Enhanced Semantic Interoperability by Profiling Health Informatics Standards

D. M. Lopez; B. Blobel

eHealth Competence Center, University of Regensburg Medical Center, Regensburg, Germany

Keywords
Integrated health care systems, standards, semantic interoperability, profiles, reusability, UML, HL7, Java components

Summary
Objectives: Several standards applied to the healthcare domain support semantic interoperability. These standards are far from being completely adopted in health information system development, however. The objective of this paper is to provide a method and suggest the necessary tooling for reusing standard health information models, by that way supporting the development of semantically interoperable systems and components.

Methods: The approach is based on the definition of UML Profiles. UML profiling is a formal modeling mechanism to specialize reference metamodels in such a way that it is possible to adapt those meta-models to specific platforms or domains. A health information model can be considered as such a metamodel.

Correspondence to:
Diego M. López, MSc.
eHealth Competence Center
University of Regensburg Medical Center
Franz-Josef-Strauß-Allee 11
93053 Regensburg
Germany
E-mail: diego.lopez@ehealth-cc.de
dmlopez@unicauca.edu.co

Results: The first step of the introduced method identifies the standard health information models and tasks in the software development process in which healthcare information models can be reused. Then, the selected information model is formalized as a UML Profile. That Profile is finally applied to system models, annotating them with the semantics of the information model. The approach is supported on Eclipse-based UML modeling tools. The method is integrated into a comprehensive framework for health information systems development, and the feasibility of the approach is demonstrated in the analysis, design, and implementation of a public health surveillance system, reusing HL7 RIM and DIMs specifications.

Conclusions: The paper describes a method and the necessary tooling for reusing standard healthcare information models. UML offers several advantages such as specification maintenance, tooling support, widely spread use, graphical notation, exchangeability, extensibility, semi-automatic code generation, etc. The approach presented is also applicable for harmonizing different standard specifications.

1. Introduction

Especially in the context of eHealth applications such as Electronic Health Record (EHR) systems to be used in long terms, at least two basic requisites for semantic interoperability need to be met: i) agreement on a standardized set of domain-specific conceptual models, and ii) agreement on standardized terminologies associated with controlled vocabularies [1].

Different health informatics standards currently exist defining domain models, such as ISO HL7 21731 “Health informatics – HL7 version 3 Reference Information Model” (RIM) [2], ISO EN 13606 “Health Informatics – EHR communication” [3], OpenEHR Reference Model [4], etc. These standards are far from being completely harmonized and adopted, however. Apart from political reasons, principal barriers for standards adoption include: lack of tooling and methodologies, methods for bridging between the different domains involved, missing implementation guidelines, incompleteness of standards, lack of conformance mechanisms, incomplete harmonization, etc.

Commonly, Standards Development Organizations (SDOs) use their own notation and an underlying development paradigm to describe their own standards. The Unified Modeling Language (UML) as the de-facto industry standard for software modeling has been used for many of those organizations as the language to formally represent the standards semantics. UML offers several advantages such as specification maintenance, tooling support, widely spread use, graphical notation, exchangeability, extensibility, code generation, etc. Unfortunately, UML has not been consistently used in healthcare standards development and – without appropriate adaptation – presents some limitations when describing specific domain concepts.

The objective of this paper is to formulate and demonstrate a method for reusing standard information models applied in the healthcare domain by exploiting the extensibility mechanisms of UML, and by that way supporting the development of semantically interoperable systems and components. If system models are based on such standard information models (e.g. specifying a platform-independent model), those system models can easily be transformed into different implementation platforms supported by Model Driven Development (MDD) approaches. Furthermore, when the resulting systems and components conform to the information structures and behavior described in the standard, semantic interoperability will be supported.
ability at functional and services levels is thereby supported.

2. Definitions and Methodology

Profiling in software engineering is defined as a mechanism for extending existing meta-models to adapt them to other purposes. In health informatics, the term “profile” has extensively been used in different ways. HL7, for example, uses profiles to constrain and extend HL7 specifications such as the RIM, Domain Information Models (DIMs), services specifications, etc. [5]. Profiling HL7 specifications allows using them in particular environments and/or defined communities of users (e.g. addressing specific requirements in some countries). The Integrating the Healthcare Enterprise (IHE) initiative [6] defines Integration Profiles for describing the solution to a specific integration problem (use case) in different domains, e.g. the IHE profiles for cardiology, radiology, IT infrastructure, etc., thereby reusing existing standards. OpenEHR defines the Archetype Profile which is a mechanism to specify customized Archetypes for certain information models.

The concept of profiling deployed in this paper was defined in the UML standard [7] as a formal mechanism to specialize a reference meta-model. This is done in such a way that it is possible to adapt that model to a specific platform or domain. Profiles are defined as UML packages containing stereotypes representing the meta-classes of the meta-model that can be extended. The stereotype is represented as a UML class along with its attributes (formerly known as “tagged values”), and is represented prepending the tag <<stereotype>> to the class name or using a customized shape (see class diagram in Fig. 3). A profile is formally defined as a UML model enabling to interchange such profiles between tools (using XML Metadata Interchange, XMI), together with the models applied. Constraints (restrictions) are also part of the stereotype definition as they provide additional information about the validity of the model instances.

In order to realize semantically interoperable solutions, HIS architectures have to conform to health informatics standards. This challenge implies the use, as much as possible, of those standards in the complete HIS architectural development cycle consisting of business modeling, requirements, analysis, design, implementation, deployment and maintenance phases [8, 9]. A more formal, customizable, scalable, and shareable methodology for Health Information System (HIS) architectural development existing at the moment is the Health Information Systems Development Framework (HIS-DF) [9]. HIS-DF is based on the Generic Component Model (GCM) and deployed as a specialization of the Rational Unified Process. It defines the activities to be performed within the course of the HIS development process (tasks), responsibilities within the software development team (roles), resulting models created in the different tasks (artifacts), as well as specific methods and examples (guidance) supporting the tasks in the development process. The framework also includes the definition of phases (time planning), workflows, etc., necessary for the architectural analysis of healthcare software development projects. Concretely, HIS-DF defines a task called “Harmonize Information Models” which specifies a set of generic steps to support conformance to standard health information models. The first step in this task is the specification of domain concepts, followed by the identification of related health informatics standard specifications. The final step is the harmonization (specialization) of the system analysis and design models with the identified standard specifications.

However, it is out of the scope of the HIS-DF methodology to define a specific technique (or method) to perform the process of harmonization to standards. To overcome this issue, the method presented in this paper formalizes and extends the harmonization process described in HIS-DF. It formulates and demonstrates a novel technique for system analysis and design models conformance to health informatics standards based on UML Profiles. Referring to the aforementioned HIS architectural development cycle, the difference between both approaches is that HIS-DF covers business modeling, requirements, analysis, and partially the design phases in a HIS architectural development project. Meanwhile, the method proposed in this paper addresses design, implementation, deployment, and maintenance phases of HIS architectures development.

3. A Method for Formally Profiling Health Informatics Standards

The introduced method was created with the objective of facilitating software architects and analysts to conform to health informatics standards in the process of healthcare development projects. The approach is integrated into the HIS-DF as a guidance artifact which supports its “Harmonize Information Models” task. The method formalizes and partially also automates this harmonization process defined in the harmonization task by formalizing existing health informatics standards using UML Profiles. It provides adaptations to the tooling necessary for identifying and creating those profiles, and demonstrates how to apply the profiles to the system UML models (e.g. class diagrams, sequence diagrams). The method is summarized in the following subsections.

3.1 Identify Standard Models and Artifacts in the Software Development Process

The profiling mechanism supports any of the three RM-ODP views described in the HIS-DF architectural development process. In the Enterprise Viewpoint, profiles can be defined for Use Case Models, Activity Diagrams, Business Rules, Storyboards, Business Use Case Models and other artifacts of the HL7 Development Framework (HDF) [5]. In the Information Viewpoint, such profiles can be defined to formalize HL7 RIM, DIMs, Refined Message Information Models (R-MIMs), Common Message Element Types (CMETs) or Templates. Finally in the Computational Viewpoint, profiles can be specified for HL7 Services Specifications, Application Roles, Trigger Events, Interactions, and Messages Specifications.

The proposed process for identifying the specific HL7 artifacts to be profiled consists of two steps. First, the HL7 Universal Domain(s) and/or the concrete HL7 sub-domains (Topics) for each one of system’s use cases (normally one per use case) are identified. Second, useful HL7 artifacts within the domain or sub-domain (R-MIM, CMET, Services Specifications, Application Roles, etc.) are identified.
3.2. Implement the Standard as a UML Profile

After having identified HL7 standard information models to be reused, the models are represented as UML Profiles. The classes and properties of the information model are described as stereotypes and attributes (tagged values), respectively. The Object Constraint Language (OCL) can be used to define restrictions to the stereotypes. In this stage, UML tools supporting the creation of profiles are necessary. Requisites for such tools include:

- capability to export the profile in XMI format for supporting further processing in other environments;
- ability to allow the definition of stereotypes and tagged values as modeling elements, being possible to define, e.g., optional and mandatory attributes;
- support of the creation of OCL expressions as part of the Profile, and
- validation of the correct application of tagged values and constraints (well-formedness) of the profiled models.

3.3 Apply the Profile to System Models

The previously defined UML system models are annotated with the standard information models by applying the profile. Conformance to the standard is guaranteed as the UML Profiles impose certain restrictions on how the UML model can be modified, excluding further changes to the profile. These restrictions should be supported by the UML tools, however. A mapping (applying rules either manually or tool-assisted) describes which classes of the original model correspond to the base classes in the profile. It is furthermore possible to automate the mapping process using model transformations. Transformation rules, however, need to always be defined by domain experts. Automatic transformation of models is out of the scope of this paper.

4. Using UML Profiles in a Health IT Development Project

In order to demonstrate how the introduced method can be used in a real software development project, a prototype of a health information system has been developed. The project supports the Integrated National Health Information System in Colombia [10] in the process of reporting public healthcare events (mainly mandatory reports of communicable diseases) from primary care providers (healthcare offices, clinics, hospitals) to the regional public health authorities. The system consists of three high-level use cases. “Receive Case” reflects a primary care provider reporting to the regional public health authority using a Web Form the occurrence of a Case. The use case “Look for Case” is initiated by a legacy Electronic Health Record system (EHR-S) or hospital information system running in the primary care provider. It allows the healthcare service provider to send a Case report derived directly from the EHR-S data repository, without any human intervention. Finally, the use case “Delivery Consolidate Report” aggregates and consolidates the information received about Cases, and delivers it to the local public health authority. For demonstration purposes, the “Look for Case” use case is the only one considered. The specification, analysis, design, and implementation stages of the complete information system have been developed and described. They are presented in [11].
4.1 Identify Standard Models and Tasks in the Software Development Process

Within the project development process, the proposed profiling method supports the analysis of the Informational Viewpoint as described in the HIS-DF methodology, more precisely the development of the System Analysis Model. Figure 1 describes the UML Analysis Model for the "Look for Case" use case. The automatic reporting process of a communicable disease event is initialized by the EHR-S interface which sends the information of a possible public health event (Case) to the "CaseInvestigationControl" class. This information is obtained from an ambulatory encounter event in the premises of a primary care provider, by matching the ICD-10 code that corresponds to the reported communicable disease. The CaseInvestigationControl class receives the message, extracts pertinent information, and creates a local instance of the public health event in its local data repository (database). The information about the event is made persistent through the entity classes IndividualCaseReport, SivigliaEvent, Patient and EventPlace.

In the Figure 1 diagram, the RUP Profile for domain analysis separating interface, control, and entity domain classes is used. The RUP Boundary stereotype represents the system’s interfaces whereas the RUP Control stereotype represents the classes realizing the use case functionality. Consequently, the RUP Entity stereotype represents persistent information in the system. It’s worth noting that while using UML Profiles, different purposes (semantics) of UML classes can be expressed in the same diagram as it will be demonstrated in a further stage of the application of the method.

The next step is the identification of the HL7 information models to be profiled. As a result of this analysis, the Case Management Topic as part of the HL7 Public Health Universal Domain was identified as candidate source of information models for the addressed problem. Concretely, the Notifiable Condition Report R-MIM (PORR_RM100001UV01) was identified as the information model that matches the problem domain.

4.2 Implement the Standard as a UML Profile

The UML Profile for the HL7 RIM is imported from the HyperModel tool [12], whereas the R-MIM Profile is created using the IBM Rational Software Architect modeling tool. The structure of the R-MIM Profile...
is shown in Figure 2. Hereafter, a UML stereotype is created for each entity, act, and role of the R-MIM, along with its attributes. The specialization of HL7 classes as defined in the HL7 specifications (RIM ← D-MIM ← R-MIM) is also expressed via stereotypes specialization/generalization associations. For instance, in the Figure 2 diagram, the PublicHealthCaseEvent stereotype generalizes the PublicHealthCase stereotype; it in turn generalizes an Observation stereotype which then generalizes an Act stereotype. The Profile is made available for reuse as an Eclipse plug-in and an XMI file.

### Table 1

<table>
<thead>
<tr>
<th>HIS Class</th>
<th>R-MIM Class</th>
<th>HL7 RIM Class (Stereotype)</th>
<th>HIS Attribute</th>
<th>R-MIM Attribute (Tagged Value)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>caseName</td>
<td>Text</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>type</td>
<td>detectionMethodCode</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dateOfnotification</td>
<td>effectiveTime</td>
</tr>
<tr>
<td>Patient</td>
<td>Person</td>
<td>InvestigatedPerson</td>
<td>patientID</td>
<td>Id</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>name</td>
<td>Name.given</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>secondName</td>
<td>Name.given</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>lastName</td>
<td>Name.family</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>secondLastName</td>
<td>Name.family</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>sex</td>
<td>administrativeGenderCode</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ethnicity</td>
<td>raceCode</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>populationGroup</td>
<td>etnicGroupCode</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>finalCondition</td>
<td>IndividualCaseReport.statusCode</td>
</tr>
<tr>
<td>IndividualCase Report</td>
<td>Observation</td>
<td>PrimaryCaseNotification</td>
<td>currentStatus</td>
<td>statusCode</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>reportID</td>
<td>Id</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>caseReportDate</td>
<td>effectiveTime</td>
</tr>
<tr>
<td>EventPlace</td>
<td>Place</td>
<td>Associated Place</td>
<td>PlaceCode</td>
<td>Code</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>placeName</td>
<td>Name</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>typeOfPlace</td>
<td>directionsText</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>geographicalArea</td>
<td>directionsText</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>municipality</td>
<td>addr.county</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>department</td>
<td>addr.state</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>country</td>
<td>addr.country</td>
</tr>
</tbody>
</table>

**4.3 Applying the Profile to System Models**

Following the harmonization process defined in the HIS-DF methodology, the analysis classes, attributes and data types in Figure 1 are mapped to classes, attributes and data types defined in HL7 information models, respectively. Considering that, and being HL7 information models restricted to static informational aspects of healthcare information, Boundary and Control classes in the model (e.g. EHR-S Interface, CaseInvestigationControl and CaseRegistrationConfirmationUI) are not mapped to HL7 classes.

Table 1 summarizes the mapping to classes in the HL7 RIM and the successive refinement to classes and attributes in the Notifiable Condition Report R-MIM Profile. The table furthermore defines the mapping rules necessary for the model transformation. Note that all the classes were matched to a stereotype in the Profile. For a few attributes, however, it was not possible to find an exact matching with the stereotype's attributes—the ones marked in italics. In those cases, the attributes needed to further be processes at programming level, for instance by splitting the attribute “Name.given” into a “name” and a “secondName” attributes. One alternative to this is the “localization” (adaptation) of the HL7 Information Model to the specific requirements of the surveillance system, which is however a task of the respective National HL7 organizations.

Figure 3 shows the resulting analysis model for the public health system after having applied the HL7 Profiles. For each class in the model, the applied stereotype is represented, along with the HL7 attributes and
data types extended from the HL7 RIM and R-MIM information models. For example, the SivigilaEvent class extends the <<PublicHealthCase>> stereotype defined in the R-MIM Profile. The classes are colored using the HL7 color coding functionality provided by the HyperModel tool. Also note that the RUP stereotypes (Boundary, Control and Entity) remain in the diagram.

5. Code Generation from the Profiled Models

Following, the implementation of applications based on the profiled models will be demonstrated. The Rational Software Architect (RSA) functionalities for MDA transformation were used to support the process of transforming the profiled analysis model into a set of Java implementation classes (platform-independent to platform-specific transformation). The MDA tool provides a layered, model-driven development approach which automates the transition between models at different levels of abstraction (for example, from analysis to design) and ultimately to code. ►Figure 4 shows the system’s Design Model after the UML to Enterprise Java Beans (EJB) MDA transformation has been applied. The Entity and Boundary classes in Figure 3 are transformed into Entity and Session EJBs components. As a result, the MDA tool generates skeleton Java classes and their correspondent code, which includes all HL7 attributes and data types. Getter and setter methods are also created for each attribute allowing the component’s information being manipulated. Furthermore, the UML to EJB transformation also implements the so-called component’s life cycle methods, e.g., ejbCreate(), ejbActivate(), setEntityContext(), etc. It’s important to mention that RUP Entity and Boundary Profiles previously applied to the classes facilitate its transformation to EJBs. This is because, according to the UML-EJB transformation rules, all RUP Entity classes are profiled with the stereotype "<<Entity Bean>>, being transformed into Entity EJBs and correspondently, all RUP Boundary classes are profiled with the stereotype "<<Service Bean>>, being transformed into Session EJBs.

Apart from the obvious benefit of automatically creating implementation components from UML analysis classes, the presented approach is potentially significant for creating HL7 implementation frameworks for different technologies. EJB, CORBA or .NET components can be created for each HL7 class (concept); conforming to the HL7 semantics (class types, entities, data types). Moreover, following the same model-driven approach, behavioral components (e.g. Control classes in the example) can easily be implemented, which manage any kind of HL7 messages using the available HL7 XML Schema Definition (XSD) documents. Implementation details of this approach defining generic HL7 components and the complete surveillance system are presented in [11].

6. Discussion

Few approaches suggest the use of UML Profiles to support the development of semantically interoperable systems. In [13], a method for profiling healthcare standards using UML is proposed. The method is not integrated into a development process, and the creation of ad-hoc profiles is enforced. In [14] an analysis of the German Health Card architecture based on the UMLsec Profile is presented, but restricted to security mechanisms and security policies. Carlson [15] developed a tool for MIF to UML transformation. Although UML Profiles for the HL7 HDF and RIM have recently been developed, they have not yet been used for HIS modeling. CaCORE [16] provides an infrastructure to semantically annotate UML information models with specific domain standards and vocabulary, but not exploiting UML profiling mechanisms. The openEHR archetype approach [17] proposes a model-based methodology which uses a proprietary domain-specific language, the Archetype Definition Language (ADL), to represent clinical or other domain-specific concepts by constraining instances of the openEHR information model, but not defining a formal development process nor conforming to the OMG MDA and UML profiling standards. Other approaches deal with the analysis and design of semantically interoperable HIS based on MDA or MDD approaches, e.g. [18–20], or development processes based on UML modeling [21, 22]; but none of them demonstrates a formal approach for health
informatics standards reuse. A similar approach was previously proposed by the authors [23], but the information models were manually extended using HL7 Visio tools. Furthermore, the approach did not follow the UML profiling mechanism, resulting in HL7 models incompatible with the UML 2.0 standard. Considering the relationships to HIS-DF [9], it also deals with methodological aspects of HIS development but at a higher level of abstraction (HIS-DF is a generic methodology covering theoretical aspects of HIS development based on established software development processes, meanwhile the proposed approach is a concrete method for health informatics standards formalization). They also provide different worked scenarios.

The method proposed in this paper for identifying and creating UML Profiles to healthcare standards offers the following advantages:

- **Support for Semantic Interoperability.** System models are annotated, extending standard information models and intrinsic vocabulary. HL7 version 3 information models were used to demonstrate the feasibility of the method because they are already defined based on a model-driven approach for message design – despite some fundamental inconsistencies –, and because of the broad availability of standard information models covering a wide spectrum of clinical information, along with business processes, and services [24]. Another reason is that it is the only approach at the moment that partially covers the business, informational and computational viewpoints defined in the HIS-DF methodology.

- **Formal Mechanism for Models Specialization.** A UML Profile provides mechanisms for specializing healthcare standards in such a way that the specialized semantics does not contradict the semantics of the generalized standard. The profile establishes restrictions on how the standard should be specialized; avoiding changes in the meta-class definitions.

- **Standard and Shareable Models.** UML is the de-facto standard language for software system modeling. UML models and profiles can be created and shared using the XMI standard. Most UML tools support the implementation of profiles. A challenge in the direction of realizing semantic interoperability is the deployment of repositories for the created models, profiles and components; so they can be easily located, retrieved and reused.

- **Support of Model-Driven Development.** The UML system models described are platform-independent information models. They can be transformed into platform-specific models and code (Web Services, J2EE, and CORBA) using UML transformations. Profiles can also support transformations (mapping) between standards.

The major limitation of commonly-used UML modeling tools (even the IBM Rational family) is the non-existing support of functionalities to validate well-formedness of the profiled models. Visio R-MIM Designer provides such functionality, but the resulting models cannot be processed by conventional UML tools, however. Further limitations result from incompatibility between existing
tools. For example, while using the specified tools in this project (despite being Eclipse-based tools), incompatibility between the latest versions of the HyperModel and the IBM Software Architect occurred due to incompatibilities between the different Eclipse SDK distributions. It can be expected that with HL7’s tooling migration towards the Eclipse platform, the R-MIM Designer will be extended by functionalities to process UML Models and Profiles.

Harmonization between different health informatics standards beyond the HL7 specifications and their integration into the development process is still an open challenge which can be met following a similar approach. Further research is necessary on the formalization of the model transformations and mappings. It includes the validation of the syntactic and semantic correctness of the model transformation with regards to the standards. It moreover embraces the creation and adaptation of tooling for domain-specific modeling and model-driven development. Furthermore, the proposed methodological approach opens the door towards self-organizing health information systems in the sense of autonomous computing approaches. Improvements of the formalization level of both the domain knowledge and the methodology by using, e.g., universal logics, formal languages, ontological languages need to be considered [8].

7. Conclusion

Analysis and design of health information systems, especially their underlying business and informational models describing basic concepts, business and relation networks, have to be based on standards. The presented paper formulates and demonstrates a method and the necessary tooling for reusing standard healthcare information models, by that way supporting the development of semantically interoperable systems and components. The approach is based on UML extensibility mechanisms by defining profiles. Using UML Profiles provides several advantages such as tooling support, graphical notation, exchangeability, extensibility, code generation, etc., deployable in the next generation of HL7 tools.

Acknowledgments

The authors are indebted to thank their partners from HL7, OMG, and other organizations for collegial collaboration. This work was partially founded by the German Academic Exchange Service DAAD and the University of Cauca in Colombia.

References