Ensuring HL7-based information model requirements within an ontology framework

Ouagne, David; Nadah, Nadia; Schober, Daniel; Choquet, Rémy; Teodoro, Douglas; Colaert, Dirk; Schulz, Stefan; Jaulent, Marie-Christine; Daniel, Christel

How to cite

OUAGNE, David et al. Ensuring HL7-based information model requirements within an ontology framework. In: Studies in health technology and informatics, 2010, vol. 160, n° Pt 2, p. 912–916. doi: 10.3233/978-1-60750-588-4-912

This publication URL: https://archive-ouverte.unige.ch/unige:157946

Publication DOI: 10.3233/978-1-60750-588-4-912

MEDINFO 2010
C. Safran et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-588-4-912

Ensuring HL7-based information model requirements within an ontology framework

David Ouagne[a], Nadia Nadah[a,b], Daniel Schober[c], Rémy Choquet[a], Douglas Teodoro[d], Dirk Colaert[e], Stefan Schulz[c], Marie-Christine Jaulent[a], Christel Daniel[a,f,g]

[a] INSERM, UMR_S 872, Eq. 20, Paris, F-75006 France; Université René Descartes, Paris, F-75006 France; [b] Heudiasyc CNRS/UMR 6599, Université de Technologie de Compiègne, Centre de Recherche de Royallieu BP 20529, Compiègne, F-60205 France; [c] University Medical Center Freiburg, Germany; [d] SIM, University Hospitals of Geneva, Geneva, Switzerland; [e] AGFA Healthcare, Sint-Martens-Latem, Belgium; [f] AP-HP, Hôpital George Pompidou, Département d'Informatique Hospitalière, Paris, F-75015 France; [g] ASIPSanté, Paris, France

Abstract

This paper describes the building of an HL7-based Information Model Ontology (IMO) that can be exploited by a domain ontology in order to distribute querying over different clinical data repositories. We employed the Open Medical Development Framework (OMDF) based on a model driven development methodology. OMDF provides model transformation features to build an HL7-based information model that covers the conceptual scope of a target project. The resulting IMO is used to mediate between ontologically queries and information retrieval from semantically less defined Hospital Information Systems (HIS). In the context of the DebugIT project - which scope corresponds to the control of infectious diseases and antimicrobial resistances – Information Model Ontology is integrated to the DebugIT domain ontology in order to express queries.

Keywords:

Information systems, Knowledge, Software design, Semantics, Information systems/methods.

Introduction

An important limit of most existing surveillance systems in healthcare is that they rely on local clinical data repositories based on proprietary data models without any semantic interoperability. The main objective of the DebugIT (Detecting and Eliminating Bacteria UsinG Information Technology) project [1], a 7th EU Framework Program, is to build an interoperability platform able to share heterogeneous clinical data sets from different European hospitals for the monitoring and control of infectious diseases and antimicrobial resistances. In this context, Semantic Web technologies are used to aggregate and query heterogeneous distributed clinical data in a unified view via ontologies. In particular, SPARQL[1] are Semantic Web services implemented to interface local clinical data repositories with ontologies.

A challenging issue is to ease the querying process between proprietary information models (IM) and domain ontologies. Indeed, it seems easier, as a first step, to map proprietary IMs to "mediator" ontologies that are instantiated by information entities. The binding of a domain ontology to an IM ontology has then to be done in a second step [2] in order to improve the coverage of the domain ontology.

This paper describes the methods and tools for building an HL7 IM based ontology that serves ensuring coverage of the conceptual scope within DebugIT by amending the DebugIT domain ontology (DCO)[2] with new "information entity" concepts. In this way the Information Model Ontology (IMO) ensures ontological coverage needed to query heterogeneous clinical data repositories while still largely complying to HL7.

Background

Healthcare standardization efforts

One major contribution of the standardization bodies in the health care domain (Health Level 7 (HL7)[3], CEN TC251, International Health Terminology Standards Development Organization (IHTSDO)[4], etc.) is to define the domain knowledge (reference business models, reference IMs, reference health care services and reference terminologies/ontologies) to enable semantic interoperability.

HL7 is an important standard for encoding clinical information. Its Reference Information Model (RIM) has been developed through a consensus process including harmonization activities. The RIM is the general structure that guarantees the coherence of the complex set of HL7 version 3 models that may be used in many contexts to describe particular administrative or clinical health care information. Besides IMs, HL7 also provides a controlled vocabulary that has been developed for coded properties of the HL7 IMs. Vocabulary specifications often refer to standard biomedical terminologies such as SNOMED CT.

---

1 http://www.w3.org/TR/rdf-sparql-query/

2 http://www.imbi.uni-freiburg.de/~schober/dco_owlDoc/

Model-Driven Architecture for ontology building

A new generation of Hospital Information Systems (HIS) integrate more and more standard IMs linked to biomedical terminologies, using state-of-the-art software development process methods, such as the Unified Process or Model-Driven Architecture (MDA) proposed by the Object Management Group (OMG)[3]. MDA defines a software development approach based on modeling and automated mapping of models [5]. Consequently, an increasing number of development frameworks are available allowing the use of MDA approaches into the HIS development process [6, 7, 8].

Only few of them provide the use of MDA associated to a healthcare standard into the HIS development process and investigate the use of ontologies mapped to information models [9, 10]. But, according to us, none of those development framework allow to build ontologies. Moreover, the use of the MDA approach for designing an ontology development platform is addressed by recent initiatives such as the OMG's one. Although ontologies and the Model-Driven Architecture (MDA) are approaches largely developed in parallel, many authors have attempted to bridge these modeling approaches [11, 12]. A software tool called DUET[4] enables the importing of DAML ontologies into IBM Rational Rose and ArgoUML and the exporting of UML models into the DAML ontology language. The tool is actually implemented as an add-in for IBM Rational Rose and as a plug-in for ArgoUML. It is freely available. XPetal[5] is a freely available tool implemented in Java that transforms mdl, the IBM Rational Rose model format, to the RDF and RDFS ontologies [13, 14].

In a previous work, we have developed the Open Medical Development Framework[6] (OMDF) that is a UML editor that supports adaptation of HL7 IMs according to local constraints. In this paper we present an additional extension that supports the building of an IM ontology.

Material

Conceptual scope of the DebugIT project

The objective of the DebugIT project is to aggregate clinical data, stored in different hospitals, in a unified system dedicated to the control of infectious diseases and antimicrobial resistances. The infection control scope of the stakeholders of the project is delineated via several queries. At the current stage, nine queries are being tested, e.g "percentage of patients with a given infection type (e.g urinary tract infection) by a given pathogen (e.g E.Coli) resistant to a given antibiotic (e.g trimethoprime (TMP)". The DebugIT data catalog includes 58 items covering the conceptual scope of the queries considered within the project. As a typical feature of IMs, it mixes information artifacts with references to real-world entities and is divided into four domains: microbiology lab results ("CULTURE") (11 items), antibiotherapy ("PATHOGEN TREATMENT") (7 items), patient data ("PATIENT DATA, PATIENT TREATMENT, PHYSIOLOGY & LAB FINDINGS") (31 items) and encounter ("EPISODE OF CARE") (9 items). Those items have been used as a resource in the development of DCO.

HL7 information models

We considered the January 2009 version of the HL7 ballot as the standard IMs source [3]. This ballot version includes 1285 IMs (e.g Administrable Medication or Result Event) that are available in the section Universal Domains, including 30 domain chapters (e.g. Clinical Statement, Laboratory, Medication, Orders, Pharmacy, etc.).

Open Medical Development Framework

The OMDF is a methodology and a platform dedicated to medical artifacts development. It uses the Model-Driven approach in the sense that it takes platform independent UML models and exports them into various output format. We have extended the functionalities of the UML editor Topcased[7] by integrating the following features: i) import of selected HL7 models, ii) model adaptation, iii) model transformation into various languages (html, sql, c#, etc.).

The "HL7 model import" functionality takes XML files provided by HL7 content information relating to the previously selected models to transform HL7 IM into UML [15]. The "models adaptation" function allows the designer to select the relevant classes to keep in the intended model. Only classes inheriting from "Entity", "Role" and "Act" of HL7 RIM are presented to the designer; other kinds of classes can be considered as structured relations. Relations are added automatically whenever two classes involved are part of the selection, thus simplifying specialization, aka model scope refinement, by the designer. The last feature, "model transformation" is the result of the integration of the Atlas Transformation language (ATL) engine[8]. This engine allows to implement a transformation (from one model to another) by describing the transformation process in an ATL file.

Method

We have extended OMDF in order to use the model-driven development methodology for designing an HL7-based ontology in the specific context of DebugIT. The ATL transformation engine of OMDF allowed the transformation of an IM in XMI syntax into an IM expressed in OWL syntax, which was then "ontologized" using Protégé and DCO. The Eclipse Modeling Framework for Semantic Web[9] was used to define the ATL syntax mapping file. For example, the UML classes will be mapped with OWL classes.

We followed a 3-steps methodology to derive the IMO from HL7 IMs.

[3] http://www.omg.org/
[4] http://marinemetadata.org/references/duet
[5] http://www.langdale.com.au/styler/xpetal/
[6] https://gforge.spim.jussieu.fr/projects/omdf/
[7] http://www.topcased.org
[8] http://www.eclipse.org/m2m/atl/
[9] http://code.google.com/p/eclipseuml2owl/

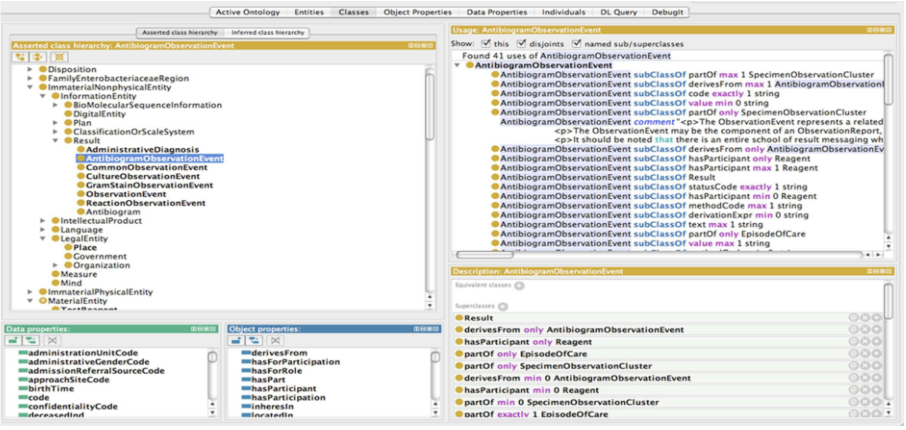

Figure 1- Protégé 4 GUI showing the classes and properties of IMO.owl file with DCO imported concepts (as "dco:InfomationEntity"). In this example, the right panel shows the usage of a highlighted "AntibiogramObservationEvent" class and below its auto converted formal definition

Identifying relevant sub-models

We investigated relevant standard IMs in the following HL7 Domains: Common Message Element Type, such as i) for encounter "EPISODE OF CARE", ii) for patient administrative data "PATIENT DATA", iii) for microbiology lab results "CULTURE", iv) for antibiotherapy orders "PATHOGEN TREATMENT", and v) for patient clinical data "PATIENT DATA & TREATMENT". We organized meetings with medical experts, business analysts and designers in order to browse the HL7 IMs of the chosen HL7 domains and select sub-models that covered the conceptual scope of the DebugIT domain.

Each HL7 domain is characterized by a representative model (Domain Information Model - DIM). We first compared the DebugIT catalog to each DIM to validate the choice of HL7 domains. Since the DIM frequently provides overly abstract information to be aligned directly with a data catalog, we then looked for more specific models within each domain. We evaluated the relevancy of a model according to the rate of mapping between the properties of the classes of the model and the items of the DebugIT data catalog.

Designing a DebugIT conceptual information model

The scope of DebugIT covers more than one HL7 domain (e.g. Laboratory, Order, etc.) and sometimes addresses more specific information. Therefore, we have been using OMDF to i) aggregate selected models into one model, ii) specialize this model to retain only the information relevant in the DebugIT conceptual scope.

Deriving IMO and completing DCO

We used OMDF in order to automatically transform the DebugIT conceptual IM into an OWL file. The problem with automatically generating OWL models from representational formalisms with different expressiveness is that the syntactical transformation may lead to semantically invalid statements. Therefore the target representation must be manually validated and if possible adapted aka "ontologized".

The initially exported OWL file is a semi-formal ontology in the sense that its semantics does not use DL expressiveness. The auto generated list of concepts, properties and attributes needs to be formalized manually.

The OWL file is edited into Protégé environment to add constraints on properties and classes. Then DCO is imported into the same OWL file to be completed with IMO's concepts by hand.

Experimenting queries using the ontology

Evaluating an ontology can be divided in two parts: verification and validation [16]. Verification refers to a technical process that aims at checking the correctness of the ontology. It deals with the formalization of the ontology and may be guaranteed by a reasoner. For instance to check the consistency of our ontology, we used a reasoner included into the Protégé editor (FaCT++[10]).

While the ontologization enabled us to use a classifier to check the logical consistency, the adequacy must be checked manually by a domain expert. An ontology is valid if it is useful to execute the task it has been built for.

The queries defined in the scope of the DebugIT project have to be expressed by IMO and DCO. As a matter of fact, the validation of IMO consists in testing the queries defined for the scope of the project.

[10] http://owl.man.ac.uk/factplusplus/

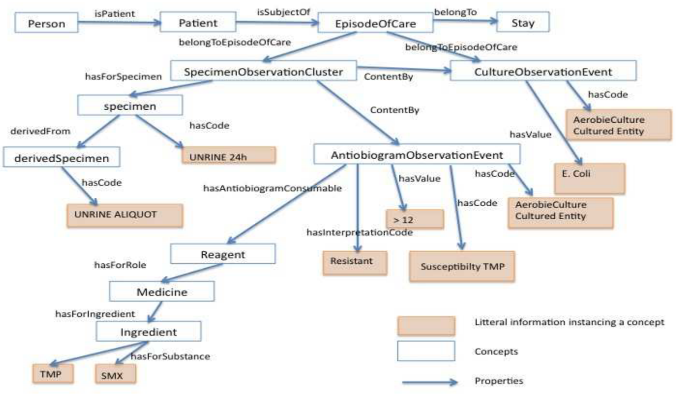

Figure 2- Instances of IMO concepts as needed within the scope of the query: "Percentage of patients with a given infection type (e.g. urinary tract infection) by a given pathogen (e.g. E.Coli) resistant to a given antibiotic"

Results

DebugIT HL7-based information model

OMDF enables us to design the conceptual model describing the information entities needed for the DebugIT project according to the HL7 standard. At this stage, this model is available as XML Metadata Interchange or HTML file. The six relevant HL7 IMs are: A_Encounter universal (COCT_RM010000UV01); Result Event (POLB_RM004000UV01); Composite Order (POOR_RM200999UV); Common Observation (POOB_RM410000UV); Adverse Reaction (REPC_RM000022UV) and A_BillableClinicalService Encounter (COCT_RM290004UV06). The IM includes 61 classes and 262 properties. The classes consist in Entity type classes (n=7) such as "Natural", "Person" or "Organization"; Act type classes (n=22) such as "ObservationEvent", "ObservationCluster", Role type classes (n=9) such as "Specimen", "Derived Specimen", "Patient", "Assigned Organization"; Participation type classes (n=10) and ActRelationship type classes (n=13). We have been able to express the totality of the DebugIT data catalog in the resulting HL7-based model. There is a one to one mapping with an HL7 property in 84% of cases. Seven per cent of the items were expressed using more than one HL7 property. For 4 items (9%), more than one item corresponded to one single HL7 property in the HL7 source IM. For example, EncounterStay.effectiveTime represents two items (admission date and discharge date).

The DebugIT HL7-based ontology

The OMDF extension enables us to get all the information we need about concepts and relations (cardinality, lexical information items as comments or definitions, domains and ranges for the properties) as shown in figure 1. The ontologization was done as follows : i) restructuring the concepts by adding subsumption properties between IMO concepts and concepts from DCO. So we imported and used classes as "clinical administration activity" "InformationEntity" "LegalEntity" or "PatientRole", ii) it is important to add constraints on the concepts and relations. "Admitter" and "Discharger" are both defined as linked to the concept "AssignedPerson" by a property. Nothing in the ontology states that a discharger is different from an admitter so that the reasoner defines both concepts as equivalents. So we have to create a disjointness axiom between those two concepts. The current version of the DebugIT HL7-based domain information includes 40 classes and 41 properties (10 Object properties and 31 Data properties). We used the reasoner (FaCT++) included into the Protégé editor. It took around twenty iterations the FaCT++ reasoner to obtain a consistent ontology. After each step, constraints were adjusted manually. Then, medical experts validated the completeness of IMO using the ontology to express the nine example questions defined in the DebugIT project. Figure 2 shows a sample of this validation process. It represents one of the queries expressed by IMO's concepts and properties.

Conclusion

Large-scale data integration efforts to support clinical and biological research are greatly facilitated by the adoption of standards for the representation and exchange of data. As part of the DebugIT project dedicated to multi-institutional sharing of disparate data on infectious diseases, we have explored the potential of the standard HL7 information models (IMs) for representing medical information entities and makes them usable through ontologies. We adopted state-of-the-art software development methods, such as the MDA approach proposed by OMG and have used the Open Medical Development Framework, developed in our laboratory, to support software designers and developers in adapting (importing and specializing) relevant HL7 IMs and transforming IMs into ontologies than can be mapped to domain ontologies.

We found that HL7 was a valuable source of artifacts and knowledge such as domain use cases, IMs and vocabularies. In our study, 100% of the DebugIT data catalog has been covered by the standard IMs selected in the HL7 Ballot. We experienced that HL7 IMs, though available in a specific format not handled by the UML modeling tools, could be converted to UML and used to build an IM ontology. HL7 IMs include many explicit comments about the meaning of classes and/or properties that are carried on from conceptual models to an OWL file, across all the transformations performed with Open Medical Development Framework. We experienced that the syntactical transformation that occurs while automatically generating OWL models from other representational formalisms may lead to semantically invalid statements. Indeed, the IMO derived from HL7 IMs does only yield a semantic network like OWL file including both information entities and real-world entities. Manual adaptation and validation are required so that this IMO conforms to the ontological assumptions of OWL. A potential outcome of this effort is an example of how epistemological assertions, as occurring in information models, can be modeled within an ontological framework. Thus, using one single representation formalism, our proposal brings together HL7 information models and philosophically founded ontologies. At this first stage of the DebugIT project, we focused on integrating HL7 and not CEN TC251 artifacts to the MDA approach. CEN TC 251 also provides reference models (the openEHR Information Model [17]) and defines in additional constrained models how the general reference model is used to describe particular administrative or clinical health care information. Although, integrating specific plug-ins dedicated to the use of HL7 artifacts, the Open Medical Development Framework is not structurally dedicated to HL7 standards and we will develop specific extensions dedicated to support CEN TC251 IMs as long as these IMs conform to a meta-model conforming itself to the Meta Object Facility (MOF).

Acknowledgments

This work was funded by the EU 7th FP project DebugIT (ICT-2007.5.2-217139).

References

[1] Lovis C, Colaert D, Müller H, Stroetmann VN. DebugIT for patient safety - improving the treatment with antibiotics through multimedia data mining of heterogeneous clinical data. Stud Health Technol Inform. 136 (2008), 641-6.

[2] Sundvall E, Qamar R, Nyström M, Forss M, Petersson H, Karlsson D, Ahlfeldt H., Rector A. Integration of tools for binding archetypes to SNOMED CT. BMC Med Inform Decis Mak. 2008 Oct 27;8 Suppl 1:S7

[3] HL7 version 3, Jan 2009 ballot package, 2009, http://www.hl7.org/v3ballot2009jan/html/welcome/environment/index.htm accessed August 2009.

[4] American Medical Informatics Association and American Health Information Management Association Terminology and Classification Policy Task Force, Healthcare Terminologies and Classifications: An Action Agenda for the United States, 2007, http://ncvhs.hhs.gov/080221p4.pdf accessed August 2009.

[5] Frankel D. Model Driven Architecture: Applying MDA to Enterprise Computing, John Wiley & Sons. 2003.

[6] Spronk R. The RIMBAA Technology Matrix, 2008, http://www.ringholm.de/docs/03100_en.htm accessed August 2009.

[7] Tuomainen M, Mykkanen J, Luostarinen H, Poyhola A, Paakkanen E. Model-centric approaches for the development of health information systems. Stud Health Technol Inform. 129 (2007), 28-32.

[8] Walderhaug S, Mikalsen M, Hartvigsen G, Stav E, Aagedal J. Improving systems interoperability with model-driven software development for healthcare. Stud Health Technol Inform. 129 (2007), 122-6.

[9] Lopez D, Blobel B. A development framework for semantically interoperable health information systems, Int J Med Inform. 78(2) (2009), 83-103.

[10] Oemig F, Blobel B. Semantic interoperability between health communication standards through formal ontologies. Stud Health Technol Inform. 150 (2009), 200-4.

[11] Baclawski K, Kokar M, Kogut ., Hart L, Smith JE, Letkowski J, Emery P. Extending the Unified Modeling Language for ontology development, Software and Systems Modeling, (2002) vol. 1, no. 2, pp. 142-156.

[12] Djuric D, Gasevic D, Devedzic V. Ontology Modeling and MDA. Journal of Object Technology, (2005) vol. 4, pp. 109-128.

[13] deVos A, Rowbotham CT. Knowledge representation for power system modelling, Proceedings of the 22nd International Conference on Power Industry Computer Applications, Sydney, 2001, pp. 50-56.

[14] deVos A, Widergren SE, Zhu J. XML for CIM Model Exchange, Proceedings of the 22nd International Conference on Power Industry Computer Applications, Sydney, 2001, pp. 31-37.

[15] Carlson D, Singureanu I. MIF to UML mapping tool, http://www.xmlmodeling.com/papers/ImportingMIFtoUML accessed January 2009.

[16] Fernandez M., Gómez-Pérez A, Juristo N. METHONTOLOGY: From ontological art to ontological engineering, in Spring Symposium Series, (1997) 33–40.

[17] Beale T, Heard S, Kalra D, Lloyd D. The openEHR EHR information model. (Revision 5.1.0), 2007, http://www.openehr.org/releases/1.0.1/architecture/rm/ehr_im.pdf accessed January 2009.

Address for correspondence

David Ouagne, CRC, Centre de Recherche des Cordeliers, 15 rue de l'école de Médecine, F-75006, Paris, France
Email: david.ouagne@crc.jussieu.fr