Simplifying HL7 Version 3 messages

Robert WORDEN[a] and Philip SCOTT[b, 1]
[a] *Open Mapping Software Ltd, Cambridge, UK*
[b] *Centre for Healthcare Modelling and Informatics, University of Portsmouth, UK*

Abstract. HL7 Version 3 offers a semantically robust method for healthcare interoperability but has been criticized as overly complex to implement. This paper reviews initiatives to simplify HL7 Version 3 messaging and presents a novel approach based on semantic mapping. Based on user-defined definitions, precise transforms between simple and full messages are automatically generated. Systems can be interfaced with the simple messages and achieve interoperability with full Version 3 messages through the transforms. This reduces the costs of HL7 interfacing and will encourage better uptake of HL7 Version 3 and CDA.

Keywords. HL7, interoperability, standards, XML, mapping

Introduction

HL7 Version 3 is a semantic standard for healthcare messaging, in which the meanings of messages are defined in terms of a UML-based Reference Information Model (RIM) and its specializations for particular domains and communications, called Refined Message Information Models (RMIMs) [1]. RMIMs are defined in the Model Interchange Format (MIF). An XML Implementation Technology Specification (ITS) defines how instances of this model are serialised as XML documents or messages.

HL7 Clinical Document Architecture (CDA) is an application of HL7 Version 3 that defines the form and meaning of clinical documents [2]. CDA has been adopted as a standard by the NHS in England [3], it features in the US 'meaningful use' criteria for electronic health records [4] and has been used in eighteen other countries [5].

HL7 Version 3 has been criticized as overly complex and expensive to implement [6]. Local interoperability projects outside the scope of centrally-funded national programmes have generally lacked the technical and organizational infrastructure and rather arcane expertise that seems to be necessary to support a full Version 3 development and deployment. Consequently, the semantically inferior HL7 Version 2 has tended to remain the *de facto* standard for local level interactions. This paper presents a novel approach to Version 3 message simplification that promises to enable wider adoption and thereby improve semantic interoperability in healthcare.

[1] Corresponding Author: Dr. Philip J. Scott, Centre for Healthcare Modelling and Informatics, University of Portsmouth, Buckingham Building, Portsmouth PO1 3HE; E-mail: philip.scott@port.ac.uk.

1. Why HL7 Version 3 Messages are Complex

RIM-based information modelling is semantically robust but practically complex. Models are constructed as networks of a few core RIM classes (notably the classes Entity, Role, Participation, Act and ActRelationship), linked together by a few types of association. 'Structural attributes' on the classes define their role in a particular model. This obligatory design pattern leads to a consistent style of models across all domains. Common data modelling pitfalls (such as a failure to distinguish between individuals and their roles) are exposed and avoided. For a RIM expert, semantic insights are easily transferred from one domain to another, leading to sound and consistent models.

RIM-based semantic models are substantially more complex than 'natural' data models or class models in UML that are readily understood by domain experts. (These simpler models, called Domain Analysis Models (DAMs), are intended to be used as a first stage of analysis and to be traceable to the RIM-based models; in practice they are rarely maintained or used after initial analysis).

A RIM-based model typically has two or three times as many classes and associations as the domain model that conventional analysis would produce. It also has many fixed attributes whose values do not change from one model instance to another. Applying a 'one size fits all' semantic modelling approach across diverse healthcare domains has produced models which are overly complex. This extra complexity makes the models quite unapproachable for RIM non-experts and it makes the XML serialisation of the model instances large and complex to read and write.

CDA has a further level of complexity in that it requires specialization using various forms of template. For instance, a CDA 'Entry' may contain a wide range of discrete pieces of clinical information, but templates must be specified to say what type of information is in each entry. Many hundreds of CDA templates have been designed, and their management is a major source of complexity.

An XML instance of HL7 Version 3 or CDA can be deeply nested – twenty levels of nesting is not unusual, with each level representing an association in the RIM-based model of the domain – and typically has large numbers of fixed attributes, whose values do not change from one message instance to the next. Scattered across this large XML structure are a comparatively small number of items of variable information, constituting the actual information content of the message.

2. Approaches to Message Simplification

The aim of message simplification is to define a simple XML message format – with shallower nesting and with fewer fixed attributes than a Version 3 or CDA message – but which conveys the same variable information, and can be reliably transformed to and from a full Version 3 message. Once a simplified message has been defined (with transforms to and from the full messages), then a developer can interface systems to full Version 3 messages just by writing software to read and write the simplified messages and using the transforms to convert to and from the full messages. This could greatly reduce the difficulty of building HL7 Version 3 interfaces.

It has been recognised for some years that HL7 message simplification is feasible and useful. We shall briefly note work done in Canada and the USA, before proceeding to give a fuller account of work in the UK.

The Structured Documents workgroup of HL7 International has sponsored a project called 'greenCDA' whose aim is to provide simpler ways of writing and parsing CDA documents. This project has focused on a profile of CDA called the Continuity of Care Document (CCD) which is widely used in the US. The 'meaningful use' criteria set by the US federal government have aroused commercial interest in interfacing systems to CCD and a specialization of it called C32. The greenCDA project has produced a simplified CCD using a hand-designed XML format with meaningful business names for elements and attributes and an XSLT to transform the simple form to a conformant instance of CCD/C32. The US Centers for Disease Control (CDC) is working on a greenCDA version of their Healthcare Associated Infection report.

There is significant effort involved in developing and testing the XSLT transforms. It remains to be seen how well these techniques scale to the volume of CDA templates in use, and how feasible it is to maintain the simplified message formats and transforms. This may be important in the light of (a) development of the underlying standards and profiles through successive versions, and (b) the iteration and experiment that may be needed to develop appropriate simplifications. We expect that the transforms can be developed in reusable pieces. The question remains how much this modularisation will save effort and reduce errors in transform development.

Canada Infoway has been applying message simplification since 2007, using Version 3 reshaping rules [7]. These rules allow associations in RIM-based models with maximum multiplicity 1 to be 'flattened' (removed from the model, removing a level of XML nesting from the message) wherever doing so would allow the original model instances to be recovered precisely. The tooling to do this in built into the Version 3 generator (the design tool for Version 3 messages). It allows the renaming of elements and attributes to give meaningful business names, and allows some associations which are eligible to be flattened to be preserved, if so wished.

A feature of this approach is that the tools produce not only simplified messages, but also a simplified object/class model and programming interfaces to develop software against that model. Run-time translation in both directions between the simplified object model and the full Version 3 messages is done by automatically generated code so the translation between the simplified model and the full message is guaranteed to be accurate. The simplified model produced in this approach is also useful for domain experts to validate that the model meets their business requirements without having to master the technical complexity of a Version 3 RMIM. These simplification techniques are applicable to messages defined in the static model designer as RMIMs, but do not appear to yet support templates as required for CDA.

3. Mapping-Based Message Simplification

This section reports novel work initiated in the UK to simplify Version 3 messages using a semantic mapping approach. This method maps different data structures not directly to one another, but each to a common UML class model. If any two data structures can be mapped to a shared class model, then open-source tools exist [8] to automatically generate an accurate transformation between the two data structures.

The data structures are defined as XML schemas. The tooling is built on the Eclipse tool framework, using the Eclipse Modelling Facility (EMF). EMF has a notation called Ecore for representing UML class models. Mappings define how each XML data structure represents information in the class model. Transforms between the

structures can be generated automatically from the mappings. When a simplified message is defined, both the simplified message and the full message are automatically mapped to the same simplified class model. The tools can generate and execute reliable transformations in both directions between the simplified message and the full message. This is shown in Figure 1.

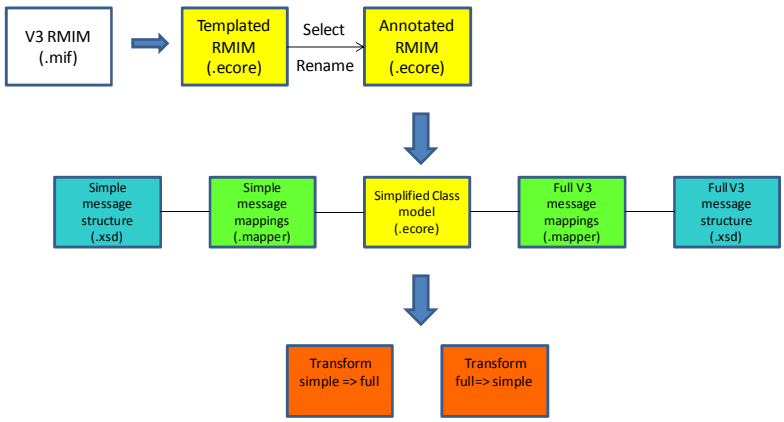

Figure 1 HL7 Version 3 message simplification process using semantic mapping.

In Figure 1, the only manual step is that marked 'select/rename'. The definitions of Version 3 messages are supplied in the HL7 Message Interchange Format and the tools automatically convert these to an EMF Ecore model. If the Version 3 message is a CDA with templates, the tools apply the templates to produce a templated Ecore model.

The templated RMIM is a very large tree structure, which corresponds precisely to the tree structure of the XML messages. For some CDA applications, the tree may have millions of nodes, even without permitted recursive self-nesting of subtrees (which makes the allowed number of nodes infinite). The HL7 analyst who defines a message simplification uses the tools to navigate this large RMIM tree, adding annotations to define the simplified message. This involves:

- Marking those leaf nodes that must be populated for the given interchange
- Marking the internal nodes to be 'flattened' to make a shallower XML structure
- Defining meaningful business names for all retained nodes.

After annotating the class model, further user interaction is minimal. Simplified messages are typically about three times smaller than canonical messages [9]. Simplification works best for tightly defined profiles of Version 3 or CDA messages, where the information to be transferred between systems forms a well-bounded set.

This approach is being tested and validated in a UK project to support the Care Assessment Framework (CAF) processes as part of Healthcare and Social Care Integration (HSCI), for which the NHS has defined a set of five CDA-based messages.

4. Discussion

We believe that the approach outlined here satisfies the following critical success factors for message simplification:

- **Scope**: applicable to any types of HL7 message or documents;
- **Transform Reliability**: demonstrably reliable and testable two-way between simplified and canonical Version 3 messages;
- **Semantic Precision**: fully and clearly defined in HL7 semantics;
- **Ease of Use**: straightforward to interface systems to the simplified messages, significant cost savings compared to direct V3 interfacing;
- **Breadth of Use**: applicable for model-based development, model-based comparative query, or validation of domain models;
- **Development of Simplified Message Definitions:** process of defining simplified messages automated and reliable, scales well and is easily repeatable as message definitions change between versions.

A broader consideration is how the document messaging paradigm (simplified or not) will fit with service oriented architectures (SOA) for healthcare interchanges [10].

5. Conclusions

Message simplification can greatly reduce the costs of HL7 Version 3 interfaces, hiding the technical complexity of the RIM while preserving its robust semantics. We expect to see continuing development and use of HL7 simplification tools and techniques.

References

[1] Hinchley A. *Understanding Version 3*. 4th ed. Munich: Alexander Mönch; 2007.

[2] Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, et al. HL7 Clinical Document Architecture, Release 2. J Am Med Inform Assoc. 2006;13(1):30-9.

[3] Information Standards Board for Health and Social Care. Interoperability. n.d. [cited 12 April 2011]. Available from: http://www.isb.nhs.uk/use/baselines/interoper

[4] Office of the National Coordinator for Health Information Technology. Health Information Technology: Initial Set of Standards, Implementation Specifications, and Certification Criteria for Electronic Health Record Technology. Federal Register. 2010;75(144). [cited 17 January 2011]. Available from: http://www.gpo.gov/fdsys/pkg/FR-2010-07-28/pdf/2010-17210.pdf

[5] Kaminker D. HL7 Clinical Document Architecture Ambassador Briefing. 11th International HL7 Interoperability Conference, Rio de Janeiro, Brazil; 2010.

[6] Browne E. openEHR Archetypes for HL7 CDA Documents. 2008. [cited 17 January 2011]. Available from: http://www.openehr.org/

[7] McKenzie L, Stephens M. Message reshaping rules. 2007 [cited 17 January 2011]. Available from: http://wiki.hl7.org/index.php?title=Message_reshaping_rules

[8] V2 and V3 mapping tools. 2009 [cited 17 January 2011]. Available from: http://gforge.hl7.org/gf/project/v2v3-mapping/frs/

[9] Worden R, Scott P. Fragmentary example of simplified XML. 2011 [cited 28 April 2011]. Available from: http://userweb.port.ac.uk/~scottp/MIE2011/Example.pdf

[10] Kawamoto K, Lobach DF. Proposal for fulfilling strategic objectives of the U.S. Roadmap for national action on clinical decision support through a service-oriented architecture leveraging HL7 services. J Am Med Inform Assoc. 2007 Mar-Apr;14(2):146-55.