

FROM STORED CLINICAL DATA TO INTEROPERABLE CLINICAL KNOWLEDGE

Idoia Berges, Jesús Bermúdez, Alfredo Goñi and Arantza Illarramendi
University of the Basque Country, Paseo Manuel de Lardizabal 1, Donostia-San Sebastián, Spain

Keywords: Semantic Interoperability, Electronic Health Records.

Abstract: Health Information Systems deal with a great volume of digital clinical data. Although this management has brought several advantages to the healthcare domain, a more intelligent management of those stored data can provide further benefits. In this paper we present a proposal which introduces two main new advantages to those Health Information Systems: First, the possibility of a semantic interoperability among them and second, the ability to share the medical knowledge generated by each system—or by specialized organizations. Our proposal makes use of an ontology to describe the terms used by the different Electronic Health Record standards and the terms used by specific Health Information Systems (not forcing them to use a fixed standard), and a reasoner that allows to interpret on the fly the data of a particular Health Information System by another one, even when they use different data representations. Moreover, our solution provides a rule-based formalism for representing medical knowledge of each system. Thanks to this representation, one system can benefit from the knowledge which is stored on other systems and that was not initially aware of.

1 INTRODUCTION

The use of Electronic Health Records (EHRs) has brought multiple advantages to the healthcare domain. The problem of poor legibility that might occur when using handwritten paper records is avoided. Moreover, decision support can be added to the medical systems to help the professionals in the decision process as well as to detect possible inconsistencies. These advantages lead to less error-prone systems which undoubtedly increase the quality of healthcare. Another benefit of using information technologies in this area is the possibility of exchanging EHRs between different organizations. Over his lifetime, a patient is likely to receive medical attention from several institutions, and it seems reasonable for each institution to have access to the previously recorded data of the patient. Nowadays, however, this interoperability is still hard to achieve. Health Information Systems used within the organizations have been independently developed, which results in a high number of heterogeneous models for representing and recording EHRs.

In order to try to solve this situation, several standards have arisen to represent EHRs. Most of them follow a dual model approach, which separates

the information level from the knowledge level. The information level is represented by the Reference Information Model (RIM), which contains the generic classes and properties that allow the writing of any clinical annotations. For example, the classes representing Tables, Lists and Entries will be found at this level. The RIM is considered a stable model that is not expected to change. Since the RIM is composed of a small number of classes and they are too general to describe the semantics of the clinical terms, archetypes are used in the knowledge level to create those descriptions. Archetypes impose restrictions over the classes of the RIM to create new clinical terms. These terms can be linked to clinical terminologies, such as SNOMED (SNOMED, 2009) or LOINC (LOINC, 2009) in order to represent their meaning. For example, the term “Blood Pressure Reading” could be described by an archetype.

The best known EHR standards are openEHR(openEHR, 2009), CEN13606(CEN-13606, 2007) and HL7 CDA(HL7-CDA, 2009). The openEHR standard has been developed by the openEHR Foundation, an international not for profit foundation that aims at improving healthcare by developing specifications, software and knowledge resources with the purpose of achieving EHR in-

teroperability. It follows the aforementioned dual model approach and has already done a major work in designing archetypes and templates (compositions of archetypes). CEN 13606 is the proposal of the European Committee for Standardization to represent EHRs. It follows also the dual model approach and provides by now a quite simple RIM and few archetypes based on those of openEHR. Finally, HL7 CDA has been developed by HL7 and provides a RIM and a draft template specification, which in this standard represents the same idea of the openEHR archetypes. For example, let us suppose that the archetypes for representing the concept of “Risk Score”(RS) are described as in Fig. 1 for openEHR and CEN13606, respectively. Notice that some of the classes to build the archetypes are different in both standards (e.g. Observation vs. Entry in Fig. 1).

However, none of these standards has been universally adopted, so the interoperability problem remains unsolved. Moreover, as most medical organizations in the world have their own developed EHR models, it would be too expensive and time consuming to change all their Health Information Systems, migrating the stored data instances to the new specifications, and training the medical staff to use the new system. We advocate for another solution which is transparent to the medical staff and where only the essential information is transformed into another representation. This solution goes far beyond the use of XML for the interchange of data—because even if it has been proved relevant for this issue, it does not deal with the semantics of the exchanged data (Hefflin and Hendler, 2000). On the contrary, our proposal benefits from emerging semantic web technologies, and more specifically ontologies, which can play a relevant role in the development of frameworks to facilitate semantic interoperation between heterogeneous Information Systems (Obrst, 2003).

In this paper we present a framework that favours the semantic interoperability between heterogeneous Health Information Systems. Two main research issues are tackled by the proposal: the query interpretation problem (Sections 2 and 3) and the clinical knowledge sharing (Section 4). Considering the first one, we have built an ontology, named EHRONT, in which the RIMs and archetypes from the EHR standards and specific EHRs are described. This ontology has been built following well-known methodologies for building ontologies (Corcho et al., 2003). Thanks to the ontological axioms defined between the terms of different standards, and those between terms of standards and terms of specific EHRs, a record of a specific Health Information System *A* (corresponding

to a particular patient) will be interpreted on the fly by another specific Health Information System *B*. With regard to the second issue, our approach allows the sharing of medical knowledge between systems by using Semantic Web rules.

Achieving real interoperability among EHRs is a research issue into which great effort is being put ((European Community, 2009), (Hoffman, 2009)). Among the related works closer to our proposal, the following ones can be mentioned: (Kilic and Dogac, 2009) provides a solution that uses ontological reasoning. In this case, communication between systems that follow different standards under the same RIM (HL7-RIM) is supported, but communication between systems that use their proprietary EHR specifications is not considered. Moreover, in the transformation process of an instance of a source archetype $Arch_A$, the target archetype must be explicitly declared, which requires the sender to know specific knowledge about the receiver. In (Martínez-Costa et al., 2009) a software architecture is presented for transforming a source ADL archetype description that follows openEHR into a target ADL description that follows UNE-EN 13606. Ontologies describing archetype models of both standards, in addition to an integrated ontology, are used in the process. Notice that neither of both works considers the feature of knowledge sharing.

2 THE EHRONT ONTOLOGY

The EHRONT ontology is the central element of the framework and it is essential to achieve semantic interoperability between heterogeneous EHR systems. The terms of the EHRONT ontology are described as classes and properties using the Web Ontology Language, OWL (OWL, 2009), and more precisely, OWL-DL. EHRONT is composed of two interrelated layers (standards layer and application layer) that classify the EHR contents regarding different levels of abstraction, being the standards layer the most general and the applications layer the most concrete. Each Health Information System will have its own version of the EHRONT ontology. The standards layer will be the same for all versions, while the applications layer will be proper to each system. In the standards layer, the classes and properties (RIM) that are specific to each EHR standard are specified, as well as their archetypes. Up to now, the terms of openEHR RIM, CEN13606 RIM and HL7-CDA and their archetypes belong to this level.

Following, a fragment of the logical representa-

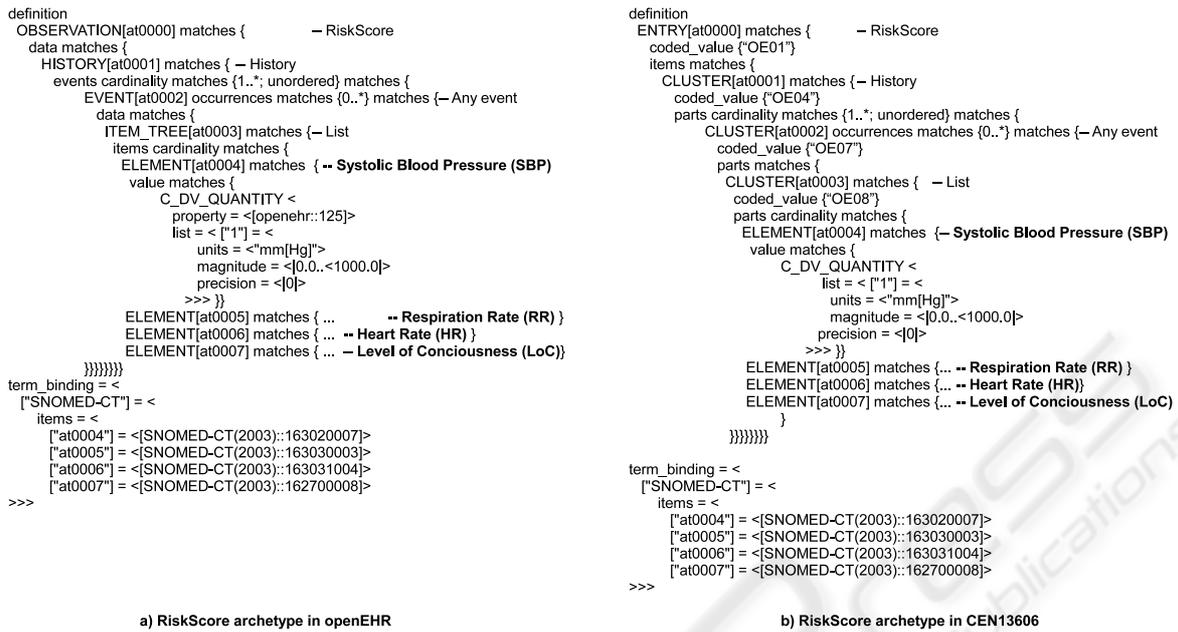


Figure 1: Archetypes for representing the Risk Score.

tion of archetypes in Fig. 1 can be found¹.

oe:RS ≡ oe:Observation ⊃ oe:hasObsData.oe:RSHistory
 oe:RSHistory ≡ oe:History ⊃ oe:hasHistoryEvent.oe:RSEvent
 oe:RSEvent ≡ oe:Event ⊃ oe:hasEvent.oe:RSData
 oe:RSData ≡ oe:ItemTree ⊃ oe:hasItem.oe:SBP
 ⊃ oe:hasItem.oe:RR ⊃ oe:hasItem.oe:HR
 ⊃ oe:hasItem.oe:LoC
 oe:SBP ≡ ∃oe:hasSNOMEDCode. { '163030007' }
 ⊃ oe:units. { 'mmHg' } ⊃ oe:magnitude
 cen:RS ≡ cen:Entry ⊃ cen:codedV. { 'OE01' }
 ⊃ cen:items.cen:RSHistory
 cen:RSHistory ≡ cen:Cluster ⊃ cen:codedV. { 'OE04' }
 ⊃ cen:parts.cen:RSEvent
 cen:RSEvent ≡ cen:Cluster ⊃ cen:codedV. { 'OE07' }
 ⊃ cen:parts.cen:RSData
 cen:RSData ≡ cen:Cluster ⊃ cen:codedV. { 'OE08' }
 ⊃ cen:parts.cen:SBP ⊃ cen:parts.cen:RR
 ⊃ cen:parts.cen:HR ⊃ cen:parts.cen:LoC
 cen:SBP ≡ ∃cen:SNOMEDCode. { '163030007' }
 ⊃ cen:units. { 'mmHg' } ⊃ cen:magnitude

Moreover, relationships between terms of the different standards are defined in this layer. These rela-

¹We prefer this logical notation instead of the more verbose RDF/XML syntax. Terms belonging to the openEHR standard are prefixed by **oe:**, while terms belonging to the CEN13606 standard are prefixed by **cen:**

tionships are essential to achieve interoperability between medical systems based on different EHR standards.

oe:Observation ≡ cen:Entry ⊃ cen:codedV. { 'OE01' }
 oe:History ≡ cen:Cluster ⊃ cen:codedV. { 'OE04' }
 oe:Event ≡ cen:Cluster ⊃ cen:codedV. { 'OE07' }
 oe:ItemTree ≡ cen:Cluster ⊃ cen:codedV. { 'OE08' }
 ∃oe:hasObsData ≡ ∃cen:items
 ∃oe:hasHistoryEvent ⊃ ∃cen:parts
 ∃oe:hasEvent ⊃ ∃cen:parts
 ∃oe:hasItem ⊃ ∃cen:parts
 ∃oe:hasSNOMEDCode ≡ ∃cen:SNOMEDCode
 ∃oe:units ≡ ∃cen:units
 ∃oe:magnitude ≡ ∃cen:magnitude

The application layer extends the level above with classes and properties that are specific to each Health Information System. All these classes can be defined as subclasses of the terms in the standards layer of the EHRONT ontology. Semantic Web Rules can be added to enrich instance relationships. Obviously, the more relationships and rules are defined with the terms of the standards layer, the easier it will be to achieve interoperability among systems. Moreover, we propose a module that explores a schema of a particular system and extracts all its information (fields, datatypes, restrictions, etc.) to create a new class in the application layer of the EHRONT used by that

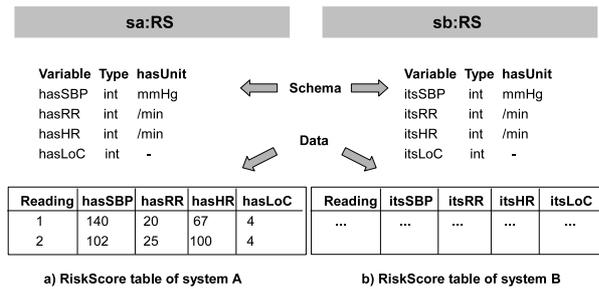


Figure 2: Risk Score tables of particular systems.

system. This new class is indeed the definition of an archetype. Let us imagine a Health Information System *A* whose table for storing Risk Score readings is the one in Fig. 2a. The representation of this archetype in the application layer of the EHRONT ontology, using OWL, could be the following:

```

sa:RS ≡ oe:Observation ⊓ ∃sa:hasSBP.sa:SBP
      ⊓ ∃sa:hasRR.sa:RR ⊓ ∃sa:hasHR.sa:HR
      ⊓ ∃sa:hasLoC.sa:LoC

sa:hasUnit ⊆ oe:units
sa:hasMagnitude ⊆ oe:magnitude
sa:hasSnomedCode ⊆ oe:hasSNOMEDCode

sa:SBP ≡ ∃sa:hasUnit. { 'mmHg' }
      ⊓ ∃sa:hasSnomedCode. { '163020007' }

```

In order to map instances of this archetype to, for example, the Risk Score archetype of the openEHR standard, the following SWRL(SWRL, 2009) rule must be used, because the openEHR Risk Score archetype is much more complex than this one.

```

sa:RS(?x) ∧ sa:hasSBP(?x,?y) ∧ sa:hasRR(?x,?z)
∧ sa:hasHR(?x,?a) ∧ sa:hasLoC(?x,?b)
∧ swrlx:createOWLThing(?x,?c) ∧ swrlx:createOWLThing(?x,?d)
∧ swrlx:createOWLThing(?x,?e) →
oe:RS(?x) ∧ oe:RSHistory(?c) ∧ oe:RSEvent(?d) ∧ oe:RSData(?e)
∧ oe:hasObsData(?x,?c) ∧ oe:hasHistoryEvent(?c,?d)
∧ oe:hasEvent(?d,?e) ∧ oe:hasItem(?e,?y) ∧ oe:hasItem(?e,?z)
∧ oe:hasItem(?e,?a) ∧ oe:hasItem(?e,?b)

```

3 INSTANCE INTEROPERABILITY

The first contribution provided by our proposal is the capability of a particular system *B* to interpret information sent by another system *A* on the fly.

Let $T_1(i)$ be a tuple stored on the T_1 table of a database of system *A*. Let T_2 be a table of a database of system *B*. Let \mathcal{A} and \mathcal{B} the versions of EHRONT ontologies used by systems *A* and *B* respectively. In

addition, let us assume that the module mentioned in the previous section has converted the table schema from T_1 to an OWL archetype class in the application level of \mathcal{A} , creating the class $\mathcal{A}:appT_1$. In the same way, the table schema from T_2 has been converted to an OWL archetype class in the application level of \mathcal{B} , creating the class $\mathcal{B}:appT_2$. Furthermore, specialization relationships have been created between $\mathcal{A}:appT_1$ and the standard level class $\mathcal{A}:openehrT_1$ and between $\mathcal{B}:appT_2$ and the standard level class $\mathcal{B}:centT_2$. Finally, in the standards level relationships have been defined between $\mathcal{A}:openehrT_1$ and $\mathcal{B}:centT_2$.

The process that needs to be carried is composed of several steps:

1. One specific module is used in system *A* to convert $T_1(i)$ into an individual of the class $\mathcal{A}:appT_1$, namely $\mathcal{A}:appT_1(i)$. In our example, if the second tuple of the table in Fig. 2a is to be migrated, individual *sa:rs-2* is created. Moreover, the following OWL assertions will be created, among others:

```

(sa:rs-2 is-a sa:RS)
(sa:rs-2 is-a oe:Observation)
(sa:rs-2 sa:hasSBP sa:sbp-1)
(sa:sbp-1 sa:hasUnit "mmHg")
(sa:sbp-1 sa:hasMagnitude 102)
(sa:sbp-1 sa:hasSnomedCode "163020007")

```

2. At this point, with the help of an OWL reasoner, all the information related to $\mathcal{A}:appT_1(i)$ is extracted and converted into a set of OWL assertions. Thanks to the specialization relationships and rules that have been defined between the application layer class $\mathcal{A}:appT_1$ and the standards layer class $\mathcal{A}:openehrT_1(i)$ of the \mathcal{A} ontology, $\mathcal{A}:appT_1(i)$ will be also related to the terms of the standards layer. More precisely, it will also be an individual of the $\mathcal{A}:openehrT_1$ class. In our example, thanks to the rule described in the previous section, the following assertions will be created:

```

(sa:rs-2 is-a oe:RS)
(sa:rs-2 oe:hasObsData oe:rs-hist-1)
(oe:rs-hist-1 is-a oe:RSHistory)
(oe:hist-1 oe:hasHistoryEvent oe:rs-event-1)
(oe:rs-event-1 is-a oe:RSEvent)
(oe:rs-event-1 oe:hasEvent oe:rs-data-1)
(oe:rs-data-1 is-a oe:RSData)
(oe:rs-data-1 oe:hasItem sa:sbp-1)
(sa:sbp-1 is-a oe:SBP)
(sa:sbp-1 oe:units "mmHg")
(sa:sbp-1 oe:magnitude 102)
(sa:sbp-1 oe:hasSNOMEDCode "163020007")

```

3. The OWL assertions are sent to system *B* and asserted into the \mathcal{B} ontology. Again, with the help of a reasoner, new information is obtained. Thanks to the horizontal relationships defined in the standards

layer of the EHRONT ontology, the $\mathcal{A}:appT_1(i)$ individual is transformed into an individual of the standards layer class $\mathcal{B}:cenT_2$. Moreover, thanks to the specialization relationships defined between the standards and application layers of the EHRONT ontology, $\mathcal{A}:appT_1(i)$ will become also an individual of the $\mathcal{B}:appT_2$ class.

In our example, let us imagine that in system B , the Risk Score table is the one in Fig. 2b and that there is a description of its archetype in the application layer in the same way that has been done in section 2 for the table of system A . Among others, the following assertions will be inferred:

```
(sa:rs-2 is-a cen:RS)
(sa:rs-2 cen:items oe:rs-hist-1)
(oe:rs-hist-1 is-a cen:RSHistory)
(oe:hist-1 cen:parts oe:rs-event-1)
(oe:rs-event-1 is-a cen:RSEvent)
(oe:rs-event-1 cen:parts oe:rs-data-1)
(oe:rs-data-1 is-a cen:RSData)
(oe:rs-data-1 cen:parts sa:sbp-1)
(sa:sbp-1 is-a cen:SBP)
(sa:sbp-1 cen:units "mmHg")
(sa:sbp-1 cen:magnitude 102)
(sa:sbp-1 cen:SNOMEDCode "163020007")
```

Furthermore, thanks to the rules that may be defined to relate the standards layer classes with the application layer, the following assertions are created:

```
(sa:rs-2 is-a sb:RS)
(sa:rs-2 sb:hasSBPReading sa:sbp-1)
(sa:sbp-1 is-a sb:SBPReading)
(sa:rs-2 sb:itsUnit "mmHg")
(sa:rs-2 sb:itsMagnitude 102)
(sa:rs-2 sb:itsSNOMEDCode "163020007")
```

4. As $\mathcal{A}:appT_1(i)$ is now an individual of the class $\mathcal{B}:appT_2$, one specific module can be used to convert the individual to an instance of the T_2 table.

Notice that thanks to the relationships that have been defined in the EHRONT ontology, a tuple of a table of a particular system has been interpreted by another system.

4 KNOWLEDGE SHARING

EHRs hold great potential for clinical decision support, for example by translating practice guidelines into automated reminders and actionable recommendation. Usually, medical experts are in charge of those translation tasks. An additional advantage provided by our proposal is the possibility of defining and sharing obtained medical knowledge, expressed as rules, between Health Information Systems. In order to define the knowledge rules we have chosen again SWRL, because it uses the declarative form to

Table 1: Expected values for the variables of the Risk Score.

Variable	Min. value	Max. value
Systolic Blood Pressure (SBP)	91	199
Respiration Rate (RR)	8	30
Heart Rate (HR)	50	119
Level of Consciousness (LoC)	4	4

express rules (which is suitable for obtaining conclusions from a set of data) and moreover it is thought to be used along with the OWL language.

For example, SWRL rules can be used to calculate the Risk Score (RS) value of a given patient once the Risk Score variables are registered. Table 1 shows the values that are considered as normal for each variable.

If any of the variables is outside its limits, there is an anomaly with that patient. For each anomaly that is detected, the Risk Score value increases by 1. The only good Risk Score for a patient is to have value 0. Otherwise, there is some risk and medical staff must be warned.

The knowledge could be expressed at the standard level using SWRL rules similar to the following ones. From this moment on, those rules can be shared with other systems.

- $$\begin{aligned} &oe:RSData(?x) \wedge oe:hasItem(?x, ?y) \wedge oe:SBP(?y) \wedge \\ &oe:magnitude(?y, ?a) \wedge swrlb:lessThan(?a, 91) \rightarrow \\ &oe:hasRS-SBP(?x, 1) \end{aligned}$$
- $$\begin{aligned} &oe:RSData(?x) \wedge oe:hasRS-SBP(?x, ?y) \wedge oe:hasRS-RR(?x, ?z) \\ &\wedge oe:hasRS-HR(?x, ?a) \wedge oe:hasRS-LoC(?x, ?b) \\ &\wedge swrlb:add(?c, ?y, ?z) \wedge swrlb:add(?d, ?c, ?a) \\ &\wedge swrlb:add(?e, ?d, ?b) \rightarrow oe:hasRS(?x, ?e) \end{aligned}$$

5 CONCLUSIONS

Although great efforts are being made in order to achieve interoperability between Health Information Systems, in our humble opinion there is still much work to be done. Our approach takes into account the real problem of systems that were not developed under EHR standards and integrates them into a semantic-based framework. This framework allows interoperability of EHR records (supported by a reasoning process that decreases human intervention) and, additionally, the shareability of encoded knowledge about clinical processes. As future work we plan to precisely identify the types of queries that can be totally or partially interpreted on the fly.

REFERENCES

- CEN-13606 (2007). EN 13606-1: Electronic Health Record Communication.
- Corcho, Ó., Fernández-López, M., and Gómez-Pérez, A. (2003). Methodologies, tools and languages for building ontologies: Where is their meeting point? *Data and Knowledge Engineering*, 46(1):41–64.
- European Community (2009). Semantic Interoperability for Better Health and Safer Healthcare. Research and Deployment Roadmap for Europe. ISBN-13: 978-92-79-11139-6.
- Hefflin, J. and Hendler, J. (2000). Semantic Interoperability on the Web. In *Proceedings of Extreme Markup Languages 2000*, pages 111–120. Graphic Communications Association.
- HL7-CDA (2009). Available at <http://www.hl7.org>.
- Hoffman, L. (2009). Implementing Electronic Medical Records. *Communications of the ACM*, 52(11):18–20.
- Kilic, O. and Dogac, A. (2009). Achieving Clinical Statement Interoperability using R-MIM and Archetype-based Semantic Transformations. *IEEE Transactions on Information Technology in Biomedicine*, to appear.
- LOINC (2009). Available at <http://loinc.org>.
- Martínez-Costa, C., Menárguez-Tortosa, M., Valencia-García, R., Maldonado, J., and Fernández-Breis, J. T. (2009). Transformación Automática de Arquetipos UNE-EN 13606 y openEHR para Facilitar la Interoperabilidad Semántica. In *Inforsalud 2009*, Madrid, Spain.
- Obrst, L. (2003). Ontologies for Semantically Interoperable Systems. In *Proceedings of the 2003 ACM CIKM International Conference on Information and Knowledge Management*, pages 366–369, New Orleans, Louisiana, USA. ACM.
- openEHR (2009). Available at <http://www.openehr.org>.
- OWL (2009). Available at <http://www.w3.org/2001/sw/WebOnt/guide-src/Guide.html>.
- SNOMED (2009). Available at <http://www.ihtsdo.org/snomed-ct/>.
- SWRL (2009). Available at <http://www.w3.org/Submission/SWRL/>.