CEN prEN 13606 compliant export of medical data from an Entity-Attribute-Value based patient record system

Christoph Rinner, Thomas Wrba, Georg Duftschmid

Institute of Medical Information- and Retrieval Systems, Core Unit for Medical Statistics and Informatics, Medical University of Vienna, Austria

Abstract

This work describes an implementation of a proof-of-concept layer to export medical data from a health information system in a CEN prEN 13606 compliant format. The resulting EHR extracts are deposited in a local repository. The export-layer is based on Oracle® PL/SQL and the PL/SQL DOM API. EHR extracts are described by means of two XML Schemas and are transformed into Archetype compliant EHR extracts using XSLT. The XSLT needed for the transformation can be created visually using the two XML Schemas. A method is presented that solves the "unique particle attribution" problem, which would otherwise preclude the description of an EHR extract by means of XML Schema.

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 to exchange EHR segments in a semantically interoperable manner. It is based on the two-level modelling methodology which separates knowledge from information. In part one of the standard the Reference Model is described. Instances of the Reference Model can represent all medical information that may be contained within an EHR, independently of the originating system. In part two the concept of Archetypes [2, 3] and a language for specifying Archetypes is defined. Archetypes contain the knowledge, in other words they act as "construction plans" for instances of the Reference Model. By using the Reference Model and Archetypes an important step is made towards exchanging information in a semantically interoperable manner.

ArchiMed [4] is a clinical trial system developed at the Medical University of Vienna's Core Unit for Medical Statistics and Informatics. It allows to input and to statistically analyse data and further supports patient recruiting and the interactive design of case report forms (CRFs). ArchiMed uses the generic Entity-Attribute-Value (EAV) database design [5] often used with medical data.

Three elementary types of communicating health information can be distinguished. The *addressed communication* where the message is sent to a known receiver, the *directed communication* where the message can be retrieved by someone who is authorized by his role to do so, and the *non-directed communication* where the message is sent to a repository and can be retrieved by authorized individuals when needed [6]. An example of the last communication type is the IHE Cross-Enterprise Document Sharing (XDS) technical framework.

We applied the pre standard prEN 13606 and implemented a proof-of-concept export layer to export data stored in ArchiMed to a document repository as archetyped EHR extracts (see figure 1).

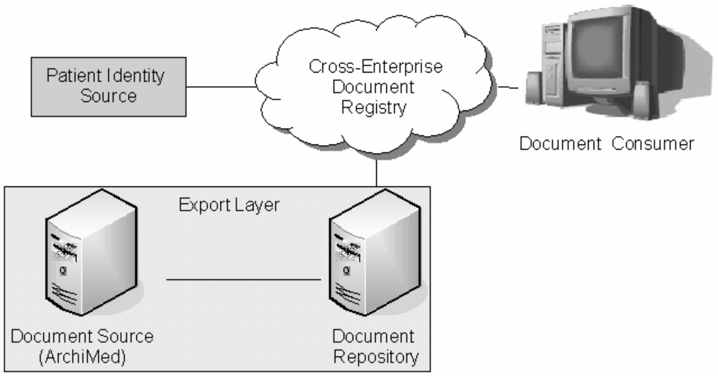

Figure 1: Cross-Enterprise Document Sharing Diagram

In this paper the term *EHR extract* refers to an XML document compliant to the prEN 13606 Reference Model but not conforming to an Archetype, whereas *archetyped EHR extract* refers to an EHR extract that additionally conforms to an existing Archetype.

Design Objectives

The objectives of this project were to explore the practical applicability of pre standard prEN 13606 and its concepts for exchanging semantically interoperable information. Only few references to prEN 13606 implementations can be found in literature. Pangea [7] is a middleware based system to export hospital data. It uses an individual schema language to import and export medical data. In contrast to them, we decided to base our publishing mechanisms on standard XML concepts such as XML Schema and the Extensible Stylesheet Language Transformations (XSLT) to allow the reuse of existing tools and knowledge gained in the last couple of years using XML. In [8] a service for out-of-hospital follow-up and monitoring of patients with chronic heart disease is described. The data are stored and retrieved compliant to prEN 13606. Our goal is to retrieve data that are not already stored in a prEN 13606 compliant system. We use existing Archetypes, developed independently of the current project to simulate a scenario where data were stored in a health information system previously and have to be published compliant to a given Archetype.

System Description

We divide the publishing of archetyped EHR extracts into two steps as depicted in figure 2.

In the first step we publish the relevant relational data from ArchiMed as XML documents compliant with the prEN 13606 Reference Model (EHR extracts). In the second step, the EHR extracts are transformed into archetyped EHR extracts. By conforming to an Archetype the EHR extract becomes semantically interoperable and can easily be shared between different health information systems.

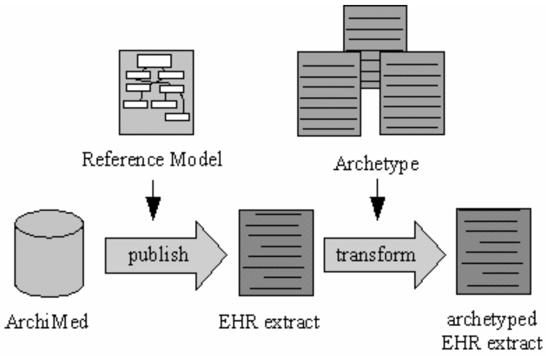

Figure 2: Two level publishing of archetyped EHR extracts

Publication of Relational Data as EHR extracts

We first extract the medical data stored in the database as EHR extract similar to the method used in the OACIS project [9]. To be able to generically publish the content of every CRF in ArchiMed as an XML document we mapped the structural concepts used within the CRFs to the classes of the Reference Model according to the schema in table 1. Using this mapping we implemented a generic PL/SQL procedure [10] in ArchiMed's Oracle® database engine to retrieve the data originating from any CRF. The CRF defines the structure of the EHR extract the same way as an Archetype is describing the structure of an archetyped EHR extract. The procedure uses the Structured Query Language (SQL) and the PL/SQL DOM API [11] in combination with the information from the Reference Model. The Entity-Attribute-Value design of the database simplified the implementation of a generic export since all data values are stored in a single table. Each document originates from a single CRF. By passing the document ID to the PL/SQL procedure the structure of the CRF is retrieved. For every structure of the CRF the corresponding values of the document are retrieved, converted into instances of the Reference Model and added to the EHR extract. The completed EHR extract is stored in a document repository.

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

ArchiMed	Reference Model
CRFs	COMPOSITIONs
Pages	SECTIONs
Tables	CLUSTERs
Entries	ELEMENTs

Instead of using a generic XML format we immediately create EHR extracts from the data retrieved from the database. Creating instances of the Reference Model ensures that the significance and context of the data are conserved and can be interpreted correctly during the transformation process. Since the structure of the EHR extract is based on the same Reference Model as the archetyped EHR extract, the transformation is facilitated. Additionally, mappings from the Reference Model to other standards (e.g. HL7, openEHR) exist and allow a reuse of the export-layer.

To be able to trace back an EHR extract and single RECORD_COMPONENTs to their origin in the database we assign an object identifier (OID) to every table of the ArchiMed Entity-Attribute-Value 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 Record Component's ID.

Transformation of EHR extracts into archetyped EHR extracts

Once the data are published from the relational database as EHR extract, the EHR extract has to be transformed into an EHR extract that complies with a previously agreed Archetype.

In the "XML world" this means that an XML document complying with XML Schema A has to be transformed into another XML document complying with XML Schema B. The Extensible Stylesheet Language Transformation (XSLT) recommendation by the World Wide Web Consortium (W3C) was developed to realize such transformations.

For every Archetype a separate XSLT script has to be created. Since manual creation of the XSLT scripts is highly laborious we use a mapping tool [12]. It allows the mapping of XML documents by visually associating the components of their underlying XML Schemas (see figure 3).

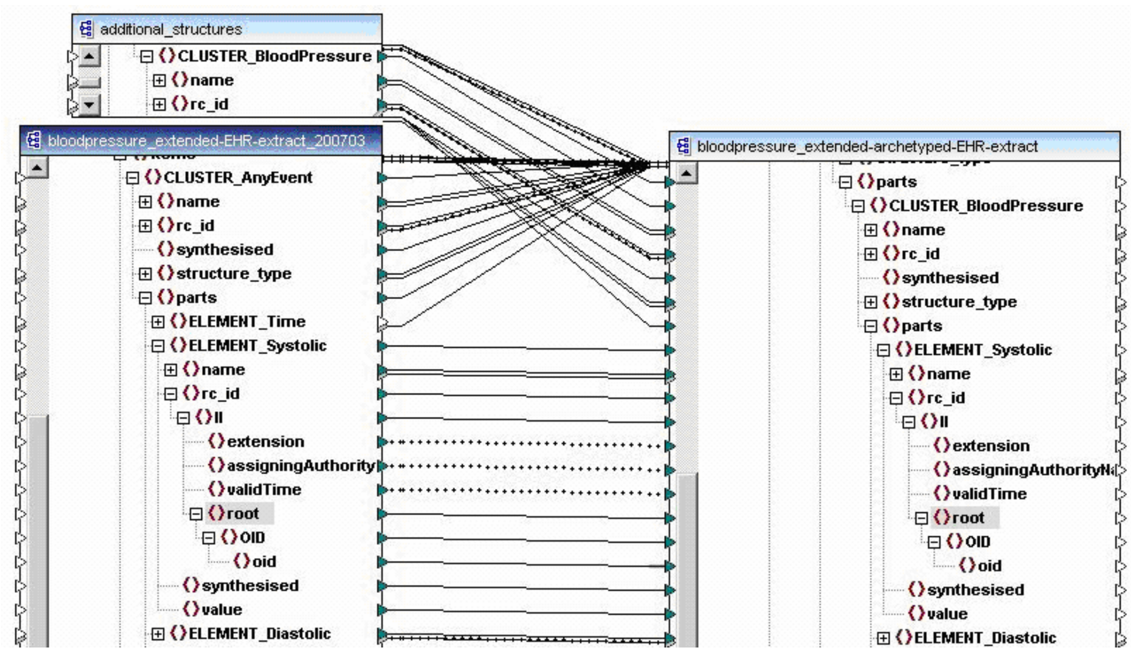

Figure 3: Visual XSLT generation using a commercial software tool

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 [13]: Two XML elements at the same hierarchy level within an XML Schema cannot be specified to have the same tag name. This constellation is, however, frequently implied by an Archetype's structure (e.g. two XML elements at the same level with 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 Archetype. These classes are virtual in the sense that their only purpose is to allow an XML Schema representation of an Archetype. 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 Reference Model classes.

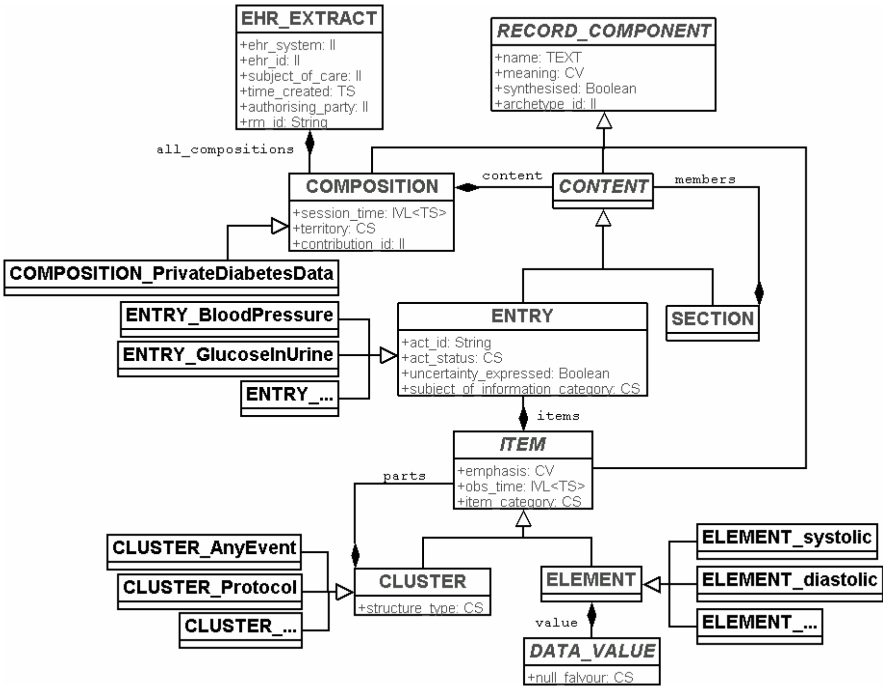

Figure 4: Extended Reference Model

The virtual subclasses are named after their superclasses concatenated with the name defined in the Archetype in camel-back notation. Naturally the subclasses inherit all the attributes and relations. An example of this “extended Reference Model” as we call it can be seen in figure 4. The original Reference Model classes are shown in gray, the classes in black are the virtual subclasses. The extended Reference Model also contains virtual subclasses for each component of the published CRF to allow pinpointing the sources for the following mapping process.

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

By referring to the virtual subclasses of the extended Reference Model 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 2.

Both XML Schemas are created manually and contain all the obligatory attributes from the Reference Model (such as *rc_id*, *name*, etc.) additionally to the ones specified either in the Archetype or the CRF.

In figure 3 these two XML Schemas are mapped visually. The output of this mapping is the needed XSLT that is applied to every extended EHR extract that is created by the PL/SQL procedure. By renaming the tags corresponding to virtual subclasses in the resulting extended archetyped EHR extract to the names of their superclasses (e.g. “ELEMENT_systolic” is renamed to ”ELEMENT”), the final archetyped EHR extract is created.

In cases where the Archetype specifies additional structures (e.g. CLUSTER_BloodPressure to further structure the EHR extract) they can be added during the mapping process.

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

Discussion

The developed method was tested using an existing SECTION Archetype called “private diabetes data” [14] consisting of six ENTRYs (blood pressure, laboratory-glucose, heart rate-pulse, body weight, dimensions and physical activity) modelled in the Archetype Description Language (ADL) [1].

The described virtual extension of the Reference Model allows us to use XML Schemas to describe the structure of EHR extracts. Relying on approved XML concepts we can take advantage of the experience gained with XML technologies and existing XML tools can be applied. A shortcoming of XML Schemas is that they are not capable to constrain values within EHR extracts to the same extent as Archetypes. As an example, constraints that involve more than one element (e.g., systolic pressure should be ≥ diastolic pressure) can be represented by means of the ASSERTION class within an Archetype but cannot be expressed within an XML Schema.

The XSLT script could also be created manually yet the visual approach using XML Schema and the extended Reference Model is easier and does not require XSLT programming skills. The XML Schema combines the information from the Reference Model and the knowledge from the Archetype in one single document and facilitates the interpretation of the EHR extract content for humans.

Conclusion

We implemented a system based on XML, XML Schema, XSLT and DOM to publish prEN 13606 compliant EHR extracts that conform to existing Archetypes. 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.

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. 238-244.

[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): pp. 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, ed. Proceedings of the Annual Symposium on Computer Application in Medical Care; November 1990: Washington, DC: IEEE Computer Society Press; 1990; pp. 335-339.

[6] Haas P. Gesundheitstelematik, Springer; 2006.

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

[8] Munoz A et al. Proof-of-Concept Design and Development of an EN13606-based Electronic Healthcare Record Service. J Am Med Inform Assoc; 2007; 14(1): pp. 118-129.

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

[10] Russell J. Oracle9 i, PL/SQL, User’s Guide and Reference, Release 2 (9.2), Oracle Corporation; 2002.

[11] Leyderman R. Oracle9 i, XML API Reference – XDK and Oracle XML DB, Release 2 (9.2), Oracle Corporation; 2002.

[12] Altova GmbH, Mapforce 2007 Enterprise Edition, free 30-day trial; 2006.

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

[14] 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.