HL7-based information model ontology for EHR querying

D. Ouagne[a], N. Nadah[a,b], R. Choquet[a], D. Teodoro[d], C. Lovis[d], D. Colaert[e], D. Schober[f], S. Schulz[f], M-C. Jaulent[a], C. Daniel[a,c]

[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]AP-HP, Hôpital George Pompidou, Département d'Informatique Hospitalière, Paris, F-75015 France
[d]SIM, University Hospitals of Geneva, Geneva, Switzerland
[e]AGFA Healthcare, Sint-Martens-Latem, Belgium
[f]University Medical Center Freiburg, Germany

Abstract

This paper describes the building of the DebugIT Information Model Ontology (DIMO), an HL7-based ontology, and its use in the distributed querying process of different clinical data repositories. We employed a model-driven development methodology and used the Open Medical Development Framework (OMDF) providing model transformation features to first build the HL7-based information model that covers the conceptual scope of the project and then to derive the corresponding DIMO which is used to express queries. Further work is needed to evaluate to what extend this HL7-based information model ontology supports more complex queries corresponding to the infection control purposes of the stakeholders of the DebugIT project.

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 framework. The main objective of the DebugIT (Detecting and Eliminating Bacteria UsinG Information Technology) project [1], a 7th EU Framework Program (FP7) is to build a technical and semantic information technology platform able to share heterogeneous clinical data sets from different hospitals in different countries for the monitoring and the control of infectious diseases and antimicrobial resistances in Europe. In this context, Semantic Web technologies are used to aggregate and query heterogeneous distributed clinical data in a unified view thanks to an ontology. In particular, SPARQL Endpoints are Semantic Web services implemented to interface local clinical data repositories with the ontology.

A challenging issue is to ease the querying process between proprietary information models and an ontology. Indeed, it seems easier, as a first step, to map proprietary information models to ontologies that are instantiated by information entities (e.g. UML-derived ontology) rather than to ontologies that are instantiated to objects of the world (e.g. BioTop).

This paper describes the methods and tools for building an HL7-based information model ontology that covers the conceptual scope of the DebugIT project. This ontology may then be used to query heterogeneous clinical data repositories.

Background

Healthcare standardization efforts

One major contribution of the standardization bodies in the health care domain (Health Level 7 (HL7)[2], CEN TC251, International Health Terminology Standards Development Organization (IHTSDO)[3], etc.) is to define the domain knowledge (reference business models, reference information models, reference health care services and reference terminologies/ontologies) to enable semantic interoperability.

HL7 is one of the important sources of knowledge about medical information. HL7 provides a reference model - HL7 Reference Information Model (RIM) – that has been developed through a consensus process including harmonization activities. The RIM reference model is the general structure that guaranties 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 information models, HL7 also provides a controlled vocabulary that has been developed for coded properties of the HL7 information models. Vocabulary domain specifications often refer to standard biomedical terminologies such as SNOMED CT.

Model-Driven Architecture approach for ontology building

New generation of HIS designers and developers integrates more and more standard information models linked to biomedical terminologies in their software development processes. They use the state of the art of software development process methods, such as the Unified Process or Model-Driven Architecture (MDA) approach proposed by the Object Management Group (OMG). MDA defines a software development approach based on modeling and automated mapping of models [4].

Consequently, an increasing number of development frameworks are available allowing the use of MDA approaches into development of medical system [5][6][7].

But, our knowledge of a single development framework proposes the use of MDA in conjunction with a standard health [8].

Moreover, the use of the MDA approach for designing ontology development platform is addressed by recent initiatives such as the OMG's one. Although, ontologies and the Model-Driven Architecture (MDA) are two modeling approaches that are being developed in parallel, many authors have attempted to bridge the gaps and [9][10]. For example, Kateryna Falkovych et al. [11] have proposed a transformation of UML diagrams into DAML+OIL ontologies which enables the semantics of UML concepts to be preserved.

A software tool called DUET [12] 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 is a freely available tool implemented in Java that transforms mdl format, the IBM Rational Rose models format, the to RDF and RDFS ontologies [13][14][15].

In a previous work, we have developed the Open Medical Development Framework (OMDF) that is a UML editor that supports adaptation of HL7 information models according to local constraints [16]. In this paper we present an additional extension that allows derivation of a semi-formal ontology from adapted information models.

Material

Conceptual scope of the DebugIT project

The infection control purposes of the stakeholders of the DebugIT project correspond to several indicators. At the current stage, nine queries are being tested, e.g the query 5: "rate of patients with a given infection type (e.g urinary tract infection) by a given bug (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 current indicators of the project. It mixes information artifacts with reference 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).

HL7 information models

We considered the January 2009 version of the HL7 ballot as the standard information models source for achieving our objective. This ballot version includes 1285 in formation models (e.g Administrable Medication or Result Event) 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 Open Medical Development Framework (OMDF) is a methodology and a platform dedicated to medical artifacts development using the Model-Driven approach. We have extended the functionalities of the UML editor Topcased [17] by integrating the following features: i) import HL7 models, ii) HL7 models specialization, iii) transform specialized models into various languages.

The "import HL7 models" functionality is based on Carlson et al. work [18]. It is to take XML files provided by HL7 content information relating to the previously selected models to transform this information into UML.

The second function added allows the designer to select the classes to keep in 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 if two classes involved are part of the selection and thus simplified the work of specialization by the designer.

The last feature is the result of the integration of ATL transformation engine [19]. This engine allows to implement a transformation (such as a language to another) by describing the transformation process in a file (in ATL language).

Method

We extended OMDF in order to use the model-driven development methodology for designing an HL7-based semi formal ontology in the specific context of DebugIT.

We defined a mapping file between the meta models of UML and OWL so that the ATL model transformation engine of OMDF allows to export UML models in a format usable by an OWL ontology editor as Protégé. For example, the UML classes will be mapped with OWL classes.

We followed a 3-steps methodology to derive the DebugIT Information Model Ontology (DIMO) from HL7 information models.

Identifying relevant models

We investigated for relevant standard information models in the following HL7 Domains: Common Message Element Type (for encounter ("EPISODE OF CARE") and patient administrative data ("PATIENT DATA'), Laboratory (for microbiology lab results ("CULTURE")), Orders (for antibiotherapy ("PATHOGEN TREATMENT")), Clinical Statement (for patient clinical data ("PATIENT DATA & TREATMENT")). We organized meetings with medical experts, business analysts and designers in order to browse the HL7 information models of the chosen HL7 domains and select the models that covered the best 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 provides frequently too 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 the mapping between the properties of the class of the model and the items of the DebugIT data catalog.

Designing the DebugIT conceptual information model

The scope of the DebugIT project covers more than one HL7 domain (e.g. Laboratory, Order, etc). and addresses sometimes more specific information. Therefore, knowledge engineers and medical experts used 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 DIMO

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

The exported OWL file is a semi-formal ontology including the material (a list of concepts, properties and attributes) that has to be formalized. Protégé is then used to add constraints on properties and classes and if necessary add classes and properties to organize the semi-formal file into a formal and consistent ontology.

Experimenting queries thanks to DIMO

Evaluating an ontology can be divided in two points: verification and validation [20].

The first one refers to a technical process that aims to check the correctness of the ontology. It deals with the formalization of the ontology and may be guarantees by a reasoner. For instance to check the consistency of our ontology, we used a reasoner included into the Protégé editor.

The validation process concerns the usefulness of the ontology. An ontology is valid if it is useful to execute the task it has been built for.

The aim of DIMO is to express and run queries such as the query 5. In the scope of the DebugIT project, 9 queries have been defined. They have to be expressed by DIMO.

As a matter of fact, the validation of DIMO consists in testing queries as the previous one.

Results

We first present the extension of the Open Medical Development Framework (OMDF). Then, we describe the HL7-based information model and HL7-based ontology.

UML to OWL extension of OMDF

The ATL transformation engine of OMDF allows to implement a transformation (such as a language to another) by describing the transformation process in a file (in ATL language). The engine was used to transform a model in OWL file usable by Protégé. We used the Eclipse Modeling Framework for Semantic Web to define the ATL file [21].

DebugIT HL7-based conceptual information model

The conceptual model describing the information needed for the DebugIT project according to the HL7 standard was designed using UML and is available as XML Metadata Interchange file or HTML file. The model was derived form the six following HL7 information models: 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), that cover the conceptual scope of the DebugIT project.

The information model 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 Act relationships type classes (n=13).

One hundred per cent of the DebugIT data catalog was expressed in the HL7 based conceptual model. There was a one to one mapping with an HL7 property in 84% of the case. 7% of the items was expressed using more than one HL7 property. For 4 items (9%), more than one item corresponded to a single HL7 property in the HL7 source information model. For example, EncounterStay.effectiveTime represents two items (admission date and discharge date).

The DebugIT HL7-based semi formal ontology

As a matter of fact, OMDF enables us to get all the informations we need about concepts and relations (cardinality, lexical informations as comments or definitions, domains and ranges for the properties) as it is shown in the figure 1.

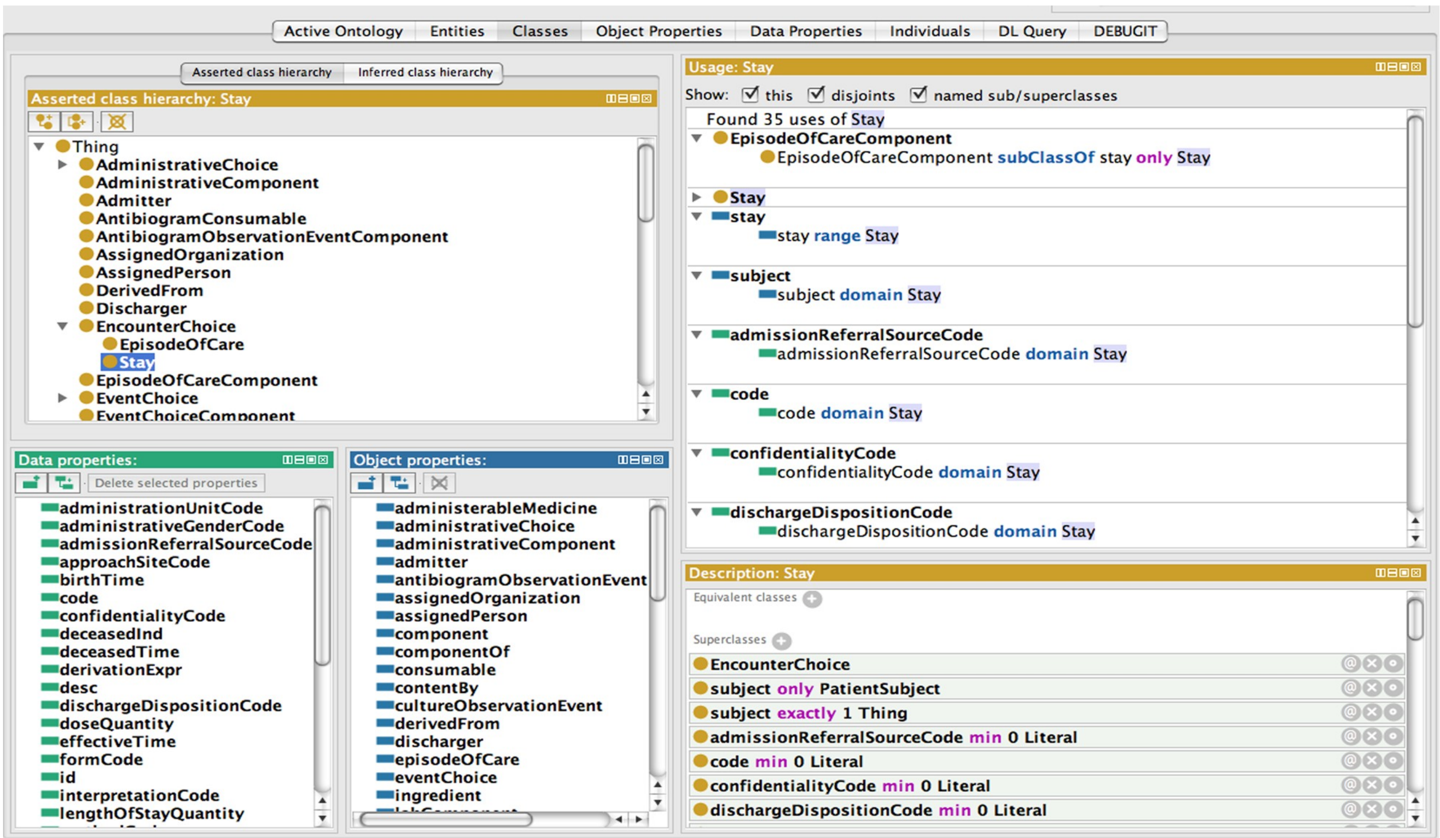

Figure 1: The OMDF exported OWL file opened into Protégé editor

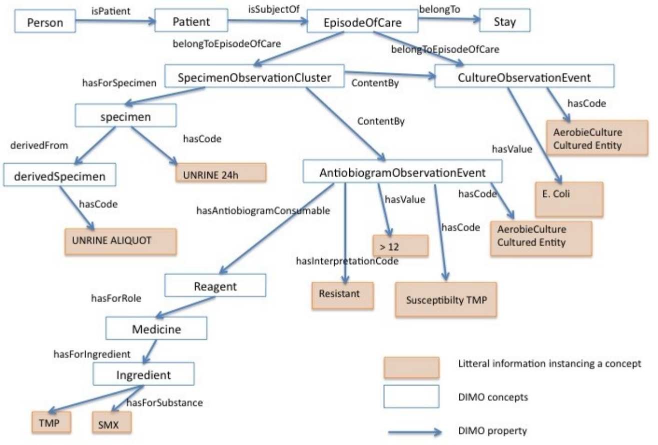

Figure 2: Instantiation using concepts and properties from DIMO corresponding to scope of the query: "Rate of patients with a given infection type (e.g urinary tract infection) by a given pathogen (e.g E.Coli) resistant to a given antibiotic"

The two next steps are the followings :

- restructuring the concepts by adding subsumption properties. For instance, "Admitter" and "Discharger" are both two kinds of participation. So we have to use the class "PARTICIPATION" and define "Admitter" and "Discharger" as subclasses of this concept;
- 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 semi-formal ontology precise that a discharger is different than the admiter. As well, if we try to infer from the export, the reasoner will define both concepts as equivalents. So we have to create a disjunction between those two concepts.

The current version of the DebugIT HL7-based domain information includes 61 classes and 79 properties (47 Object properties and 32 Data properties).

To check the consistency of the ontology, we used a reasoner included into the Protégé editor. After around twenty iterations (using the reasoner and adding constrains), we obtain a consistent ontology.

Then, medical experts validated the completeness of DIMO expressing the 9 queries defined in the DebugIT project by the mean of concepts and properties of the ontology.

Figure 2 shows a sample of this validation process. It represents the query 5 expressed by DIMO 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 biomedical data in the infectious diseases domain, we explored the potential of the standard HL7 information models for representing medical information concepts and provided a unified view of the clinical data of heterogeneous clinical data.

We adopted the state of the art of software development process methods, such as the Model-Driven Architecture (MDA) approach proposed by the Object Management Group (OMG). We used the Open Medical Development Framework, developed in our laboratory, to support software designers and developers in (i) adapting (importing and specializing) relevant HL7 information models, (ii) transforming information models into semi-formal ontologies.

We found that the HL7 was a valuable source of artifacts such as domain use cases, information models and vocabularies. In our study, we found that 100% of the DebugIT data catalog of the project was covered by the standard information models selected in the HL7 Ballot. We experienced that HL7 information models, though available in a specific format that is not handled by the UML modeling tools, could be converted to UML and be usable to build an information model ontology. HL7 information models include many explicit comments about the meaning of classes and/or properties that are carried on from conceptual model to the semi-formal ontology, 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 DIMO derived from HL7 information models doesn't adequately represent an OWL ontology. This semi-formal ontology is a semantic network that including both information entities and real-world entities and manual adaptation and/or validation are required so that this DIMO conforms to the ontological assumptions of OWL.

Like other related works, we propose a development framework to support the use of recent standard software engineering process concepts [6][7], and to make use of available standard medical knowledge artifacts [8]. 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 [22]) 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-in dedicated to the use of HL7 domain knowledge artifact, the Open Medical Development Framework is not structurally dedicated to HL7 standards and we could develop specific extensions dedicated to support CEN TC251 information models as long as these information models conform to a meta-model conforming itself to the Meta Object Facility (MOF).

Further work is needed to evaluate to what extend this DIMO supports all queries corresponding to the infection control purposes of the stakeholders of the DebugIT project. We will focus on positioning the DIMO to the DebugIT Core Ontology (DCO) developed in the project.

Acknowledgements

This work was supported by grant from FP7-ICT-2007-1- « Patient Safety » (DebugIT project).

References

1. C Lovis et al., 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. HL7, HL7 version 3, Jan 2009 ballot package, 2009, http://www.hl7.org/v3ballot2009jan/html/welcome/environment/index.htm accessed August 2009

3. 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

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

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

6. M Tuomainen et al., Model-centric approaches for the development of health information systems. Stud Health Technol Inform. 129 (2007), 28-32

7. S Walderhaug et al., Improving systems interoperability with model-driven software development for healthcare. Stud Health Technol Inform. 129 (2007), 122-6

8. D Lopez et al., A development framework for semantically interoperable health information systems, Int J Med Inform, 2008

9. Baclawski K, Kokar M, Kogut P, 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

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

11. Falkovych K, Sabou M, Stuckenschimidt H, UML for the Semantic Web: Transformation-based approaches. In Knowledge Transformation for the Semantic Web, eds. B Omelayenko & M Klein, Frontiers in Artificial Intelligence and Applications, (2003) vol. 95, IOS, Amsterdam, pp. 92-106

12. DAML-UML Enhanced Tool (DUET) Documents. http://marinemetadata.org/references/duet accessed August 2009

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. The Xpetal Tool. http://www.langdale.com.au/styler/xpetal/ accessed August 2009

16. Open Medical Development Framework (OMDF). https://gforge.spim.jussieu.fr/projects/omdf/ accessed September 2009

17. TOPCASED, Toolkit in OPen source for Critical Application on SystEm Development, http://www.topcased.org accessed August 2009

18. D Carlson et al., MIF to UML mapping tool, http://www.xmlmodeling.com/papers/ImportingMIFtoUML accessed January 2009

19. ATL, Atlas Transform Language, http://www.eclipse.org/m2m/atl/ accessed August 2009

20. M Fernandez, et al. METHONTOLOGY: From ontological art to ontological engineering, in Spring Symposium Series, (1997) 33–40.

21. EMF4SW: Eclipse Modeling Framework For Semantic Web, http://code.google.com/p/eclipseuml2owl/ accessed August 2009

22. T Beale et al., 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