Computable Declarative Representation of Clinical Assessment Scales in EHRs

Mercedes Argüello Casteleiro[1], Nicolas Matentzoglu[1], Bijan Parsia[1], and Sebastian Brandt[1]

[1]School of Computer Science, The University of Manchester, United Kingdom
m.arguello@computer.org

Abstract. Clinical assessment scales, such as the Glasgow coma scale, are a core part of Electronic Health Records (EHRs). However, fully representing them in an OWL ontology is challenging: In particular, the determination of a score from patient's observations and clinical findings requires forms of aggregation and addition which are either tedious in OWL 2 or merely impractical due to combinatorial explosion. To solve this problem, we propose to separate the representation of the structure and content of an assessment scale from its enactment with the former being captured in OWL 2 and the latter being determined by a SPARQL query. The paper reports the results of a systematic review of 104 well-established clinical assessment scales along with the performance of the SPARQL queries proposed when executed with the query engine ARQ for Jena.

Keywords. Clinical assessment scales, OWL 2, SPARQL 1.1, HL7 CDA, SNOMED CT

1   Introduction

Over the years health measurement methods have become a consolidated part of healthcare practice and research. In the literature, there is an increasing amount of studies that test the validity and reliability of measurements of health. For example, within the field of mental health, there is an ongoing debate about the optimal use of clinical rating scales and outcomes assessment tools [1]. Despite the large volume of research into clinical decision making in general, and clinical assessment scales in particular, there is still a lack of full integration between Electronic Health Records (EHRs) and evidence-based medicine. Some of the unaddressed challenges for the systematic incorporation of clinical assessment scales within EHRs are the following:

- SNOMED CT has been acknowledged as the most comprehensive, multilingual clinical healthcare terminology in the world [2]. SNOMED CT (January 2013) contains 890 assessment scales concepts within the *staging and scales* hierarchy, but these are just the name of the scales. They are unrelated to other SNOMED biomedical concepts from other top-level hierarchies, such as *clinical finding* or *observable entity*. Thus, to fully represent a clinical assessment scale like Glasgow

coma scale [3], the SNOMED CT biomedical concept 386554004|Glasgow coma scale (assessment scale)| should be related to its components, such as 281396004|Glasgow Coma Score motor response subscore (observable entity)|, and to its scale items, such as 85157005|Decorticate posture (finding)|. Overall, to represent assessment scales components as well as assessment scale items, both SNOMED CT pre- and post-coordinated expressions are needed.

- There are connections between elements of a specific terminology (e.g. SNOMED CT) and an information model for EHRs (e.g. HL7 V3). These connections are now widely recognised and known as the *terminology binding process*. Indeed, the *assessment scale result pattern* is an example of a common pattern found in EHRs. Therefore, it is essential not only to representation of such pattern, but also to establish a suitable mechanism to retrieve and manipulate its components and items. The latter is a cornerstone in EHRs standards, which are XML-based. Numerous studies have highlighted the problem of extracting information from patients' EHRs [4]. As Hristidis et al. [5] highlights information discovery on XML documents is not adequate due to the domain-specific semantics and the frequent references to external information sources like dictionaries.
- For assessment scales like the Apgar score [6] that assesses the health of a newborn; the Barthel index [7] that measures the performance in activities of daily living; the APACHE II [8] that estimates mortality in critically ill patients, the Glasgow coma scale [3] that assesses level of consciousness; and the Hamilton anxiety rating scale [9] that rates the severity of a patient's anxiety, the aggregation of the total score cannot be straightforward modelled in OWL 2 [10], as it is either tedious or merely impractical due to combinatorial explosion.

This research study proposes to separate the representation of the structure and content of a scale from its enactment with the former being captured in OWL 2 and the latter being determined by a SPARQL 1.1 [11] query. Although the *terminology binding process* is properly documented, its exploitation with the aim of facilitating query building has not been studied before.

2 Preliminaries

The following subsections offer an overview of the foundations of the research presented.

2.1 SNOMED CT

SNOMED CT is formulated in the description logic EL++ [12], which corresponds to the OWL 2 EL profile [10]. The January 2013 edition of SNOMED CT terminology contains 890 assessment scales concepts within the *staging and scales* hierarchy, which has no defining attributes assigned [13]. Thus, an ontology that models assessment scales needs to extend the current SNOMED CT ontology in OWL 2 to connect names of scale-based assessments within the *staging and scales* hierarchy with biomedical pre- and post-coordinated expressions that use biomedical concepts from

other top-level hierarchies, such as *clinical finding*, *procedure*, *observable entity*, or *situation with explicit context*.

SNOMED CT concepts are pre-coordinated when a single concept identifier is used. SNOMED CT post-coordinated expressions use multiple concept identifiers and are underpinned by a compositional grammar. This research study adopts the compositional grammar for SNOMED CT expressions in HL7 V3 [14].

On the one hand, this study is aligned with the increasing interest and discussion of post-coordination [15], due to the impossibility of enumerating all combinations of medical concepts without causing a combinatorial explosion [16]. On the other hand, several authors like Nadkarni [17] have recognized the serious limitations of a pure terminological approach to model assessment scales like the Apgar score [6] without combinatorial explosion or major information loss. Thus, an ontology that envisions to model assessment scales needs to go beyond EL++ and allows concept descriptions where disjunctions and data ranges are supported. Hartela et al. [18] already hinted some of the limitations of EL++ for building large complex terminologies. Indeed, they built the National Cancer Institute (NCI) Thesaurus [19] using the description logics Ontylog language. It should be noted that SNOMED CT is also developed in Ontylog. However, data ranges that are needed for modeling the assessment scales are not in the Ontylog language [18]. Data ranges can be modeled in OWL 2. However, the aggregation (mathematical addition) of the total score cannot be modeled in OWL 2. This research study has looked at the SPARQL 1.1 Query Language [11] for representing assessment scales in a declarative way as well as computing them, i.e. aggregating the scores.

2.2 Terminology binding process: how to bind SNOMED CT with HL7 V3

HL7 Version 3 (HL7 V3 for short) is a lingua franca used by healthcare computers to talk to other computers [2]. The HL7 Clinical Document Architecture (CDA) is a document markup standard that specifies the structure and semantics of a clinical document, e.g. progress note, for the purpose of exchange [20]. Worldwide HL7 CDA is the most widely adopted application of HL7 V3 [2], and it is one of the standardization efforts towards providing the interoperability of EHRs. It is now recognized that healthcare terminologies like SNOMED CT [13] and information models like HL7 V3 can not be separated. There is a terminology binding process [21] that specifies how to establish connections between elements of a specific terminology and an information model. The HL7 and IHSTDO report [22] provides guidelines on how to bind SNOMED CT with HL7 V3. The report also acknowledges that assessment scales share certain characteristics and introduces an *assessment scale result pattern*.

The *assessment scale result pattern* is an example of a common pattern. The HL7 and IHSTDO report [22] defines common patterns as: "clinical statements that are used frequently, often in many different applications, for a wide variety of communication use cases". A formal representation for the *assessment scale result pattern* requires to take into account: 1) assessment scales have one or more component observations, which can be aggregated to provide an overall score like in the Glasgow coma scale [3]; and 2) assessment scale component observations can be represented:

a) as an observable entity with result; or b) as an assertion of a clinical finding. Either or both may be needed depending on the concrete scale.

Hence, an ontology that models assessment scales needs to incorporate axioms that represent the *assessment scale result pattern* along with its components and items. There is a range of reasoners that can enable the automatic validation of the axioms of the common patterns, i.e. clinical statements, related to the clinical assessment scales.

2.3   SPARQL 1.1

SPARQL 1.1. is a W3C candidate recommendation towards a standard query language for the Semantic Web. RDF is a directed, labeled graph data format for representing information in the Web [11]. SPARQL is defined in terms of the W3C's RDF data model and will work for any data source that can be mapped into RDF. SPARQL contains capabilities for querying an RDF Schema or an OWL 2 model to filter out individuals with specific characteristics. This is particularly relevant to match specific views facilitating medical tasks.

SPARQL can be used for real-time querying and retrieval of information from clinical and research datasets. This study exploits SPARQL 1.1 to query the OWL 2 ontology instances related to clinical assessment scales in EHRs, and therefore, it plays pivotal role to organize and organize clinical information for assessment scales.

SPARQL has four query forms. The SELECT form returns all, or a subset of, the variables bound in a query pattern match [11]. Therefore, knowing before hand the assessment scale components or items, specific SELECT queries can be created for an assessment scale. This study has found particularly useful when creating those SELECT queries for assessment scales: a) the IF function form; and b) the Sum SPARQL algebra operator.

Although the *terminology binding process* is well known and properly documented, its exploitation to create more generic queries has not been studied before. This study demonstrates how it is possible to build even more abstract queries so that the aggregation of the total score for different assessment scales can be performed with a single terse query.

3   Using OWL 2 to represent assessment scales in EHRs

As Dolin et al [20] emphasises: "The CDA R2 model is richly expressive, enabling the formal representation of clinical statements (such as observations, medication administrations, and adverse events) such that they can be interpreted and acted upon by a computer". In HL7 CDA R2 there is a tie-in between a document section and HL7 Reference Information Model (RIM) [23] Act classes, like Observation. The HL7 RIM together with the HL7 V3 data types [2] supply a powerful mechanism to incorporate concepts from standard coding systems, such as SNOMED CT or LOINC [24], into CDA clinical statements.

The *assessment scale result pattern* from the HL7 and IHSTDO report [22] reuses two common patterns proposed for the HL7 RIM Observation entries. This study proposes to distinguish clearly between these two patterns:

• *The assessment scale components* – these are individual observations that should appear in the SNOMED CT *observable entity* hierarchy.
• *The assessment scale items* – these should appear in the SNOMED CT *clinical finding* hierarchy or in the SNOMED CT *situation with explicit context* hierarchy, which enables asserting the absence of a clinical finding. Typically, coded scores are assigned to assessment scale items.

Taking into account the work described in [7], it is possible to map HL7 RIM Observation entries from a HL7 CDA document to instances of an OWL 2 ontology. This research reuses two ontologies: 1) an OWL ontology for HL7 CDA that appears in [20]; and 2) a SNOMED CT ontology in OWL 2 created by means of the Simple SNOMED Module Extraction [27]. Additionally: a) an afresh built ontology for LOINC [24]; b) the OWL ontology for HL7 CDA was refactored to tackle more effectively with the *terminology binding process*; and 3) the SNOMED CT ontology was extended to facilitate building SNOMED CT post-coordinated expressions.

As OWL allows the import of the contents of entire ontologies, the three above-mentioned ontologies are imported into each HL7 CDA document (e.g. consultation note), i.e. the OWL 2 file that contains the clinical information about a particular patient. Table 1 shows two OWL 2 instances in the Manchester OWL Syntax [28] that correspond to two different common patterns from the HL7 and IHSTDO report [22].

Table 1. Examples of HL7 RIM Observation entries formally represented in OWL 2

Common Pattern	OWL 2 instance

Observable entity with result – Example of scale component	Individual: actObservation2
	  Types:
	    ActObservation
	  Facts:
	    actCd 'Glasgow coma score motor response subscore (observable entity)',
	    actMoodCd EVN,
	    XFRM actObservation4,
	    actClassCd OBS,
	    actObservationValueInt "3"

Assertion of a clinical finding – Example of scale item	Individual: actObservation4
	  Types:
	    ActObservation
	  Facts:
	    actMoodCd EVN,
	    actClassCd OBS,
	    actCd ASSERTION,
	    actObservationValueCD 'Decorticate posture (finding)'

To represent assessment scale components as well as assessment scale items, SNOMED CT pre- and post-coordinated expressions are needed. Table 2 illustrates

the SNOMED CT concepts and expressions that are needed to represent Glasgow coma scale [3] components and items.

Table 2. Examples of SNOMED CT pre- and post-coordinated expressions

Assessment Scale Component or Item - Textual description	SNOMED CT pre- or post-coordinated expression
Assessment scale component Best Motor Response (M) – 6 grades	**SNOMED CT concept (pre-coordinated expression)** 281396004|Glasgow Coma Score motor response subscore|
Assessment scale item Localizing response to pain	**SNOMED CT post-coordinated expression** 450847001|Responds to pain|: 363713009|Has interpretation| = 255471002|Localized|

Table 3 shows how the SNOMED CT concepts and expressions from table 2 are represented in the Manchester OWL Syntax [28]. Following Rector and Iannone [15], RoleGroups have been used, which were originally designed as an extension to the Ontylog description logic [18].

Table 3. Examples of SNOMED CT pre- and post-coordinated expressions in OWL 2

SNOMED CT pre- or post-coordinated expression	Manchester OWL Syntax
SNOMED CT pre-coordinated expression 281396004|Glasgow Coma Score motor response subscore|	Class: 'Glasgow coma score motor response sub-score (observable entity)' SubClassOf: 'Component of Glasgow coma scale (observable entity)'
SNOMED CT post-coordinated expression 450847001|Responds to pain|: 363713009|Has interpretation| = 255471002|Localized|	Class: 'Localizing response to pain (SNOMED CT Expression)' EquivalentTo: 'Responds to pain (finding)' and (RoleGroup some ('Has interpretation (attribute)' some 'Localized (qualifier value)')) SubClassOf: 'SNOMED CT Expression (SNOMED CT Expression)'

The extension of the SNOMED CT ontology incorporates three classes: *assessment scale score*; *assessment scale component*; and *assessment scale item*. For these classes and their subclasses disjunctions and data ranges are allowed, and thus, they go beyond EL++. Figure 1 shows an example of a concept definition that is an *assessment scale component subclass*. SNOMED CT concepts belong to the SNOMEDCT namespace, while SNOMED CT expressions belong to the SNOMEDCT-EXPext namespace (the extended SNOMED CT ontology) and are underpinned by the compositional grammar for SNOMED CT expressions in HL7 V3 [14].

```
Class: SNOMEDCT-EXPext:SCT-EXPext_000000007

  Annotations:
    rdfs:label "GCS Best Motor Response (M) - 6 grades (SNOMED CT Expression extension)"@en

  EquivalentTo:

    SNOMEDCT:SCT_281396004
    and (SNOMEDCT:SCT_280434008 some xsd:integer[>= 1 , <= 6])
    and (SNOMEDCT-EXPext:SCT-EXPext_000000010
    or SNOMEDCT-EXPext:SCT-EXPext_000000011
    or SNOMEDCT-EXPext:SCT-EXPext_000000012
    or SNOMEDCT-EXPext:SCT-EXPext_000000014
    or SNOMEDCT-EXPext:SCT-EXPext_000000015
    or SNOMEDCT-EXPext:SCT-EXPext_000000016)

  SubClassOf:
    SNOMEDCT-EXPext:SCT-EXPext_000000003
```

Fig. 1. Example of assessment scale component subclass – The Glasgow coma scale (GCS) Best Motor Response (M) is a component observation and has 6 scale items: obeys commands; localises to central pain; withdraws from pain; flexion to pain; extension to pain; no response to painful stimuli.

4 Using SPARQL 1.1 to assign and calculate scores for assessment scales in EHRs

Two types of SPARQL 1.1 SELECT queries have been considered: 1) queries that assign the scores to the assessment scale items; and b) queries that calculate the total score for the scale. While the previous are scale-dependent, the latter can be formulated in a more abstract way and be scale-independent.

4.1 Assign scores for assessment scales items in EHRs

It is feasible to create SPARQL 1.1 SELECT queries that assign points (scores) for the assessment scale items. These queries make use of the IF function to assign a numeric value, i.e. the points or scores, to SNOMED CT *findings* (pre- or post-coordinated expressions). Thus, a repository of SPARQL queries for representing assessment scales in a declarative way can be created, where the queries stored are scale-dependent and will automatically assign the scores to the assessment scale items of a particular assessment scale.

The SPARQL 1.1 SELECT query from figure 2 illustrates the automatically assignment of points (scores) to assessment scales items that are related to a particular assessment scale component, i.e. the subscale 281396004|Glasgow coma score motor response subscore (observable entity)|. This subscale is one of the three subscales, i.e. assessment scale component, for the Glasgow coma scale [3].

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX RIMV3OWL: <http://wopeg.he.agfa.be/rules/RIMV3OWL#>
PREFIX SNOMEDCT: <http://www.ihtsdo.org/SNOMEDCT#>
PREFIX SNOMEDCT-EXP: <http://www.ihtsdo.org/SNOMEDCT-EXP#>

SELECT (
if(?type = SNOMEDCT-EXP:SCT-EXP_450847001__363713009__2667000, 1,
if(?type = SNOMEDCT:SCT_23073007 , 2,
if(?type = SNOMEDCT:SCT_85157005 , 3,
if(?type = SNOMEDCT-EXP:SCT-EXP_247755007__42752001__22253000 , 4,
if(?type = SNOMEDCT-EXP:SCT-EXPext_000000015, 5,
if(?type =SNOMEDCT-EXP:SCT-EXP_302000008___365452001 , 6, 0)))))) as
?motor)
{
      ?Act1 RIMV3OWL:actCd SNOMEDCT:SCT_281396004;
            RIMV3OWL:transformation ?Act2.

      ?Act2 RIMV3OWL:actObservationValueCD ?type.
}
```

Fig. 2. Example of SPARQL 1.1 SELECT query to assign scores – This query assigns points (scores) 1 to 6 for the six scale items of the Glasgow coma scale (GCS) Best Motor Response (M). Each scale item is either a SNOMED CT pre-coordinated expression or a SNOMED CT post-coordinated expression.

4.2   Aggregate scores for assessment scales in EHRs

SPARQL has algebra operators. Aggregates defined in version 1.1 of SPARQL [11] are COUNT, SUM, MIN, MAX, AVG, GROUP_CONCAT, and SAMPLE. Aggregates can be useful for obtaining a result that is computed over a group of solutions, instead of a single solution. Indeed, for calculating the total score of a scale, like the Glasgow coma scale [3], it is necessary to add the points (scores) for the three behavioral responses, i.e. assessment scale components, that Glasgow coma scale assesses: the 281395000|Glasgow coma score eye opening subscore (observable entity)|, the 281397008|Glasgow coma score verbal response subscore (observable entity)|, and the 281396004|Glasgow coma score motor response subscore (observable entity)|.

Figure 3 shows the SPARQL SELECT query that retrieves the assessment scale components and uses the SPARQL algebra operator SUM to calculate the total score for the scale 248241002|Glasgow coma score (observable entity)|.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX RIMV3OWL: <http://wopeg.he.agfa.be/rules/RIMV3OWL#>
PREFIX SNOMEDCT: <http://www.ihtsdo.org/SNOMEDCT#>
PREFIX SNOMEDCT-EXP: <http://www.ihtsdo.org/SNOMEDCT-EXP#>

SELECT (sum(?Val) AS ?summedval)
WHERE {
    ?Act RIMV3OWL:actCd SNOMEDCT:SCT_248241002;
                    RIMV3OWL:isDerivedFrom ?Observation.

    ?Observation RIMV3OWL:actObservationValueInt ?Val.
}
```

Fig. 3. SPARQL 1.1 SELECT query to aggregate the total score for Glaucoma coma scale [3] – This query exploits the terminology binding process that exists between SNOMED CT and HL7 CDA R2. The query is tailor-made for the scale 248241002|Glasgow coma score (observable entity)| and it can not be used for any other assessment scale.

As it can be seen from figure 3, this SPARQL 1.1 SELECT query only refers to one single SNOMED CT pre- or post-coordinated expression, in the example given the 248241002|Glasgow coma score (observable entity)|, and all the other assessment scale components involved in the calculation are not explicitly mentioned (see figure 3). This is a substantive difference with respect the first SPARQL SELECT query from figure 2, where all the assessment scale items are explicitly mentioned along with the relevant assessment scale component. Therefore, the SPARQL SELECT query from figure 3 is easier to maintain than the SPARQL SELECT query from figure 2, particularly if new components are added to the assessment scale or even if the assessment scale components are re-assigned, for example when a SNOMED CT concept becomes obsolete, the SPARQL SELECT query still will work.

Intuitively, the style of query in figure 3 is the "right way" to cope with assessment scales. It hosts the representation of the scale in its natural home, the ontology itself, which supports definition of particular items in the scale as well as additional constraints on the values or connections to other concepts. The representation of the map between observations and findings inside the query separates this critical bit of information about the assessment scale from the rest of the modeling of the scale. Figure 3 isolates the computation of the scale from the representation of the scale. Indeed, it is possible to build even more abstract queries so that all assessments can be performed with a single terse query.

Figure 4 shows the SPARQL 1.1 SELECT query that can calculate the total score of more than one assessment scale that may appear in EHRs (e.g. CDA R2 consultation note).

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX RIMV3OWL: <http://wopeg.he.agfa.be/rules/RIMV3OWL#>
PREFIX SNOMEDCT: <http://www.ihtsdo.org/SNOMEDCT#>
PREFIX SNOMEDCT-EXP: <http://www.ihtsdo.org/SNOMEDCT-EXP#>
PREFIX SNOMEDCT-EXPext: <http://uk.ac.manchester.cs.owl/SNOMEDCT-EXPext#>

SELECT (?Act2 as ?Scale) (SUM(?Val) as ?Score) {
SELECT  ?Act1 ?Act2 ?Observation ?Code ?Val ?EKA
WHERE {
    ?Act1 RIMV3OWL:actCd ?Act2;
        RIMV3OWL:isDerivedFrom ?Observation.

   ?Act2 owl:equivalentClass ?EKA.

   ?EKA rdfs:subClassOf SNOMEDCT-EXPext:SCT-EXPext_000000020.

    ?Observation RIMV3OWL:actObservationValueInt ?Val;
            RIMV3OWL:actCd ?Code.
}}
GROUP BY ?Act2
```

Fig. 4. SPARQL 1.1 SELECT query to calculate the total score of assessment scales that can be scored using simple addition and that appear within a HL7 CDA document – This query exploits the terminology binding process that exists between SNOMED CT and HL7 CDA R2. This query is "abstract" as it refers to an OWL class whose subclasses are the assessment scale scores that can be calculated by using simple addition.

5   Execution of SPARQL 1.1 SELECT queries

To test the computational feasibility of the SPARQL 1.1 SELECT queries proposed in the previous section, a simple test harness on top of the query engine ARQ [29] for Jena [30] is implemented.

The anonymised consultation notes (HL7 CDA R2 documents) selected for the test harness exemplify the use of the *assessment scale result pattern* from the HL7 and IHSTDO report [22]. Through this pattern, assessment scales that differ in complexity are incorporated within the consultation notes, such as: Apgar score [6], the Barthel index [7], the APACHE II [8], the Glasgow coma scale [3], and the Hamilton anxiety rating scale [9]. It should be noted that *Glasgow coma scale* appears within *APACHE II* as an assessment scale item. However, *Glasgow coma scale* is quite complex in itself with a modelling that involves the incorporation of: 1 assessment scale score; 3 assessment scale components; and 15 assessment scale items. Each of these 15 assessment scale items is a HL7 RIM Observation entry with SNOMED CT pre- or post-coordinated expressions.

Jena [30] is based on Java. In ARQ [29], a SPARQL 1.1 SELECT query under the RDF entailment regime [31] is created from a string using the `QueryFactory`. The query and model (or RDF dataset) to be queried are passed to `QueryExecutionFactory`. The `ResultSetFormatter` class has methods to write out the SPARQL Query Results XML Format [32]. Time counters are incorporated into the Java code to estimate the execution time needed per query.

```xml
<?xml version="1.0"?>
<sparql xmlns="http://www.w3.org/2005/sparql-results#">
 <head>
  <variable name="Scale"/>
  <variable name="Score"/>
 </head>
 <results>
  <result>
   <binding name="Scale">
    <uri>http://www.ihtsdo.org/SNOMEDCT#SCT_420195005</uri>
   </binding>
   <binding name="Score">
    <literal datatype="http://www.w3.org/2001/XMLSchema#integer">40</literal>
   </binding>
  </result>
  <result>
   <binding name="Scale">
    <uri>http://www.ihtsdo.org/SNOMEDCT#SCT_248241002</uri>
   </binding>
   <binding name="Score">
    <literal datatype="http://www.w3.org/2001/XMLSchema#integer">7</literal>
   </binding>
  </result>
 </results>
</sparql>
```

Fig. 5. Example of SPARQL Query Results in XML Format for query in figure 4

Using a MacBook Pro with a processor 2.7 GHz Intel Core i7 and 16GB of RAM, the SPARQL 1.1 SELECT query from figure 4, which is the more generic query and allows calculating the total score of more than one assessment scale, is executed in 301 milliseconds for a consultation note (HL7 CDA R2 document) for a particular patient. For this patient, the concrete results of the query appear in figure 5 in the XML Format [32], where the total score of two assessment scales are being calculated: 248241002|Glasgow coma score (observable entity)|; and the 420195005|Component of Barthel index score (observable entity)|.

6 Results of the review of 104 clinical assessment scales

There are several ways to classify health measurements [33]: a) by their function taking into account the purpose or application of the method, i.e. functional classifications; b) by the focus on their scope also known as descriptive classifications; c) by

technical aspects like the techniques used to record the information, these are known as methodological classifications; d) by their scope or the range of topics they cover, i.e. descriptive classifications.

In this paper, the 104 clinical assessment scales selected by McDowell [33], which are the leading health measurement methods, are systematically reviewed to determine: 1) how widely applicable is the representation of the structure of the assessment scale proposed; 2) the variety of queries needed to assign the scores to the assessment scale items; and 3) to what extend the algebra operators for SPARQL 1.1 can be useful to calculate the *health index* or *profile* of an assessment scale.

A hierarchy of mathematical adequacy for the assignment of numerical scores has been considered: 1) *nominal or categorical scales*, which use numbers as mere labels for categories; 2) *ordinal scales*, where numbers are an indication of the quantity of the characteristic being measured and their assignment is arbitrary; 3) *interval scales*, where it is feasible to interpret differences in scores, as well as performing addition, subtraction and average calculation; 4) *ratio scales*, where numbers are used in measuring physical characteristics, e.g. time.
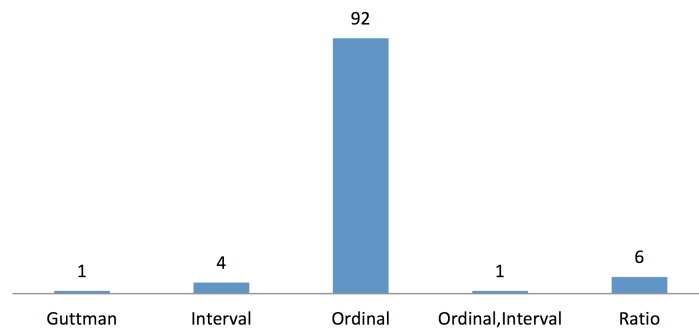

Fig. 6. Numerical characteristics of the 104 clinical assessment scales reviewed

Figure 6 shows the above-mentioned numerical characteristics of the 104 assessment scales reviewed. These numerical characteristics form the basis for assigning the scores to the assessment scale items of a particular assessment scale. 88% of the assessment scales are ordinal (92 out of 104), very few are interval scales (4 out of 104) or ratio scales (6 out of 104). Only the McGill Pain Questionnaire [34] is considered as both ordinal and interval. Only the Physical Self-Maintenance Scale [35] is judged as Guttman, i.e. nominal or categorical scale.

Another characteristic that this study considers is the depth or number of levels, which unfolds the complexity of the assessment scale and is paramount in the representation of the structure of the assessment scale proposed. Depth 1 is a single item value. From a modelling point of view, this is the simplest. A typical representative of assessment scales of depth 1 is the visual analogue rating scales (VAS) [36], as they provide a simple way to record subjective estimates of pain intensity. Depth 2 is typically a flat list of single items that are aggregated to obtain an overall score. For example, the Barthel index [7] has depth 2 and is used within some of the consultation

notes for the test harness. Depth 3 usually presents the items grouped into subscales. For example, the Glasgow coma scale [3] and Apgar score [6] have depth 3 and are used within some of the consultation notes for the test harness. Depth 4 is the most complex, where there is another dimension of grouping or depth 3 scales are nested into a bigger scale. For example, the APACHE II [8] has depth 4, makes use of Glasgow coma scale [3], and appears within some of the consultation notes for the test harness.

Figure 7 displays the depth of the 104 assessment scales reviewed. 66% of the assessment scales have depth 3, i.e. they have three levels: *assessment scale score*, *assessment scale component*, and *assessment scale item*.
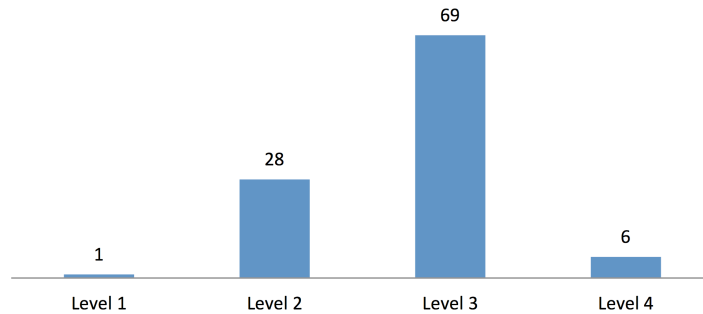

Fig. 7. Depth levels found in the 104 clinical assessment scales reviewed

52 out of 104 (50%) assessment scales can be scored using simple addition (this number includes counting). 94% of the scores (98 out of 104) can be calculated by SPARQL 1.1 exploiting its algebra operators, although some of them require complex arithmetic operations.

37 assessment scales are summarised as a single overall scores only (sometimes call *health index*). Another 34 assessment scales are summarised as a set of scores (sometimes called *profile*), but also do have an overall score. The remaining 32 scales (1 scale was not conclusively classifiable) are generally presented as a set of scores only (no summary overall score exists).

7   Conclusions

Clinical assessment scales, like Glasgow coma scale, present unaddressed challenges related to data management and information retrieval, as their representation and assessment procedure needs to tackle with connections between elements of a specific terminology (e.g. SNOMED CT) and an information model (e.g. HL7 V3). These connections are now widely recognised and known as the *terminology binding process*. This paper proposes: a) exploiting the expressive capabilities of standard languages as the W3C's Web Ontology Language (OWL) to capture key aspects of the *terminology binding process*, i.e. the structure and content of the assessment scale;

and b) using the query language SPARQL to drive the assessment procedure for those health measurements that require the assignment of numerical scores. On the one hand, the systematic review of 104 well-established clinical assessment scales corroborates how profitable it can be to calculate with a single query the total score of more than one assessment scale that may appear in EHRs (e.g. CDA R2 consultation note), as 50% of the assessment scales reviewed can be scored using simple addition. On the other hand, the test harness implemented on top of the query engine ARQ for Jena proves the computational feasibility to execute the more abstract queries presented here, so that all scale assessments requiring simple addition can be performed with a single terse query in less than half a second for a patient's CDA R2 consultation note.

References

1. Weiss, A.: The role of outcomes assessment in clinical quality improvement. In: L. Baer, M.A. Blais (eds.), Handbook of Clinical Rating Scales and Assessment in Psychiatry and Mental Health, Current Clinical Psychiatry, DOI 10.1007/978-1-59745-387-5_15 (2010)
2. Benson, T.: Principles of Health Interoperability HL7 and SNOMED, HI, DOI 10.1007/978-1-84882-803-2_12 (2010)
3. Glasgow coma scale, http://www.patient.co.uk/doctor/glasgow-coma-scale-(gcs).htm
4. Sucurovic, S.: An approach to access control in electronic health record. Journal of Medical Systems, vol. 34(4), pp. 659-666 (2010)
5. Hristidis, V., Farfán, F., Burke, R.P., Rossi, A.F., White, J.A.: Information Discovery on Electronic Medical Records. In: National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation (2007)
6. Apgar, V.: A proposal for a new method of evaluation of the newborn infant. Curr Res Anesth Analg, vol. 32, pp. 260-267 (1953)
7. Barthel index of activities of daily living, http://www.patient.co.uk/doctor/Barthel's-Index-of-Activities-of-Daily-Living-(BAI).htm
8. Knaus, W.A., Draper, E.A., Wagner, D.P., Zimmerman, J.E.: APACHE II: a severity of disease classification system. Crit Care Med, vol. 13, pp. 818–829 (1985)
9. Hamilton, M.: The assessment of anxiety states by rating. Br J Med Psychol, vol. 32, pp. 50-55 (1959)
10. OWL 2 Web Ontology Language, Primer (Second Edition), W3C Recommendation 11 December 2012. Available at http://www.w3.org/TR/owl2-primer/
11. SPARQL 1.1 Query Language, W3C Proposed Recommendation 08 November 2012. Available at http://www.w3.org/TR/sparql11-query/
12. Baader, F., Brandt, D., Lutz, C.: Pushing the EL envelope. In: IJCAI-05 – Proceedings of the 19th International Joint Conference on Artificial Intelligence, pp. 364–369 (2005)
13. IHTSDO, SNOMED CT Editorial Guide (January 2013 International Release) - included in standard release (2013)
14. IHTSDO. Compositional Grammar for SNOMED CT Expressions in HL7 Version 3. External Draft for Trial Use. Version 0.06 (2008)
15. Rector, A., Iannone, L.: Lexically suggest, logically define: Quality assurance of the use of qualifiers and expected results of post-coordination in SNOMED CT. Journal of Biomedical Informatics, vol. 45, pp. 199–209 (2012)
16. Rector, A. Rogers, J.: Ontological and practical issues in using a description logic to represent medical concept systems: experience from GALEN. In: Barahona P, Bry F, Franconi

E, Henze N, Sattler U, editors. Reasoning web. Heidelberg: Springer-Verlag, pp. 197–231 (2006)

17. Nadkarni, P.M.: Metadata-driven Software Systems in Biomedicine. Health Informatics, DOI: 10.1007/978-0-85729-510-1_13 (2011)

18. Hartela, F.W., Coronadoa, S. Dionneb, R. Fragosoa, G., Golbeckc, J.: Modeling a description logic vocabulary for cancer research. Journal of Biomedical Informatics, vol. 38, pp. 114–129 (2005)

19. Coronado, S., Haber, M.W., Sioutos, N., Tuttle, M.S., Wright, L.W.: NCI Thesaurus: using science based terminology to integrate cancer research results. Proceeding of the 11th World Congress on Medical Informatics (Medinfo), pp. 33–37 (2004)

20. Dolin, R.H., Alschuler, L., Boyer, S., Beebe, C., Behlen, F.M., Biron, P.V., Shabo, A.: HL7 Clinical Document Architecture, Release 2. Journal of the American Medical Informatics Association vol. 13, pp. 30-39 (2006)

21. Markwell, D. Sato, L. Cheetham, E.: Representing clinical information using SNOMED Clinical Terms with different structural information models. In: Proceedings of the 3rd international conference on Knowledge Representation in Medicine (KR-MED) (2008)

22. Cheetham, E., Markwell, D., Dolin, R.: Using SNOMED CT in HL7 Version 3; Implementa-tion Guide, Release 1.5 (2009)

23. HL7 RIM, http://www.hl7.org/implement/standards/rim.cfm

24. LOINC, http://www.loinc.org

25. Arguello, M., Des, J., Fernandez-Prieto, M.J., Perez, R. Paniagua, H.: Executing Medical Guidelines on the Web: Towards Next Generation Healthcare. Journal Knowledge-Based Systems, vol. 22, pp. 545–551 (2009)

26. HL7 CDA OWL, http://www.w3.org/wiki/HCLS/ClinicalObservationsInteroperability/HL7CDA2OWL.html

27. Simple SNOMED Module Extraction, http://owl.cs.manchester.ac.uk/snomed/

28. IHTSDO. Compositional Grammar for SNOMED CT Expressions in HL7 Version 3. External Draft for Trial Use. Version 0.06 (2008)

29. ARQ, http://jena.apache.org/documentation/query/

30. Jena, http://jena.apache.org/index.html

31. SPARQL 1.1 Entailment Regimes, W3C Recommendation 21 March 2013. Available at http://www.w3.org/TR/sparql11-entailment/

32. SPARQL Query Results XML Format (Second Edition), W3C Recommendation 21 March 2013. Available at http://www.w3.org/TR/rdf-sparql-XMLres/

33. McDowell, I.: Measuring Health – a guide to rating scales and questionnaires. Oxford University Press (2006)

34. Melzack, R.: The McGill Pain Questionnaire: major properties and scoring methods. Pain. vol. 1(3), pp. 277-299 (1975)

35. Lawton, M.P., Brody, E.M.: Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist, vol. 9, pp. 179–186 (1969)

36. Huskisson, E.C.: Measurement of pain. Lancet, vol. 2, pp. 1127–1131 (1974)