specialises the triple S3 (shn:InformationItem ‘has attribute’ 0..* shn:InformationAttribute) of the parent pattern (I_CS_PT) for representing the absence aspect. Thus, the predicate ‘has situation context’ specialises the predicate ‘has attribute’ and constrains its value to sct:Absent, which is placed as subclass of shn:InformationAttribute in the ontology framework. The cardinality “1..1” indicates that this triple is obligatory when this pattern is instantiated. Note that this is a meta-description, so it has no existential import at the level of individuals.

<table>
<thead>
<tr>
<th>#N</th>
<th>Subject</th>
<th>Predicate</th>
<th>Cardinality</th>
<th>Object</th>
</tr>
</thead>
</table>
Examples of application of this top-level pattern are Diagnosis of absence of X, No Symptom X, No Allergy X, etc. The following table shows its application for representing “No severe frailty”. The Object part of the triples has been constrained by following the cardinality and value constraints indicated by the pattern.

<table>
<thead>
<tr>
<th>#N</th>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>shn:InformationItem</td>
<td>'describes situation'</td>
<td>sct:Frailty</td>
</tr>
<tr>
<td>S2</td>
<td>shn:InformationItem</td>
<td>'results from process'</td>
<td>sct:SymptomEvaluation</td>
</tr>
<tr>
<td>S3</td>
<td>shn:InformationItem</td>
<td>'has attribute'</td>
<td>sct:Severe</td>
</tr>
<tr>
<td>S4</td>
<td>shn:InformationItem</td>
<td>'has situation context'</td>
<td>sct:Absent</td>
</tr>
</tbody>
</table>

Table 4-9 No severe frailty example based on I_NCS_PT pattern

Top-level pattern OB_PT:

The observation result pattern represents the result of observing or assessing some quality related with the health status of the patient.

It consists of the following SPO triples:

- O1: The quality observed / assessed (e.g. mass intake)
- O2: The result of the observation / assessment (e.g. 15 cigarettes per day)
- O3: The observation / assessment process performed to acquire the information.

Examples of application of this top-level pattern are Clinical Test Result, Physical Examination Result, measurement result, etc. The following table shows its application for representing “LVEF 20%”. The Object part of the triples has been constrained by following the cardinality and value constraints indicated by the pattern.

<table>
<thead>
<tr>
<th>#N</th>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>O1</td>
<td>shn:ObservationResult</td>
<td>'describes quality'</td>
<td>sct:LVEF</td>
</tr>
<tr>
<td>O2</td>
<td>shn:ObservationResult</td>
<td>'has observed value'</td>
<td>VR_PT</td>
</tr>
<tr>
<td>O3</td>
<td>shn:ObservationResult</td>
<td>'results from process'</td>
<td>sct:Assessment</td>
</tr>
</tbody>
</table>

Table 4-11 LVEF 20 % example based on OB_CS_PT

<table>
<thead>
<tr>
<th>#N</th>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>btl:ValueRegion</td>
<td>'has value'</td>
<td>20</td>
</tr>
<tr>
<td>V2</td>
<td>btl:ValueRegion</td>
<td>'has units'</td>
<td>sct:Percentage</td>
</tr>
</tbody>
</table>

Table 4-12 LVEF 20 % example based on VR_PT

Next, we will apply the 4 steps previously described for modelling the atrial fibrillation use case.
Map clinical model data elements and their values to semantic patterns

In the following table the correspondences between the clinical model data elements and the semantic pattern triples are established. In bold the variable parts of the patterns that will be later instantiated with the patient data.

### Table 4-13 Definition of correspondences between clinical models and semantic patterns

<table>
<thead>
<tr>
<th>ARCHETYPE_ID / ADL Path</th>
<th>Data Element / Value</th>
<th>Triple representation</th>
<th>#N</th>
</tr>
</thead>
<tbody>
<tr>
<td>openEHR-EHR-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EVALUATION.problem‐diagnosis.v1</td>
<td>Diagnosis / Disease</td>
<td>shn:informationItem `describes situation' shn:ClinicalSituation</td>
<td>#S1</td>
</tr>
<tr>
<td>/data[@0001]/items[@0002.1]</td>
<td></td>
<td>shn:informationItem `results from process' sct:DiagnosticProcess</td>
<td>#S2</td>
</tr>
<tr>
<td>openEHR-EHR-EVALUATION.exclusion-problem_diagnosis.v1</td>
<td>Diagnosis / Disease</td>
<td>shn:informationItem `describes situation' shn:ClinicalSituation</td>
<td>#S1</td>
</tr>
<tr>
<td>/data[@0001]/items[@0003.1]</td>
<td></td>
<td>shn:informationItem `results from process' sct:DiagnosticProcess</td>
<td>#S2</td>
</tr>
<tr>
<td>/data[@0001]/items[@0008]</td>
<td>Diagnosis / Severity</td>
<td>shn:informationItem `has attribute' shn:Severity</td>
<td>#S3</td>
</tr>
<tr>
<td>openEHR-EHR-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OBSERVATION.imaging_exam-Echocardiography.v1</td>
<td>LVEF / Value, Units</td>
<td>shn:ObservationResult `describes quality' btl:Quality</td>
<td>#O1</td>
</tr>
<tr>
<td>/data[@0001]/events[@0002]/data[@0003]/items[@0015]/items[@0016]/items[@0017.2]</td>
<td></td>
<td>shn:ObservationResult `has observed value' btl:ValueRegion</td>
<td>#O2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>btl:ValueRegion `has value' xml:double</td>
<td>#V1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>btl:ValueRegion `has units' shn:Units</td>
<td>#V2</td>
</tr>
</tbody>
</table>

Represent our fictitious patient data as OWL DL instances of the patterns

First we will instantiate the pattern triples with the fictitious patient and then we will transform them into OWL DL by following the correspondences shown in the next table:

### Table 4-14 Correspondences between the Triple-based pattern representation and OWL DL

<table>
<thead>
<tr>
<th>Predicate</th>
<th>OWL DL expression</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>describes situation</code></td>
<td>SUBJ subclassOf shn:isAboutSituation only OBJ</td>
</tr>
<tr>
<td><code>describes quality</code></td>
<td>SUBJ subclassOf shn:isAboutQuality only OBJ</td>
</tr>
<tr>
<td><code>results from process</code></td>
<td>SUBJ subclassOf btl:isOutcomeOf some OBJ</td>
</tr>
<tr>
<td><code>has attribute</code></td>
<td>SUBJ subclassOf btl:hasAttribute some OBJ</td>
</tr>
<tr>
<td><code>has observed value</code></td>
<td>SUBJ subclassOf btl:Quality and btl:projectsOnto some OBJ</td>
</tr>
<tr>
<td><code>has value</code></td>
<td>SUBJ subclassOf btl:isRepresentedBy only (shn:hasValue some OBJ)</td>
</tr>
<tr>
<td><code>has units</code></td>
<td>SUBJ subclassOf btl:isRepresentedBy only (shn:hasAttribute some OBJ)</td>
</tr>
</tbody>
</table>

By applying the above correspondences the following OWL DL instances are produced for our fictitious patient:

### Table 4-15 OWL DL instances generated for the fictitious patient data

<table>
<thead>
<tr>
<th>Data Element / Value</th>
<th>OWL DL representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis / Atrial fibrillation</td>
<td>shn:informationItem and shn:isAboutSituation only sct:AtrialFibrillation and btl:isOutcomeOf some sct:DiagnosticProcess</td>
</tr>
<tr>
<td>Diagnosis / LVSD</td>
<td>shn:informationItem and shn:isAboutSituation only sct:LVSD and btl:isOutcomeOf some sct:DiagnosticProcess</td>
</tr>
<tr>
<td>Diagnosis / No severe frailty</td>
<td>shn:informationItem and shn:isAboutSituation only (shn:ClinicalSituation and not (btl:includes sct:Frailty)) and btl:isOutcomeOf some sct:DiagnosticProcess and shn:hasAttribute some sct:Severe</td>
</tr>
<tr>
<td>LVEF</td>
<td>shn:ObservationResult</td>
</tr>
</tbody>
</table>
Formulate the rule as a set of SPARQL queries

We will use SPARQL for formulating the queries corresponding to the running example guideline rule. The queries follow the same structure provided by the semantic patterns used. The following three queries are required for executing the guideline rule, each one corresponding to a recommendation as shown in Figure 4-3:

(Q1) information about patients with LVEF < 35 % and no severe frailty
(Q2) information about patients with no LVEF measurement and LVSD unknown
(Q3) information about patients with no LVEF and severe LVSD and no severe frailty

For each patient, the following queries have to be executed:

```sparql
#Q1 SELECT ?ItemLVEF ?ItemFrailty
WHERE{
  ?ItemLVEF
  a [ a owl:Class ;
    owl:intersectionOf( shn:ObservationResult
      [ a owl:Restriction ;
        owl:onProperty shn:isAboutQuality
          [ a owl:Class ;
            owl:intersectionOf ( sct:LVEF
              [ a owl:Restriction ;
                owl:onProperty btl2:projectsOnto ;
                owl:someValuesFrom [ a owl:Class ;
                  owl:intersectionOf ( bt2:ValueRegion
                    [ a owl:Restriction ;
                      owl:onProperty btl2:isRepresentedBy ;
                      owl:allValuesFrom [ a owl:Class ]
                    ]
                  ]
                ]
              ]
            ]
          ]
        ]
      ]
    ).

  ?ItemFrailty
  a [ a owl:Class ;
    owl:intersectionOf( shn:InformationItem
      [ a owl:Restriction ;
        owl:onProperty btl2:isOutcomeOf ;
        owl:someValuesFrom sct:DiagnosticProcedure
      ]
    ]
  ].
}
```
#Q2
SELECT ?ItemLVEF
WHERE{
?ItemLVEF a [a owl:Class ;
owl:intersectionOf(shn:ObservationResult
[ a owl:Restriction ;
owl:onProperty shn:isAboutQuality ;
owl:someValuesFrom sct:LVEF])]
}

SELECT ?ItemLVSD
WHERE{
?ItemLVSD a [a owl:Class ;
owl:intersectionOf(shn:InformationItem
[ a owl:Restriction ;
owl:onProperty btl2:isOutcomeOf ;
owl:someValuesFrom sct:DiagnosticProcedure
]
[ a owl:Restriction ;
owl:onProperty shn:hasInformationObjectAttribute ;
owl:someValuesFrom sct:Unknown
]
[ a owl:Restriction ;
owl:onProperty shn:isAboutSituation ;
owl:allValuesFrom sct:LVSD
]
})

#Q3
SELECT ?ItemLVEF
WHERE{
?ItemLVEF a [a owl:Class ;
owl:intersectionOf(shn:ObservationResult
[ a owl:Restriction ;
owl:onProperty shn:isAboutQuality ;
owl:someValuesFrom sct:LVEF])]
}

SELECT ?ItemLVSD
WHERE{
?ItemLVSD a [a owl:Class ;
owl:intersectionOf(shn:InformationItem
[ a owl:Restriction ;
owl:onProperty btl2:isOutcomeOf ;
owl:someValuesFrom sct:DiagnosticProcedure
]
[ a owl:Restriction ;
owl:onProperty shn:hasInformationObjectAttribute ;
owl:someValuesFrom sct:Unknown
]
[ a owl:Restriction ;
owl:onProperty shn:isAboutSituation ;
owl:allValuesFrom sct:LVSD
]
})

SELECT ?ItemFrailty
WHERE{
?ItemFrailty a [a owl:Class;
owl:intersectionof(shn:InformationItem
[ a owl:Restriction ;
owl:onProperty btl2:isOutcomeOf ;
owl:someValuesFrom sct:DiagnosticProcedure
]
[ a owl:Restriction ;
owl:onProperty shn:hasInformationObjectAttribute ;
owl:someValuesFrom sct:Severe
]
[ a owl:Restriction ;
owl:onProperty shn:isAboutSituation ;
owl:allValuesFrom sct:LVSD
]
])

SELECT ?ItemFrailty
WHERE{
?ItemFrailty a [a owl:Class;
owl:intersectionof(shn:InformationItem
[ a owl:Restriction ;
owl:onProperty btl2:isOutcomeOf ;
owl:someValuesFrom sct:DiagnosticProcedure
]
[ a owl:Restriction ;
owl:onProperty shn:hasInformationObjectAttribute ;
owl:someValuesFrom sct:Severe
]
[ a owl:Restriction ;
owl:onProperty shn:isAboutSituation ;
owl:allValuesFrom [ a owl:Class ;
owl:intersectionOf ( shn:ClinicalSituation
] ]
}
For the case in which the query asks for information about patients without LVEF measurement, we have to explicitly ask for the existence of an instance of information about LVEF. In case of not retrieving anything we interpret it as no LVEF measurement performed.

The use of logic-based reasoning would allow also dealing with some heterogeneity between the guideline rule and the EHR data such as in case in the record states chronic LVSD instead of LVSD. Since the first is a kind of LVSD, a query for retrieving all information instances of patients with LVSD, would retrieve also those encoded as chronic.

The OWL DL representation proposed also allows the representation of epistemic concepts such as “severe”, by following predefined patterns as previously described and detecting equivalences between pre/post-coordinated expressions (e.g. if a pre-coordinated term for “severe frailty” would exist, it could detect its equivalence with the corresponding post-coordinated representation).

**Run the queries against the test data instances**

When we run the previous queries against the fictitious patient data, we only get an answer for #Q1. It means that the recommendation “Consider an implantable cardiac defibrillator after adequate trial of pharmacological treatment” will be raised.

**Comments to the semantic approach:**

As main drawbacks of this approach we highlight:

- The need of mapping clinical model data elements and their values to semantic patterns as a previous step;
- The need of creating SPARQL queries for modelling the guideline rules.

Both issues need of appropriate tools that support the modeller, and hide SPARQL and OWL DL or RDF related implementation details, in a similar way to existing clinical model editors.

As main benefits we highlight:

- The independence of the method of a particular clinical modelling approach (e.g. openEHR / ISO 13606, HL7, etc.)
- The smooth integration of ontology-based terminologies and information models by using the same representation formalism (OWL DL or RDF).
- The support of DL reasoning for working with advanced queries and post-coordinated expressions
- The flexibility and expressivity offered by the SPARQL query language

As in the case of GDL, this formalization does not cover the process aspects of clinical guidelines, which have to be implemented additionally (e.g. the order in which the queries are executed).
5.6 Benefits for GDL and PROforma of implementing the SHN semantic approach

We think that both GDL and PROforma could benefit from the proposed semantic approach.

In both cases it requires that EHR data can formalized as OWL DL instances of the proposed ontological framework.

In case of GDL, openEHR clinical models used as input of guideline rules could be mapped to semantic patterns, and these later instantiated with the data.

In case of PROforma, since it does not provide a proper information model and neither is based on a standardized EHR representation, as in the GDL case we foresee two possible alternatives:

- Mapping directly EHR data to a set of semantic patterns in order to get their OWL DL representation.
- Mapping EHR data to a set of archetypes as proposed by Mar et al.\(^3\), and the latter ones to a set of semantic patterns

For both GDL and PROforma cases, providing EHR data formalized according to a set of formal ontologies allows:

- EHR data can be accessed in a homogeneous way by different clinical decision support systems.
- It provides external reasoning support that facilitates the mapping of clinical guideline rule variables with archetype concepts (e.g. dealing with pre/post coordination of concepts and dealing with concepts at different levels of abstraction).
- Both clinical guideline rule variables and archetype concepts are semantically related and not just related by certain structure, what facilitates its mapping and processing
6 Summary and conclusions

In this document we have explored the interface of the EHR with systems that implement clinical practice guideline specifications (CPGs). For that we have studied the interfaces or relationships between clinical practice guidelines, EHR information models and domain ontologies.

Although the three of them are different, they do interact each other and this interaction is not trivial since there are overlaps between them: e.g. CPGs interact with EHR information models and they also provide their own ones; CPGs refer to specific types of clinical process and they can be heterogeneously described by different systems (e.g. ContSys or local domain ontologies).

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

In our study we have analysed the main challenges that arise at each interface: (1) EHR information models – Domain ontologies; (2) Clinical practice guidelines – Domain ontologies and (3) Clinical practice guidelines - EHR information models. For each one we have explored how these challenges could be addressed by the semantic-driven approach proposed within WP4. We have also described how this interface is addressed within existing approaches. The use of the WP4 semantic-driven approach is to act as harmonization tool, without claiming to substitute any of them.

As an example of the above, for the interface between EHR information models and domain ontologies we have provided an analysis of the state of the art of current terminology binding approaches and described their relationship with the approach proposed within our workpackage describing how existing binding approaches could benefit of such an approach and vice-versa.

Another example, for the interface between process descriptions within clinical guidelines and domain ontologies, we have described our initial experiments in showing how ContSys concepts under the Health condition category could be represented as ontology concepts within the proposed ontological framework. Our hypothesis is that formalizing them by using ontology concepts will help to implement the standard by facilitating its integration with existing EHR information models and ontology-based terminologies.

Many of the results provided in this document are experimental and need further elaboration when considered worth to be done. There is also interesting future work to conduct such as the alignment of ContSys concepts with the SNOMED CT terminology, by using the ontological framework proposed within our work.

In order to better understand and analyse the challenges within each of the above interfaces, we exemplify them by describing their implementation by two different clinical guideline formalisms, GDL and PROforma, and based on a simple guideline rule extracted from the ones provided by WP1 in deliverable 1.2, for possible heart failure patients which have atrial fibrillation. We have focused on describing the capability of these formalisms for representing constraints within clinical guideline rules that trigger the medical decision steps, and their linkage with electronic health records. We also describe how clinical guideline rules could be expressed as queries and run against structured EHR data
by following the semantic approach developed within our workpackage. We do not provide any guideline formalism or the implementation of a task network or process model. Our goal is to show the benefits of working with an ontology-based representation in order to exploit clinical data (here by means of queries).

For both guideline formalisms we sketch how they could benefit from a more semantic-driven approach. Further elaboration on this is expected as result of the implementation of the evaluation plan carried out in conjunction with WP2.
7  Annex

7.1  Heart Failure Summary and ESC Guidelines (Jean Marie Rodrigues)

The following is the pragmatic European definition of heart failure in common use with the ambiguities of the narrative definition as expressed by John Cleland (WP3)

1. Typical Symptoms and signs of heart failure (see below but particularly breathlessness limiting a normal amount of exertion)
2. Evidence of structural heart disease (usually with echocardiography, but this can be surprisingly subtle and the test requires a highly skilled operator)
3. Expected response to appropriate treatment (particular an improvement in symptoms with diuretics)
4. Increases in plasma concentrations of natriuretic peptides – a useful screening blood test, which can be done by anyone who can take a blood sample. A normal test indicates either that the patient does not have a heart problem or that treatment has been highly effective. Increased levels may be due to heart problems but can also be due to kidney problems.

The New York Heart Association functional classification based on severity of symptoms and physical activity:

Class I: No limitation of physical activity. Ordinary physical activity does not cause undue breathlessness, fatigue, or palpitations.

Class II: Slight limitation of physical activity. Comfortable at rest, but ordinary physical activity results in undue breathlessness, fatigue, or palpitations.

Class III: Marked limitation of physical activity. Comfortable at rest, but less than ordinary physical activity results in undue breathlessness, fatigue, or palpitations.

Class IV: Unable to carry on any physical activity without discomfort. Symptoms at rest can be present. If any physical activity is undertaken, discomfort is increased.

7.2  Heart Failure Summary

The SHN heart failure summary is the core dataset that contains essential information to be shared across sectors and disciplines in order to optimize HF management. It has been elaborated by WP1 and a more detailed description of its data items can be found in deliverable 1.2.

HFS has been modelled by using an OpenEHR template which can be found in the project tab of the OpenEHR Clinical Knowledge Manager (CKM)37 tool, under the name “EU-SHN Heart Failure Summary”. This template uses a subset of OpenEHR archetypes, where textual descriptions for each of the information items included are provided. Fig 1 depicts the main structure of the template. Each of the sections shown in the above figure can be decomposed in subsections.

37http://www.openehr.org/ckm/
7.3 Clinical Guidelines for Chronic Heart Failure

Definitions

We will consider the following definitions from ISO 13940 as the basis to identify the different inference models or Problem Solving Method to address specifically the Clinical Guidelines.

Clinical guidelines (ISO 13940):

*Set of systematically developed statements to assist the decisions made by healthcare actors about healthcare activities to be performed with regard to specified health issues*

Clinical pathway (ISO 13940):

*Structured pattern for the healthcare activities in a core care plan to be adopted in care plans for subjects of care having similar health conditions with a predictable clinical course*

Care plan (ISO 13940)

*Dynamic, personalized plan including identified needed healthcare activities, health objectives and healthcare goals, relating to one or more specified health issues in a healthcare process*

Clinical process (ISO 13940)

*Healthcare process encompassing all healthcare provider activities and other prescribed healthcare activities that addresses health problems*
7.4 European Society of Cardiology (ESC) Chronic Heart Failure Guideline

As stated in the preamble of the updated ESC guidelines on chronic heart failure published in the European Heart Journal (2012) 33, 1787–1847 ESC guidelines summarize and evaluate all available evidence at the time of the writing process, on a particular issue with the aim of assisting physicians in selecting the best management strategies for an individual patient, with a given condition, taking into account the impact on outcome, as well as the risk–benefit ratio of particular diagnostic or therapeutic means. Guidelines are no substitutes, but are complements, for textbooks and cover the European Society of Cardiology (ESC) Core Curriculum topics.

The two main issues for Heart Failure are diagnosis and treatment. Figure 7-2 Heart Failure Diagnosis shows the diagram for the diagnosis.

![Heart Failure Diagnosis Diagram]

**Figure 7-2 Heart Failure Diagnosis**
7.5 Examples

We have selected two examples from the ESC HF Guidelines

Heart Failure Diagnosis

IF (Breathlessness or fatigue, tiredness and increased time to recover for ordinary physical activity) Then
Do measurement of BNP or NT-Pro BNP

IF (BNP>100 pg/mL or NT-ProBNP>300pg/mL) is True Then
Do Echocardiography

IF (LVEF at echocardiography < 35 %) is True Then
Diagnosis is HF-REF
Determine aetiology

IF (BNP>100 pg/mL or NT-ProBNP>300pg/mL) is True Then
Do Echography
IF (LVEF at echocardiography < 35 %) Then
Diagnosis is HF-REF
Determine aetiology

IF (LVEF at echocardiography > 35 % and < 50 %)is True Then
Diagnosis is HF-PEF
Determine aetiology

IF (LVEF at echocardiography >50 %) Then
Do ECG

IF (ECG completely normal) Then
Diagnosis HF unlikely

IF (ECG shows sino atrial or atrio ventricular or y or LV intra ventricular hypertrophy intraventricular anomaly or LV hypertrophy or Q waves ) then
Diagnosis is HF-PEF
Determine aetiology

IF (BNP< 100 pg/mL and NT-ProBNP<300pg/mL>)is True then
Diagnosis HF unlikely

- Breathlessness, fatigue and tiredness are SCT primitives or synonyms and can be found in a EHR based on the HF summary: presentation and symptoms
- Increased time to recover from ordinary physical activity is not present as such
- Measurement of BNP or NT Pro-BNP is not present as such in HF summary: physical exam
• Measurement of LVEF by Echocardiography can be extracted from HF summary: Echography by the SHN method (Del4.2)

• 5 ECG normal or abnormal can be extracted from HF summary: ECG by the SHN method (Del4.2)
# Heart Failure Treatment options

**IF** (Chronic symptomatic systolic heart failure NYHA II–IV) is true

**THEN** treatment is (Diuretics AND (ACE inhibitor OR ARB Angiotensin receptor blocker)) AND beta blocker

---

**IF** Chronic symptomatic systolic heart failure NYHA II–IV is False

**THEN** treatment is (Diuretics AND (ACE inhibitor OR Angiotensin receptor blocker)) AND beta-blocker

---

**IF** Chronic symptomatic systolic heart failure NYHA II–IV is TRUE

**THEN** treatment is (Diuretics AND (ACE inhibitor OR Angiotensin receptor blocker)) AND beta-blocker AND Mineralocorticoid receptor antagonist

---

**IF** Chronic symptomatic systolic heart failure NYHA II–IV is False

**THEN** treatment is (Diuretics AND (ACE inhibitor OR Angiotensin receptor blocker)) AND beta-blocker AND Mineralocorticoid receptor antagonist

---

**IF** Chronic symptomatic systolic heart failure NYHA II–IV is TRUE

**IF** left ventricular ejection fraction <= 35% is False

**THEN** treatment is (Diuretics AND (ACE inhibitor OR Angiotensin receptor blocker)) AND beta-blocker AND Mineralocorticoid receptor antagonist

---

**IF** Chronic symptomatic systolic heart failure NYHA II–IV is True

**IF** left ventricular ejection fraction <= 35% is True

**IF** Sinus rhythm and Heart Rate > 70 beats/min is False

**THEN** treatment is (Diuretics AND (ACE inhibitor OR Angiotensin receptor blocker)) AND beta-blocker AND Mineralocorticoid receptor antagonist
7.6 GDL capabilities to formalize ESC/Cardiac Failure Guidelines (Jean Marie Rodrigues, Rong Chen)

Capacity for Modeling the Heart Failure Summary

Model for Heart Failure Summary is already created in openEHR Reference Model (RM). In the RM, Archetypes represent clinical knowledge as shareable clinical content models.

Presentation and symptoms

Reason of encounter can be recorded as free-text by instantiating archetype openEHR-EHR-EVALUATION.reason_for_encounter.v1.

The Medical history is stored with openEHR-EHR-OBSERVATION.story.v1. All diagnosis can be recorded as instance of openEHR-EHR-EVALUATION.problem-diagnosis.v1. Primary diagnosis, Primary procedure, Co-morbidity, Complication are recorded as occurrences properties.

Note: In SHN, a diagnosis has a given certainty (hypothetic, confirmed, ...) while in openEHR is has a Status (provisional, confirmed, ...).

Symptoms such as Breathlessness, Chest pain, Fatigue, Swelling of ankles, legs, feet or belly, Palpitation and Syncope can be represented using openEHR-EHR-CLUSTER.symptom.v1 this cluster can contain further symptom clusters to capture disease-associated symptom details.

NYHA Class

The New York Heart Association (NYHA) Functional Classification is modeled in openEHR-EHR-OBSERVATION.nyha_heart_failure_score.v1

Physical exam

Examination procedures are represented using openEHR-EHR-OBSERVATION.*

<table>
<thead>
<tr>
<th>Exam</th>
<th>Instance of</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body weight</td>
<td>openEHR-EHR-OBSERVATION.body_weight.v1</td>
</tr>
<tr>
<td>Height</td>
<td>openEHR-EHR-OBSERVATION.height.v1</td>
</tr>
<tr>
<td>Blood Pressure</td>
<td>openEHR-EHR-OBSERVATION.blood_pressure.v1</td>
</tr>
<tr>
<td>Oedema</td>
<td>openEHR-EHR-CLUSTER.oedema.v1</td>
</tr>
<tr>
<td>Transcutaneous oxygen saturation</td>
<td>openEHR-EHR-OBSERVATION.indirect_oximetry.v1</td>
</tr>
<tr>
<td>Blood count</td>
<td>openEHR-EHR-OBSERVATION.lab_test-full_blood_count.v1</td>
</tr>
<tr>
<td>ElectroCardiography</td>
<td>openEHR-EHR-OBSERVATION.ecg.v1</td>
</tr>
</tbody>
</table>
Echocardiography | openEHR-EHR-OBSERVATION.imaging_exam.v1
Lung function | openEHR-EHR-OBSERVATION.pulmonary_function.v1
Coronary angiography | openEHR-EHR-OBSERVATION.imaging_exam.v1

**Laboratory test request**

Instructions can be represented using openEHR-EHR-INSTRUCTION.*

<table>
<thead>
<tr>
<th>Instruction</th>
<th>Instance of</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laboratory Request</td>
<td>openEHR-EHR-INSTRUCTION.request-lab_test.v1</td>
</tr>
<tr>
<td>Health Service Request</td>
<td>openEHR-EHR-INSTRUCTION.request.v1</td>
</tr>
<tr>
<td>Medication Order</td>
<td>openEHR-EHR-INSTRUCTION.medication_order.v1</td>
</tr>
</tbody>
</table>

**Rule writing using GDL**

Guideline Definition Language (GDL) is a rule-based language that uses archetype bindings, rule expressions and term bindings for core elements.

Rules can be written in natural-language. This is possible using archetypes that makes indirect reference to data elements in the rules.

Rule for proposing a new Medication order:

**Rule Add Ivabradine**

**When**
- Element **MR antagonists** exists
- Element **Date (time) of first administration** is less than **NYHA event time**
- Element **Ivabradine** does not exist
- Element **LVEF<=35%** equals to **Present**
- Element **Sinus rhythm and HR>=70/min** equals to **Present**

**Then**
- Create **openEHR-EHR-INSTRUCTION.medication.v1**
- Set element **Recommended medication** to **Ivabradine**
Heart Failure Diagnostic Rules

Rule for Heart Failure Diagnostic (in acute onset and non-acute onset):

**Rule Set Heart failure likely in Acute onset**

When

\[
(\text{Element ECG abnormal exists}) \quad \text{or} \quad
(\text{Element NT-pro BNP is greater than or equals to 300 pg/ml})
\]

Then

Set element **Rationale** to "Heart failure likely"
Set element **Recommendation** to "Echocardiography"

**Rule Set Heart failure unlikely in Acute onset**

When

Element **Rationale** does not exist
Element **ECG normal** exists

\[
(\text{Element NT-pro BNP is less than 300 pg/ml}) \quad \text{or} \quad
(\text{Element BNP is less than 100 pg/ml})
\]

Then

Set element **Rationale** to "Heart failure unlikely"
Rule Set Heart failure likely in Non-acute onset
When
   ( ( 
    Element ECG abnormal exists
   ) or ( 
    ( ( 
     Element NT-pro BNP is greater than or equals to 125 pg/ml
    ) or ( 
     Element BNP is greater than or equals to 35 pg/ml
    ) ) 
   ) )
Then
Set element Rationale to "Heart failure likely"
Set element Recommendation to "Echocardiography"

Rule Set Heart failure unlikely in Non-acute onset
When
   Element Rationale does not exist
   Element ECG normal exists
   ( ( 
    Element NT-pro BNP is less than 125 pg/ml
   ) or ( 
    Element BNP is less than 35 pg/ml
   ) )
Then
Set element Rationale to "Heart failure unlikely"