PUBLISHING RELATIONAL MEDICAL DATA AS CEN 13606 ARCHETYPE COMPLIANT EHR EXTRACTS USING XML TECHNOLOGIES

Rinner C[1], Wrba T[1], Duftschmid G[1]

Abstract

We implemented a proof-of-concept layer to export medical data stored in a health information system as prEN 13606 EHR extracts. The export-layer is based on SQL and the DOM API. A method is presented for the description of an EHR extracts' structure using XML Schema. Using this method EHR extracts can be transformed into Archetype compliant EHR extracts. The transformation is based on XML Schema and XSLT.

1. Introduction

The European Committee for Standardization (CEN) is currently working on a five part European Standard (prEN 13606 [1]) that defines an information architecture for communicating contents of electronic health records (EHR). This standard enables health information systems (HIS) to exchange EHR segments in a semantically interoperable manner. It is based on the two-level methodology which separates knowledge from information. In part one of the standard the Reference Model (RM) is described. Instances of the RM can represent all medical information that may be contained within an EHR, independent of the originating system. In part two the concept of Archetypes (AT) [2, 3] and a language for specifying ATs is defined. ATs contain the knowledge, in other words they act as 'construction plans' for instances of the RM. By using the RM and ATs it is possible to exchange information in a semantically interoperable manner.

ArchiMed [4] is a clinical trial system developed at the Medical University of Vienna (Core Unit for Medical Statistics and Informatics). It allows to collect and to statistically analyse data, it supports patient recruiting and the interactive design of case report forms (CRFs). The data model used in ArchiMed is based on the Entity-Attribute-Value (EAV) design that is frequently used in HISs used for clinical research [5].

We applied the pre standard prEN 13606 and implemented a proof-of-concept export layer to export data stored in ArchiMed as archetyped EHR extracts using standardized XML mechanisms. In this paper the term *EHR extract* refers to an XML document compliant to the prEN 13606 RM but not conforming to an AT, whereas *archetyped EHR extract* refers to an EHR extract that additionally conforms to an existing AT.

1 Core Unit for Medical Statistics and Informatics, Medical University of Vienna

In [6] a self-made schema is used to export existing medical data. Our export-layer is based on XML Schema and tries to use standard XML mechanism to create archetyped EHR extracts. In [7] a similar approach is used to export data from another two-level modelled information system.

2. Methods

We propose to divide the publishing of archetyped EHR extracts into two steps, see figure 1.

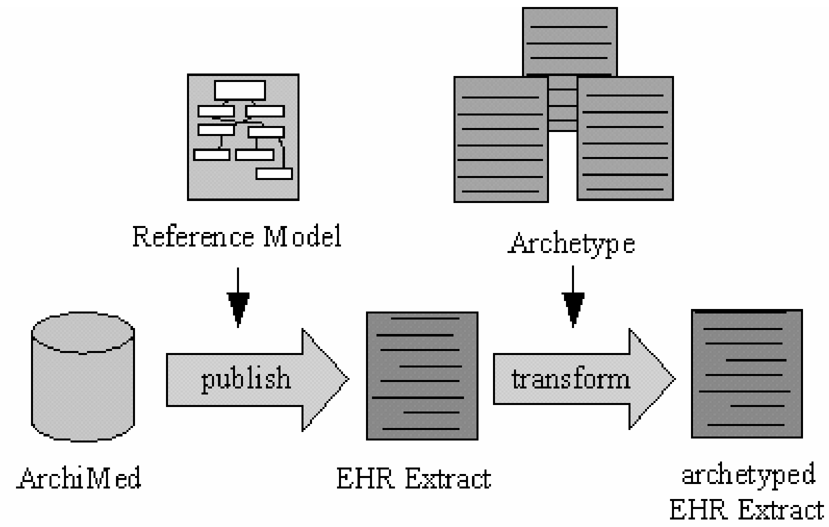

Figure 1: Two level publishing of archetyped EHR extracts

In the first step we publish the data needed from ArchiMed as EHR extract. Mapping the structure of CRFs to classes in the RM allows to generically publish medical data collected by means of any CRF as EHR extract. Publishing the data as an instance of the RM ensures that all data, independent of the originating system, can be represented as an XML document.

In the second step, the EHR extract that reflects the structure of the CRF is transformed into an archetyped EHR extract. By conforming to an AT the EHR extract is semantically interoperable and can easily be shared between different HISs. The transformation is done using Extensible Stylesheet Language Transformation (XSLT) scripts.

Since manual creation of the XSLT scripts is highly laborious we use a mapping tool [8]. It allows the mapping of XML documents by visually associating the components of their underlying XML Schemas. Due to the *unique particle attribution* however it is not possible to create an XML Schema that fully describes the structure of an archetyped EHR extract [9]. It is not possible to restrict the substructure of two XML elements at the same hierarchy level within an XML Schema if they have the same tag name (e.g. two XML elements with the tag name <ELEMENT>, one containing a systolic blood pressure and the other a diastolic blood pressure). We avoided this restriction by creating "virtual" subclasses of those record components that are referred to within the AT. These classes are virtual in the sense that their only purpose is to allow an XML Schema representation of an AT. In the final step of the publishing process, the XML elements that instantiate the virtual subclasses are typecasted back to the corresponding superclasses. Thus, the virtual subclasses only appear temporarily during the transformation process, the finally published archetyped EHR extract of course refers to the original RM classes.

The virtual subclasses are named after their superclasses concatenated with the name defined in the AT without spaces and camel-back notation. Naturally the subclasses inherit all the attributes and relations. An example of this so called "extended Reference Model" can be seen in figure 2. The original RM classes are shown in gray, the classes in black are the virtual subclasses. The extended RM also contains virtual subclasses for each component of the published CRF to allow pinpointing the sources for the following mapping process.

By referring to the virtual subclasses of the extended RM we can now create two XML Schemas

- that precisely specify the structure of the EHR extract respective the archetyped EHR extract, and
- for which a mapping can be defined as the basis of the transformation process depicted in figure 1.

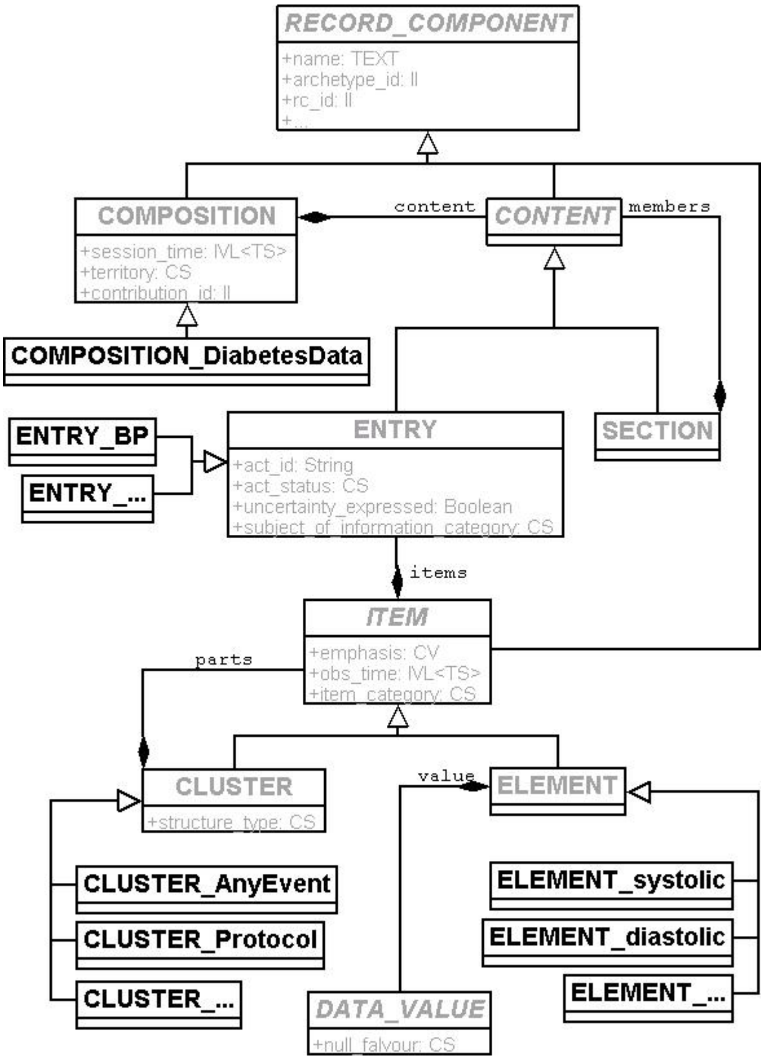

Figure 2: extended Reference Model

An EHR extract that refers to the extended RM is in the following called *extended EHR extract.*

The XML Schemas describing the extended EHR extract and the extended archetyped EHR extract can be visually mapped and result in the generation of the needed XSLT script. This XSLT script can then be applied to the extended EHR extract. By renaming the tags corresponding to virtual subclasses in the resulting extended archetyped EHR extract to the names of their superclasses (e.g. ‘ELEMENT’), the final archetyped EHR extract is created.

3. Results

An existing SECTION Archetype called “private diabetes data” [10] consisting of five ENTRYs (blood pressure, laboratory-glucose, heart-rate pulse, body weight, dimensions and physical activity) modelled in the Archetype Description Language (ADL) [1] was used. Our scenario assumes that EHR extracts have to be compliant to predefined ATs to simulate the use of existing official ATs to enable semantic interoperability.

We developed PL/SQL procedures in ArchiMed’s Oracle® database engine to retrieve the data originating from the corresponding CRFs as EHR extracts as well as extended EHR extracts. The procedure uses the Structured Query Language (SQL) and the Document Object Model (DOM) API in combination with the structural information from the RM. The structure used within CRFs was mapped to classes of the RM according to the schema shown in table 1. The EAV design of the database simplified the implementation of a generic export.

Table 1: Mapping of concepts in ArchiMed to classes in the Reference Model

ArchiMed	Reference Model
CRFs	COMPOSITIONs
Pages	SECTIONs
Tables	CLUSTERs
Entries	ELEMENTs

The XML Schema describing the extended archetyped EHR extract has to be built manually using the knowledge stored in the AT and the information from the RM. All obligatory attributes from the RM such as *rc_id*, *name* and the structure prescribed by the AT are manifested in this XML Schema.

Mapping the XML Schemas of the two extended EHR extracts was done using a mapping tool [8]. In cases where the AT specifies additional structures (e.g., SECTION to further structure the EHR extract) they were created in a separate XML document and also used in the mapping process. The resulting XSLT was applied to the extended EHR extract. For the moment renaming the subclasses to get an EHR extract from the extended EHR extract is done manually.

To be able to trace back an EHR extract and single RECORD_COMPONENTs (RCs) to their origin in the database we assign an object identifier (OID) to every table of the ArchiMed EAV model-based database. Additionally, we use the unique keys and the *version-ids* from ArchiMed’s version trail to generate the Instance Identifier (II) used in the RC’s id.

If data needed to create an archetyped EHR extract is scattered to different CRFs, their content is compiled to the same generic EHR extract into separate COMPOSITIONs. The selection of the needed data is done during the XSLT script generation process.

4. Discussion

Using the extended RM allows XML Schema to describe the structure of EHR extracts. By using XML Schema it is possible to take advantage of the experience gained with XML technologies and existing XML tools can be applied. A shortcoming of XML Schema is that it is not capable to constrain values within EHR extracts to the same extent as ATs. As an example, constraints that involve more than one extract elements (e.g., systolic pressure should be ≥ diastolic pressure) can be represented by means of the ASSERTION class within an AT but cannot be expressed within an

XML Schema. An alternative to the XML Schema approach is using XML Query to retrieve the needed parts from EHR extract. The structural information is then not stored in the document that is parsed, as done with the extended EHR extract but has to be retrieved from the AT while parsing.

5. Conclusion

We implemented a system based on XML, XML Schema, XSLT and DOM to publish prEN 13606 compliant EHR extracts that conform to existing ATs. A method is presented to describe the structure of EHR extracts using XML Schema. The next step is to implement the import-layer for prEN 13606 compliant EHR extracts.

6. References

[1] CEN/TC 251 WG1 prEN13606. Health informatics - Electronic health record communication. 2006 [cited 2006 October]; Available from: http://www.centc251.org/WGI/WGIdoclist.htm

[2] Beale T, Archetypes and the EHR, in: Blobel B, Pharow P, eds., Advanced Health Telematics and Telemedicine, IOS Press 2003; pp. 238-44.

[3] Beale T. Archetypes: Constraint-based Domain Models for Future-proof Information Systems. 2002 [cited 2006 October]; Available from: http://www.deepthought.com.au/it/archetypes/archetypes_new.pdf

[4] Dorda W, Wrba T, Duftschmid G, Sachs P, Gall W, Rehnelt C, et al. ArchiMed - A Medical Information- and Retrieval System. Methods Inf Med. 1999;38(1):16-24.

[5] Friedman C, Hripcsak G, Johnson S, Cimino J, Clayton P. A generalized relational schema for an integrated clinical patient database. In: Miller RA, editor. Proceedings of the Annual Symposium on Computer Application in Medical Care; 1990 November 1990: Washington, DC: IEEE Computer Society Press; 1990. p. 335-9.

[6] Maldonado Segura JA. Historia Clinica Electronica Federada Basada en la Norma Europea CEN/TC251 EN13606: PhD Thesis, Universidad Politecnica de Valencia; 2005.

[7] DSTC - Flinders University et al. Final Report: The OACIS-GEHR Transformation Process; 2001.

[8] Altova GmbH. Mapforce. 2006.

[9] Tun Z, Bird LJ, Goodchild A. Validating Electronic Health Records Using Archetypes and XML. CRC for Enterprise Distributed Systems: University of Queensland; 2002.

[10] Prochazkova D. Architekturen der elektronischen, lebensbegleitenden Gesundheitsakte: Ein Vergleich der Archetypansätze von CEN und HL7 am Beispiel des Konzeptes "Diabetes": Master Thesis, Medical University of Vienna; 2006.