SEMANTIC SERVICE-ORIENTED DESIGN AND DEVELOPMENT METHODOLOGY FOR ENTERPRISE HEALTHCARE INTEGRATION

Ratnesh Sahay, Ronan Fox and Manfred Hauswirth

Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland

Keywords: Health level seven (HL7), SOA, Web service, Electronic patient record (EPR), Semantics.

Abstract: The application of ontologies (semantic) to enhance any existing or proposed Service-Oriented Architecture (SOA) based software architecture has various levels of use in terms of intra and inter enterprise integration. The use of ontologies in an architectural design and development methodology of any service-oriented enterprise software holds the key to offer a dynamic, flexible and scalable solution. Current efforts in semantic Service-Oriented Architecture (sSOA) involve primarily the top-down modeling of services and data. A roadmap that meets industrial demand of existing (bottom-up) services and data is missing. This paper analyses a healthcare standard (HL7) as an integration mechanism to connect healthcare enterprises. We have applied semantics on top of HL7 profiles to fill the gap between HL7 and SOA artifacts. The results have shown that semantics can ease the integration steps and burden to connect healthcare enterprises. We have proposed an integration platform which is based on a semantic Service-oriented Architecture (sSOA). Our goal is to apply lightweight semantics that incorporate and benefit from both development methodologies (top-down and bottom-up), to create a converged approach, for enterprise healthcare integration.

1 INTRODUCTION

In recent years various research and industrial efforts have been focussed on Service-Oriented Architectures (SOAs) where Web services provide the technological foundation for implementing and delivering SOA platforms. The integration and/or interoperability requirements of enterprise information systems have resulted in the development of new types of SOAs, called semantic Service-oriented Architecture (sSOA). However, a clear design and development methodology is missing and "gaps" between domain, SOA (Web service), and semantic (ontology) artifacts exist.

A generalised integration approach for all types of enterprises is not useful as each domain has its own complexities and interoperability requirements. A domain-based balanced integration approach which includes domain knowledge (e.g. simple taxonomies, ontologies), technology (e.g. Web services, semantics tools), and domain specific development methodology (e.g. top-down/bottom-up) is required to achieve the full potential of a semantic Service-oriented architecture and to deliver a meaningful enterprise integration solution.

In a top-down ontology development methodology, ontologies are created first whereas in bottom-up methodology ontologies are created from existing syntactic format, e.g., XML, EDI, WSDL. We have introduced a Electronic Patient Record (EPR) integration platform, called PPEPR: Plug and Play Electronic Patient Records1, to connect healthcare enterprises. Our focus is on Health Level Seven (HL7) standard, which is due to the fact that it is the most widely used message based healthcare communication standard and HL7’s user base has been growing since the early 2000s.

In this paper, first we analyse the HL7 standard and its Web Service and SOA profiles from Electronic Patient Record (EPR) integration perspective. Second, we describe how ontologies are applied on top of HL7 profiles to resolve the ambiguity and heterogeneity between messages, service, and process definitions. Next, we explain our ontology development methodology. Finally, we explain PPEPR’s assessment that shows its effectiveness which is evaluated on various integration parameters.

1http://www.ppepr.com/
The main contribution of this paper is to present a design and development methodology for associating semantics with HL7’s message, service, and process definitions in a service-oriented environment.

2 HL7 AND SOA

Health Level Seven (HL7) is a standardization body, which develops data standards for storing and exchanging information across the healthcare industry. The HL7 standards cover both clinical and administrative aspects of the healthcare industry, including laboratory, clinical genomic, medical records, patient care, pharmacy, public health reporting, regulated studies, accounts and billing, claims and reimbursement, patient administration and personnel management scheduling. There are two major HL7’s versions, HL7 v2 & v3. The HL7 v2 standard was created mostly by clinical interface specialists, and the HL7 v3 standard has been influenced by medical informaticians.

HL7 v2 message is (EDI) Electronic Data Interchange based and consists of segments which are represented by rows. Segments are divided into data fields, separated by vertical bar symbols, each field containing data elements. Data elements conform to any of several HL7 v2 data types.

The HL7 v3 specifications are centered around a single, unified Reference Information Model (RIM) covering all domains of the healthcare industry. The RIM defines all data structures, data types and vocabularies, as well as the relationships between them. RIM structural arrangement is based on object-oriented paradigm and it includes several classes, representing domain models from which all healthcare-related messages and documents are built. HL7 v2 messages are unstructured and flexible involving optional fields and segments whereas HL7 v3 is structured and provides greater consistency across the entire standard.

In 2003 HL7 has published the HL7 v3 Web service profile(Ruggeri et al., ) that provides the useful capability to transport existing HL7 v3 messages using Web service protocols. The intention of this WS profile is to achieve “plug-n-play” interoperability via Web services in a healthcare environment. In this environment Independent Software Vendors (ISV) and corporate developers implementing HL7 interfaces can write generic and reusable classes, subroutines, and modules consistent with the guidelines set forth by the HL7 for Web services standards in order to handle HL7 message traffic from a potentially unlimited number of connecting applications and services.

If applications that “expose” HL7 messages follow the HL7 Web services profile (WSP) guidelines, then “consumers” of HL7 messages can be written without prior knowledge of interacting applications. A detailed description of HL7 specifications is outside the scope of this paper and we briefly explain HL7 from an integration perspective. Three major issues from an integration perspective are:

1. The service definition becomes superfluous, this leads to message definition based bottom up approach where service clients are automatically able to interoperate based on the messaging definition.
2. The WS-profile assumes that all different healthcare entities will follow the particular standard.
3. Message, service, and process definitions are tied together. Thus, there is an absence of “separation of concern”.

One major benefit of this approach is that “prior knowledge” or a single “agreed” model is not required at the communication level but still assumes a single “agreed” model at specification level where all healthcare entities should follow the Web service profile. There is a common industrial practice that people who manage information, most often have different ways of interpreting this information(Iakovidis et al., ). For example, most of IT or healthcare professionals are open to different interpretation of medical standards and diverge from the standard intended meaning and use them for different purposes, thus challenging the purpose of industry standards(Valle et al., 2005).

Recently HL7 has published SOA4HL7(Honey et al., ), a guideline for implementing healthcare services within a Service Oriented Architecture. SOA4HL7 complements the Service Specification Framework (SSF) defined within the Healthcare Services Specification Project (HSSP)2, but provides an additional interim method of defining and implementing Web services based on HL7 v3 artifacts. Two major integration issues are:

1. The SOA4HL7 profile is intended to provide a top-down “service based” approach, which means that the service definition (or service contract) becomes key and needs to be available to the client at design time. This requires a single “agreed” service definition model, in the form of a fully approved industry standard specification.
2. Even though the SOA4HL7 profile has separated the message definition from the service definition, a valuable input from the Healthcare Services Specification Project (HSSP), it still lacks

2http://hssp.wikispaces.com/
the separation of the “processes” from the services. This separation is important because each healthcare enterprise differ in their process models even if they follow same standard.

SOA-based enterprises expose their external behavior (public) without revealing the internal functionality or behavior (private). Such enterprise services interact with each other according to their behavioral description (externally and/or internally). We describe such behaviors as interaction behaviors. In this regard, the separation of interaction behavior (e.g. HL7 message exchange pattern\(^3\)) into a process (orchestration and/or choreography) layer is required to enable control and resolve the heterogeneity of interaction behavior. The semantic description of interaction behavior is called behavioral semantics. This behavioral semantics is the formal description which defines a service’s external (public) and internal (private) behavior. The external behavior describes a protocol that can be used by a client to consume the service functionality. The internal behavior describes a workflow, i.e., how the functionality of the service is aggregated out of other services (Vitvar et al., 2008). In our approach, a behavioral ontology is being developed for the semantic description of a service’s external behavior and functional ontology for internal behavior. Based on the above discussion PPEPR addresses the following integration requirements:

1. Identifies the “semantic gap” between and within SOA, HL7 WS and SOA profiles.
2. Applies ontologies (message, functional, behavioural) to resolve heterogeneity between message, service, and process definitions.
3. Provides a healthcare standard based flexible architecture that includes top-down and bottom-up development methodologies.
4. Follows a semi-automatic integration approach, where ontologies (schema level) are constructed and mapped at design time to be mediated at runtime (instance level).
5. Design and provide ontological reference to public (choreography) and private (orchestration) behavioral descriptions of healthcare services to resolve heterogeneity between process models.
6. Reduces the amount of mappings between heterogeneous messages, as compared to existing syntactic integration platforms.
7. Enables the “separation of concern”, between healthcare message, service, and process.

3\(^{http://www.w3.org/TR/wsd12-adjuncts/#meps}\)

3 sSOA FOR EPR INTEGRATION

Healthcare is a complex domain, comprising vendors, standards, legacy systems, and information systems which differ inherently from one another. Our core solution lies in enabling semantic interoperability between existing and new EPR systems. PPEPR is based on the design principles of a semantic SOA Reference Architecture and is built around semantic Web service technology(Vitvar et al., 2007). PPEPR’s architecture considers three types of integrations between EPRs based on their Web service capabilities (or lack thereof).

**EPR (non-Web service) ↔ EPR (non-Web service)**. This type of interaction is focussed on existing EPRs, which are mostly HL7 v2.x based.

**EPR (non-Web service) ↔ Web Service EPR**. This type of integration is the most complex (e.g. HL7 2.x ↔ HL7 v3), since EPRs (non-Web services) are required to communicate with the other EPRs (Web services).

**Web Service EPR (1) ↔ Web Service EPR (2)**. This type of integration offers the best interoperability solution by achieving the full potential of semantic Service-oriented Architecture (sSOA).

A detailed description of PPEPR’s architecture and its functioning (Sahay et al., 2008; Fox et al., 2008) is outside the scope of this paper. In this paper, our main focus is to present design and development methodologies for associating semantics with HL7’s message, service, and process definitions in service-oriented environment. The details of semantic Web service technologies [Web service execution environment (WSMX), Web service modeling language (WSML), Web service modeling toolkit (WSMT)] and the underlying conceptual framework Web service modeling ontology (WSMO)(Roman et al., 2005; Bussler et al., 2002) are outside the scope of this paper.

### 3.1 Semantics for Messages

Figure 1 describes how ontologies are grounded (lowering/lifting) to and from XML (Schema & Instance) and mediated by PPEPR. The grounding (lowering/lifting) task at runtime (instance level) is performed by PPEPR’s adapter framework and mediation by data mediator. Currently PPEPR can process messages in two formats, e.g., EDI and XML. The adapter framework stores the grounding rules, i.e., XSLTs\(^4\) developed at design-time for each message.

4\(^{http://www.w3.org/TR/xslt}\)
HL7 messages are uniquely defined by message-ids and the adapter uses these message-ids to identify and process each message at run-time. In figure 1 schema level integration (grounding & ontology mapping) is performed at design time and the instance level is performed at run time. To mediate a source to target message instance PPEPR requires source/target ontologies and mapping definitions defined at schema level. Section 4 describes the PPEPR’s methodology for developing, optimising HL7(v3 & v2) and mapping domain ontologies from message ontologies.

3.2 Semantics for Service & Process Definitions

As discussed above, PPEPR’s semantic Web service technologies are based on the WSMO framework(Roman et al., 2005). For a service’s external behavior, WSMO defines a choreography(Cimpian and Mocan, 2005) distinct from WS-CDL. WSMO defines a common global viewpoint of the observable behavior of collaborating services whereas in WSMO the choreography and orchestration is part of the interface definition of a service description. In PPEPR, the common global viewpoint is implicit as services are based on HL7 defined message exchange patterns and the behavioral ontology is designed for the semantic description of message exchange patterns.

The internal behavior of a service is semantically described by a functional ontology. HL7 categorises healthcare events and the PPEPR functional ontology is based on this categorisation, where each HL7 trigger event is a Web service within PPEPR. A functional ontology is a semantic description of HL7 based healthcare events. HL7 v2 & v3 differ syntactically in the structure of their trigger events. Therefore, functional ontologies are created and mapped based on their similarity. To model and execute message exchange patterns, it is necessary to employ a process modelling and execution standard which is able to reference ontological elements and allow their mapping within the model. BPEL for Semantic Web Services (BPEL4SWS(Filipowska et al., 2007; Karastoyanova et al., 2008)), a conservative set of language extensions to BPEL enables the referencing of ontological elements within a business process description. BPEL4SWS facilitates the orchestration of Semantic Web Services using a process based approach and is coupled with its ontological representation which is called sBPEL. In order to relate the semantics pertaining to one element in the BPEL4SWS description an additional attribute modelReference (like SAWSDL(Farrell and Lausen, 2007)) identifies the corresponding ontological instance in the sBPEL process model. The PPEPR mechanisms to discover and collaborate with services are end-point based (known at design time) in contrast to the WSMO goal-based mechanism.

Figure 2: Ontologising healthcare service (WSDL) & process definitions (BPEL).

Figure 2 describes the PPEPR approach for developing semantically-enabled service (WSML) and process (sBPEL) definitions. Level 3 illustrates the semantic descriptions (functional ontologies) of EPR services. Level 2 illustrates the semantic definitions of services and processes, and Level 1 is syntactic definition of services (WSSDL) and processes (BPEL)(Andrews et al., 2003). To integrate a new EPR, a semantic service definition (Level 3 → Level 2, top-down) is created first whereas existing systems are integrated in bottom-up (Level 1 → Level 2) fashion. This involves a manual process of transforming WSDL/BPEL to WSML/sBPEL. We are investigating means to incorporate the work in(Vitvar et al., 2008) to automate the WSDL ↔ WSML grounding task. Grounding and invocation of services are performed at the semantically-enabled middleware (WSMX).

PPEPR’s functional and behavioral ontologies are designed to annotate service and process definitions. Figure 3 shows how Web services (Order placer and fulfills) interfaces and internal workflows are annotated by behavioral and functional ontologies. The
bi-directional arrow indicates mapping between ontologies. Behavioral ontology provides the ontological reference to Web service interface, i.e., public external behavior. Functional ontology provides the ontological reference to internal workflows of order placer and fulfiller.

Figure 4 shows the choreography between the three actors of the PPEPR, where the Order placer (General Practitioner EPR) initiates the lab order fulfillment request. The request activates with a trigger event “Placer order activate” which maps to a similar trigger event OML^O21 (in case GP EPR (order placer) is HL7v2 compliant). Order fulfiller (Hospital Lab EPR) sends the confirmation receipt of an order followed by a trigger event (“filler promise activate(HL7v3)” or “ORU^O22(HL7v2)”) that sends a promise message (which can also be rejected) to fulfill the order. The final two messages are the lab test results sent followed by the confirmation from the order receiver (Hospital EPR) and order placer (General Practitioner EPR).

Figure 5 shows the process models of the order placer, fulfiller, and receiver required to achieve the actual healthcare process. It is sufficient if three actors, the process placer, the process fulfiller, and the process receiver model and execute a process according to the message exchange patterns defined in HL7 and shown in figure 4.

As we discussed above, HL7 not only defines the message content, but also the business logic (HL7 trigger events) to achieve certain functionality in the health care domain.

Figure 5 shows the business process models for Order Placer, Order Fulfiller, and Receiver in a HL7 Lab Test Order Request.

4 DEVELOPMENT METHODOLOGY

Figure 6 shows the development methodology for HL7 (v3 & v2) domain ontologies and their mappings. The methodology is divided into levels 1-3 (bottom-up) and levels 4-3 (top-down), where it meets at level 3. Table 1 shows the mapping between artifacts of both the HL7 versions. The Data Types, Segments, Data fields, and messages are mapped to equivalent HL7 v3 artifacts.

Initially, our approach for developing semantically-enabled messages, service, and process definitions was based on HL7’s message specifications and where all information related to message, service and process are tied together in message definition. This approach required that our previous approach to be “bottom-up” which means ontologies were developed from the existing HL7 messages (XML/XML Schema).

Our experience in PPEPR has told us that developing ontologies in a bottom-up fashion causes repetition in concepts and mapping rules. However, the bottom-up approach is still needed to connect with existing healthcare systems. HL7 profiles (Web-Service...
and SOA4HL7) open up the possibility of constructing ontologies first (top-down) for annotating message, service and process definitions. The major benefit of including both the approaches (top-down/bottom-up) is 50 percent reduction in the size of ontologies and number of mapping rules.

In figure 6 ontologies at level 4 (HL7 data types and vocabularies) were created first (top-down) as they are context-independent (deployment environment) and referenced by HL7 messages. All messages at level 1 are transformed to equivalent WSML representation. During the first stage of PPEPR development level 1 and level 2 were involved, then later we included the optimisation and merge processes, involving levels 3 and 4.

For the functional and behavioural ontologies, our development approach is top-down, since the nature of functional and behavioural ontologies is quite different from message ontologies. For example, messages are XML schema based and that involves various transformations from syntactic to semantic format and vice-versa, however functional and behavioural ontologies are mostly used for annotating semantically-enabled service (WSML), and process definitions(sBPEL) which are already in semantic format. In figure 2 transformation from level 1 to 2 is a manual task at the moment, and we are investigating means to incorporate the work in(Vitvar et al., 2008) to automate the WSDL ⇔ WSML grounding task.

A detailed description of PPEPR’s HL7 (v3 & v2) message development is outside the scope of this paper. This section focus more on domain ontology development from messages.

5 PPEPR ASSESSMENT

The following parameters are used to measure the impact of Semantics within PPEPR and effectiveness of PPEPR as an integration platform:

1. Design-Time

(a) Modeling HL7 message: The time taken for modelling HL7 ontologies, creating transformation rules (e.g., XSLT), and mapping definitions takes on average 1.5 days. A typical HL7 engine takes 0.5 days for mapping (syntactic). Similarly, PPEPR also takes 0.5 days for mapping (semantic). Therefore, extra work using PPEPR is 1 day for ontological modelling. The measurement was based on developers-recorded observations with good level of knowledge in HL7 and semantic technology tools. Each message within HL7 v3 consist of 49-51 ontological concepts. Each message within HL7 v2 consist of 36-40 ontological concepts. There are 102 mapping rules between ontological concepts of HL7 (v3 & v2) and on average similar number of rules are between other messages of each version. About 230-245 types of messages are contained in each version of the HL7 standard.

(b) Syntactic vs. Semantic Mapping: Syntactic mapping is predominantly based around the XML/XMLSchema level of expressivity. Due to the inherent nature of XML/XMLSchema, mappings are more at an implementation level and that causes a significant increase in amount of mappings. In PPEPR mappings are at the semantic(ontological) level which by nature maps two equivalent elements (concepts) at a higher level. The results have shown that the number of mappings reduced by up to 50 percent–PPEPR’s major milestone.

2. Run-Time

(a) Execution-time: The total message exchange time [message transformation, mediation and transmission] measured between two EPRs on typical broad-band connection is 2-3 seconds.

(b) Transformation: During the first stage of PPEPR development we tested the correctness of message transformation. The purpose of this test is to ensure that transformation (lifting/lowering) process is not losing the original message content and structure.

(c) Stability: In 3 months 250 messages has been exchanged on a PPEPR prototype with 100 percent success rate.
PPEPR can work as a standalone system directly interfacing with EPR systems or can be used as an addon to existing HL7 integration engines. The PPEPR software consists of two parts: The design-time and the runtime. The design-time portions of the system are used when installing PPEPR and configuring the various EPR systems which are to be made interoperable.

6 RELATED WORKS

COCOON (Valle et al., 2005)\textsuperscript{7} and ARTEMIS (Bicer et al., 2005; Valle et al., 2005)\textsuperscript{8} are 6th Framework E.U projects aimed at setting up semantics-based healthcare information infrastructure and developing semantic Web Services based Interoperability framework for the healthcare domain. The major differences between these eHealth projects and PPEPR are:

1. PPEPR requires no changes to exiting EPRs.
2. Others projects are Web-scale projects. The major focus of PPEPR is to ease the integration burden of healthcare enterprises. Additionally, PPEPR’s architecture is flexible enough to include Web-scale integration.
3. Others projects employ primarily top-down approaches as far as semantics (ontology development) for service oriented architecture is concerned. PPEPR incorporates both the methodologies (top-down/bottom-up).
4. PPEPR defines the clear “separation of concerns” for messages, services and healthcare process model.
5. PPEPR employs behavioral semantics to resolve behavioral heterogeneity of interacting healthcare services.
6. PPEPR ontologies are semantically interoperable, means that it can be easily used in other SWS frameworks like SAWSDL.
7. PPEPR ontologies are in WSML format and they are lightweight. It uses only the subset of WSML features that can be easily converted to other semantic languages [e.g. RDF/RDFS\textsuperscript{9}, OWL\textsuperscript{10}] without losing the semantics. The major motivation behind this is to be interoperable with other standard semantic Web languages.

\textsuperscript{7}http://www.cocoon-health.com/
\textsuperscript{8}http://www.srdc.metu.edu.tr/Webpage/projects/temis
\textsuperscript{9}http://www.w3.org/RDF/
\textsuperscript{10}http://www.w3.org/2007/OWL/wiki/OWL Working Group

RID\textsuperscript{11} and SemanticHEALTH\textsuperscript{12} are E.U roadmap projects with Special Emphasis on Semantic Interoperability. PPEPR has been influenced by the RIDE and SemanticHEALTH guidelines to design and develop a semantic solution to a core eHealth interoperability problem.

7 CONCLUSIONS

As we have discussed above, healthcare is a complex domain and any integration system, such as PPEPR, which connects healthcare enterprise applications must facilitate heterogeneous healthcare systems at all levels - data, services, processes, healthcare vendors, standards, legacy systems, and new information systems, all of which must interoperate to provide healthcare services.

In this paper, we describe the need of semantics in a service-oriented architecture (SOA) based healthcare integration system. This paper also presents our approach to include HL7 profiles (Web-Service and SOA4HL7) for ontologising message, service and process definitions, now we are able to deal with integration issues at separate levels. We analyse the integration requirements of HL7 compliant healthcare enterprises at message, service and process levels. We present the ontology development methodology that includes the top-down and bottom-up approaches to meet the enterprise integration requirements. The paper describes the latest results in the development of PPEPR, an integration system that connects enterprise healthcare applications at all levels (message, service, process, non-Web service, etc.). PPEPR’s architectural and ontological designs are domain based. These designs and ontologies include both standard based ontologies (message, functional, and behavioural) and the definition of approaches used to develop them. PPEPR ontologies are lightweight to be interoperable with other semantic languages and semantic Web service (SWS) framework.

In our future work we plan to focus more on optimizing ontologies. This will have the result of reducing the size of ontologies and mapping definitions. We see this as PPEPR’s core strength compared to syntactic integration solutions. We plan to automate the grounding tasks (from XML/XMLSchema/WSDL to WSML and back) for both the HL7 versions (v2 and v3). In addition, we plan to incorporate uses cases with more complex HL7 message exchange patterns within PPEPR.

\textsuperscript{11}http://www.srdc.metu.edu.tr/Webpage/projects/ride/
\textsuperscript{12}http://www.semantichealth.org/
ACKNOWLEDGEMENTS

We would like to thank Waseem Akhtar, James Cooley, and Armin Haller for their comments and input to this paper. The work presented in this paper has been funded in part by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lion-2) and the Enterprise Ireland under Grant No. CFTD 2005 INF 224 (SAOR).

REFERENCES


