Improving Usability, Safety and Patient Outcomes with Health Information Technology
F. Lau et al. (Eds.)

doi:10.3233/978-1-61499-951-5-53

Linking Health Records with Knowledge Sources Using OWL and RDF

John CHELSOM[a,b,1] and Naveed DOGAR[a,c]
[a] *Seven Informatics, Oxford, UK.*
[b] *University of Victoria, Canada.*
[c] *University of Oxford, UK.*

Abstract. This paper describes a method by which the Web Ontology Language (OWL) can be used to specify a highly structured health record, following internationally recognised standards such as ISO 13606 and HL7 CDA. The structured record is coded using schemes such as SNOMED, ICD or LOINC, with the coding applied statically, on the basis of the predefined structure, or dynamically, on the basis of data values entered in the health record. The highly structured, coded record can then be linked with external knowledge sources which are themselves coded using the Resource Description Framework. These methods have been used to implement dynamic decision support in the open source cityEHR health records system. The effectiveness of the decision support depends on the scope and quality of the clinical coding and the sophistication of the algorithm used to match the structured record with knowledge sources.

Keywords. OWL, RDF, Dublin Core, EHR, ISO 13606, HL7 CDA, Knowledge, Clinical Coding, SNOMED.

1. Introduction

The Web Ontology Language (OWL) [1] is a key component of the Semantic Web activity of the World Wide Web Consortium (W3C) [2]. OWL can be used to specify highly structured models, where the representation of the model as an ontology of distinct, yet connected, statements (axioms) adds the ability to link with other semantic data sets and to use Description Logic reasoners to check consistency and to infer relationships [3].

For the Electronic Health Record (EHR), the most widely used generic models for representation are ISO 13606 [4] and HL7 CDA [5]. We have shown previously how an OWL ontology model, encompassing both ISO 13606 and HL7 CDA, can be used as the foundation for a structured EHR [6].

Another cornerstone of the Semantic Web is the Resource Description Framework (RDF) [7] which provides a mechanism to annotate any addressable resource with meta data, regardless of whether that resource is accessible by the author of the meta data, or not. RDF is often used in conjunction with the Dublin Core standard [8], which defines a set of fifteen commonly used metadata elements, such as title, description, subject and publisher We will show in the next section how RDF and Dublin Core can be used as a standard method for adding clinical coding to any knowledge source.

1 Corresponding Author.

According to Musen at al, “the requirements for excellent decision making fall into three categories: (1) accurate data, (2) pertinent knowledge, and (3) appropriate problem-solving skills” [9]. For a clinical decision support system,as opposed to a decision making system, the problem-solving skills are left to the realm of the clinician; the job of the computer is to represent data accurately and present pertinent knowledge to the decision maker.

This paper describes a relatively simple, yet effective, method for presenting pertinent knowledge, evidence or advice to clinicians at the point where decisions are made. In cases where the EHR is well structured, carries clinical coding and is being used as the primary means of documenting the clinical process, it is possible to extract the clinical context of the EHR session, match that context with relevant, coded, knowledge sources and then retrieve the knowledge for presentation within the EHR session.

The next section outlines the method by which this is achieved using OWL, RDF and Dublin Core; the following section presents the results of applying the method in the open source health records system, cityEHR.

2. Methods

The method for linking health records with knowledge sources, as shown in Figure 1, requires a structured representation of the clinical record, dynamic coding of the record to describe the clinical context, a map of coded knowledge sources,, and use of that map to match the clinical context with the most relevant knowledge sources.

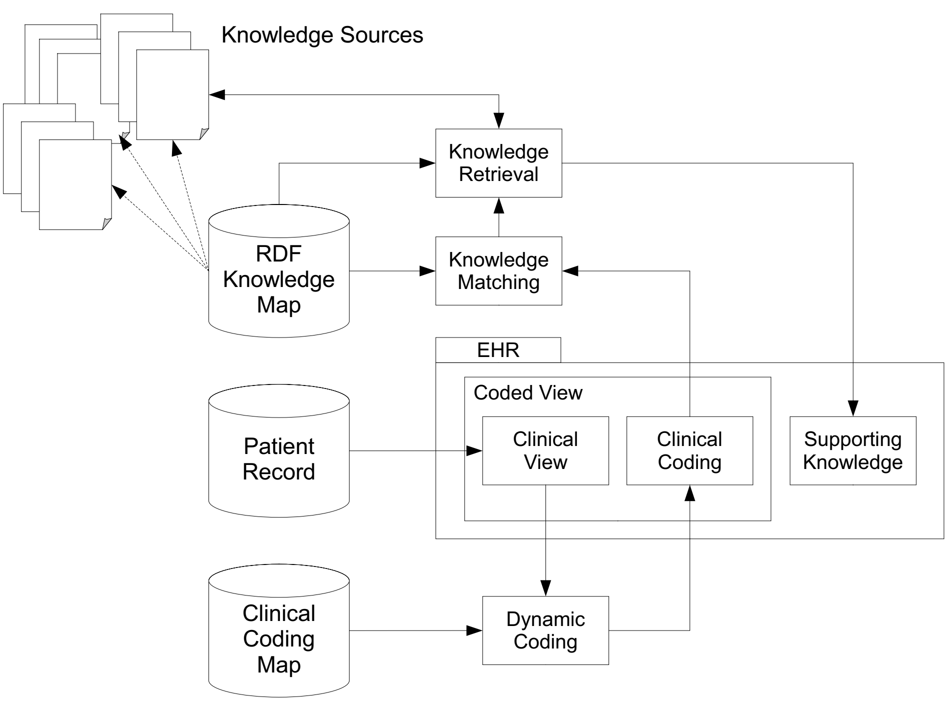

Figure 1. Clinical Decision Support Using Clinical Coding

The clinician runs a session in the EHR and interacts with a Clinical View which is a combination of information fed from the stored Patient Record and new information created within the session. The Dynamic Coding process uses the structure of the Clinical View, the specific information it contains and the Clinical Coding Map to generate a Coded View. The coding is used as input to the Knowledge Matching process which compares it with the RDF Knowledge Map to generate a set of Uniform Resource Locators for the Knowledge Sources, which are then used by the Knowledge Retrieval process to deliver Supporting Knowledge into the EHR session.

2.1. Ontology Representation of the Health Record

The ISO 13606 standard specifies a hierarchical context for any clinical data stored in the health record. The data are recorded in Elements which belong to Entries; an Entry is the lowest level of meaningful clinical context (often referred to as a 'clinical statement'). Further context can be provided by organising Entries within Sections, which are collected in Compositions. Generally, the Composition serves as the unit of storage in the EHR but further clinical context can be applied by organising Compositions within Folders. The Folder level organisation can be applied in a dynamic manner, so that a Composition may appear within multiple Folders, simultaneously. So a Folder could represent, for example, the specialty of the clinical user, an episode of care or a long term condition of the patient.

```
<Declaration>
  <Class IRI="#ISO-13606:Entry"/>
</Declaration>
<Declaration>
  <NamedIndividual IRI="#ISO-13606:Entry:BMDData"/>
</Declaration>
<ClassAssertion>
  <Class IRI="#ISO-13606:Entry"/>
  <NamedIndividual IRI="#ISO-13606:Entry:BMDData"/>
</ClassAssertion>
<DataPropertyAssertion>
  <DataProperty IRI="#hasDisplayName"/>
  <NamedIndividual IRI="#ISO-13606:Entry:BMDData"/>
  <Literal xml:lang="en-gb">DXA Scan Results</Literal>
</DataPropertyAssertion>
<ObjectPropertyAssertion>
  <ObjectProperty IRI="#hasContent"/>
  <NamedIndividual IRI="#ISO-13606:Entry:BMDData"/>
  <NamedIndividual IRI="#ISO-3606:Element:BMDMeasurement"/>
</ObjectPropertyAssertion>
<ObjectPropertyAssertion>
  <ObjectProperty IRI="#hasContent"/>
  <NamedIndividual IRI="#ISO-13606:Entry:BMDData"/>
  <NamedIndividual IRI="#ISO-13606:Element:TScore"/>
</ObjectPropertyAssertion>
<ObjectPropertyAssertion>
  <ObjectProperty IRI="#hasContent"/>
  <NamedIndividual IRI="#ISO-13606:Entry:BMDData"/>
  <NamedIndividual IRI="#ISO-13606:Element:ZScore"/>
</ObjectPropertyAssertion>
```

Figure 2. Structured Health Record Using OWL/XML

Figure 2 shows a small part of the ontology that defines a structured health record. #ISO-13606:Entry is declared to be a class and #ISO-13606:Entry:BMDData to be an individual which belongs to that class and has a display name of “DXA Scan Results” as declared in the DataPropertyAssertion. Three ObjectPropertyAssertions are then used to assert that the #ISO-13606:Entry:BMDData entry contains three elements: #ISO-13606:Element:BMDMeasurement, #ISO-13606:Element:TScore and #ISO-13606:Element:ZScore. The full model of a structured health record contains many thousands of assertions of this type, defining the full set of ISO 13606 Compositions that can be used to record patient information.

2.2. Static and Dynamic Coding of the Structured Health Record

An OWL/XML ontology representation can also be used to model clinical coding and the association of codes with ISO 13606 entries in the record, as shown in Figure 3.

```
<Declaration>
  <Class IRI="#CodingSystem:SNOMED"/>
</Declaration>
<Declaration>
  <NamedIndividual IRI="#Code:SNOMED:64859006"/>
</Declaration>
<ClassAssertion>
  <Class IRI="#CodingSystem:SNOMED"/>
  <NamedIndividual IRI="#Code:SNOMED:64859006"/>
</ClassAssertion>
<DataPropertyAssertion>
  <DataProperty IRI="#hasDisplayName"/>
  <NamedIndividual IRI="#Code:SNOMED:64859006"/>
  <Literal xml:lang="en-gb">Osteoporosis</Literal>
</DataPropertyAssertion>
<DataPropertyAssertion>
  <DataProperty IRI="#hasCode"/>
  <NamedIndividual IRI="#Code:SNOMED:64859006"/>
  <Literal>64859006</Literal>
</DataPropertyAssertion>
```

```
<Declaration>
  <Class IRI="#CodePoint"/>
</Declaration>
<Declaration>
  <NamedIndividual IRI="#CodePoint:BDMData:64859006"/>
</Declaration>
<ClassAssertion>
  <Class IRI="#CodePoint"/>
  <NamedIndividual IRI="#CodePoint:BDMData:64859006"/>
</ClassAssertion>
<DataPropertyAssertion>
  <DataProperty IRI="#hasCondition"/>
  <NamedIndividual IRI="#CodePoint:BDMData:64859006"/>
  <Literal>BMDData/TScore lt -2.5</Literal>
</DataPropertyAssertion>
<ObjectPropertyAssertion>
  <ObjectProperty IRI="#hasCode"/>
  <NamedIndividual IRI="#CodePoint:BDMData:64859006"/>
  <NamedIndividual IRI="#Code:SNOMED:64859006"/>
</ObjectPropertyAssertion>

<ObjectPropertyAssertion>
  <ObjectProperty IRI="#hasCodePoint"/>
  <NamedIndividual IRI="#ISO-13606:Entry:BMDData""/>
  <NamedIndividual IRI="#CodePoint:BDMData:64859006"/>
</ObjectPropertyAssertion>
```

Figure 3. Clinical Coding Using OWL/XML

The first set of assertions defines the clinical code for Osteoporosis from the SNOMED coding scheme (www.snowmed.org). The second set of assertions defines a CodePoint, which assigns the SNOMED code 64859006 (Osteopososis) to the ISO 13606 Entry for BMDData, with the condition that TScore is less than -2.5. This set of assertions is sufficient to implement dynamic coding of the BMDData Entry; a similar set without the hasCondition data property assertion would be used to implement static coding (i.e. the code would always be assigned to the Entry). The expression in the Literal string for hasCondition uses the W3C standard XPath language, with BMDData/TScore representing the TScore Element of the BMDData Entry.

2.3. Coding Knowledge Sources Using RDF and Dublin Core

The RDF Knowledge Map contains a set of metadata assertions as triples of subject-predicate-object. The RDF XML coding shown in Figure 4, uses the rdf:about attribute to represent the subject, which is any addressable resource (in this case a clinical guideline for Osteoporosis from the National Institute of Health and Care Excellence).

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:dc="http://purl.org/dc/elements/1.1/" dc:title="Knowledge Map">

    <rdf:Description rdf:about="http://pathways.nice.org.uk/pathways/osteoporosis">
        <dc:subject rdf:datatype="http://www.snomed.org">391068001</dc:subject>
        <dc:description rdf:datatype="http://www.snomed.org">
           Hip dual energy X-ray photon absorptiometry scan T score
        </dc:description>
    </rdf:Description>

    <rdf:Description rdf:about="http://pathways.nice.org.uk/pathways/osteoporosis">
        <dc:subject rdf:datatype="http://www.snomed.org">64859006</dc:subject>
        <dc:description rdf:datatype="http://www.snomed.org">Osteoporosis</dc:description>
    </rdf:Description>

    <rdf:Description rdf:about="http://pathways.nice.org.uk/pathways/osteoporosis">
        <dc:subject rdf:datatype="http://www.snomed.org">312894000</dc:subject>
        <dc:description rdf:datatype="http://www.snomed.org">Osteopenia</dc:description>
    </rdf:Description>

</rdf:RDF>
```

Figure 4. RDF Knowledge Map Using RDF and Dublin Core

The Dublin Core dc:subject element represents the predicate for subject metadata from the SNOMED vocabulary (defined by the rdf:datatype attribute) and the content of that element represents the object, which in this example is the SNOMED code assigned to the Osteoporosis guideline. The example in Figure 4 shows the assignment of three SNOMED codes to the Osteoporosis guideline; the full Knowledge Map for a clinical specialty would contain thousands of codes.

3. Results

The methods described above were used to implement new features in the open source cityEHR health records system, for creating and managing RDF knowledge maps, dynamic coding of the health record and presentation of supporting knowledge.

3.1. Creating and Managing RDF Knowledge Maps

In cityEHR, RDF knowledge maps can be created and edited manually or can be imported from external maps (perhaps generated as the result of Natural Language Processing on targeted web resources). The maps are stored as indexed XML documents that can be searched efficiently using the W3C standard XQuery language.

3.2. Coding of the Health Record

For the cityEHR runtime system, the ontology-based definition of the structured record is transformed from OWL/XML into a set of HL7 CDA XML documents which are linked to a corresponding set of automatically generated forms using the W3C XForms standard; these forms are then used for interaction with the clinical user.

Similarly, the ontology-based Clinical Coding Map is transformed into a set of HL7 CDA code elements, that are inserted into the XForm that represents the current clinical view. Static coding (without conditions) is inserted directly; dynamic coding is made by transforming the conditions expressed in the ontology definition into full XPath expressions that are linked to the structure of the HL7 CDA document in the XForm and are then evaluated as clinical information changes, as part of the standard XForms implementation.

3.3. Retrieval and Presentation of Knowledge

The first step in retrieving knowledge, is to extract the clinical coding associated with the clinical statements at the Entry level, then less specific context at the Section, Composition and Folder levels of the ISO 13606 structure. The extracted codes are then used to match with codes in the RDF Knowledge Map, to find relevant knowledge sources.

The algorithm used to make this match can be more or less sophisticated. The simplest algorithm merely retrieves all knowledge sources that are coded with any of the same codes extracted from the clinical context. More sophisticated algorithms can rank the knowledge sources based on the number of matching codes, or use an understanding of the relationships between clinical codes (for example the codes for Male and Female are disjoint).

The matches on knowledge sources are returned as URLs (the URLs specified by the rdf:about attributes in the RDF Knowledge Map). To present the knowledge in the cityEHR session, the URLs are passed to a web service which accesses each URL and returns the content found there. If the content is in PDF format it is returned directly; if it is in HTML format then the content is returned using the Beautiful Soup HTML scraping library [11].

4. Conclusions

Structured health records, associated clinical coding and maps of coded knowledge resources can be represented using the Semantic Web standards OWL and RDF. These representations can be transformed to drive an EHR system, where clinical coding is applied statically or dynamically to the structured health record, then extracted as a representation of the clinical context and used to retrieve relevant knowledge to support clinical decision makers. The effectiveness of this method depends on the quality and extent of the clinical coding maps and the sophistication of the algorithms used to match clinical context in the EHR session with the coded knowledge sources.

References

[1] W3C Owl Working Group. OWL 2 Web Ontology Language. Document Overview (Second Edition) *W3C Recommendation*, 11 December 2012.

[2] T. Berners-Lee, J. Hendler, and O. Lassila, 2001. The semantic web. *Scientific American* **284** (2001), 34-43.

[3] C. Tsinaraki and S. Christodoulakis, Interoperability of XML schema applications with OWL domain knowledge and semantic web tools. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems."* November 2007, 850-869, Springer, Berlin, Heidelberg.

[4] ISO 13606-1:2008 *Health informatics - Electronic health record communication - Part 1: Reference model*. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=40784

[5] R.H. Dolin, L. Alschuler and S. Boyer, et al. HL7 Clinical Document Architecture, Release 2. J Am Med Inform Assoc. **13** (2006), 30-9.

[6] J.J. Chelsom, I. Pande, R. Summers and I. Gaywood. Ontology-driven development of a clinical research information system. *24th International Symposium on Computer-Based Medical Systems*, Bristol. June 27-June 30, 2011. ISBN: 978-1-4577-1189-3.

[7] O. Lassila and R.R. Swick. Resource description framework (RDF) model and syntax specification. *W3C Recommendation*, 1999.

[8] S. Weibel, J. Kunze, C. Lagoze and M. Wolf. *Dublin core metadata for resource discovery (No. RFC 2413),* 1998.

[9] M.A. Musen, Y. Shahar and E.H. Shortliffe. Clinical Decision-Support Systems. In *Biomedical Informatics* (pp. 698-736). Springer, New York, NY. 2006.

[10] S. Gauch, J. Chaffee and A. Pretschner, 2003. Ontology-based personalized search and browsing. *Web Intelligence and Agent Systems: An International Journal*, **1** (2003), 219-234.

[11] L. Richardson, *Beautiful Soup Documentation*. 2015.
https://www.crummy.com/software/BeautifulSoup/