

VALIDIEREN VON AUF ZWEIMODELL-ANSÄTZEN BASIERENDEN, VOLLSTRUKTURIERTEN EHR-DATEN AM BEISPIEL EN/ISO13606 UND HL7 CDA

Rinner C¹, Duftschmid G¹

Kurzfassung

In dieser Arbeit wird ein Ansatz zur Validierung von HL7 CDA bzw. EN/ISO13606 formatierten EHR-Daten gegenüber in Archetypen, Templates oder Implementierungsleitfäden festgelegten Strukturvorgaben präsentiert. Zum Zweck der mittels XML-Schema und XSLT realisierten Validierung werden die den EHR-Daten zugrunde liegenden getrennten Modelle für Information und Wissen temporär zu einem Gesamtmodell zusammengeführt. Die Validierung wurde mit drei EN/ISO13606 Archetypen und einem CDA-Implementierungsleitfaden für Diabetes-Therapiedaten getestet.

1. Einleitung

Im Regierungsprogramm 2008 [1] wird der Weiterentwicklung der Informations- und Kommunikationstechnologien (e-health) im Gesundheitswesen und speziell der elektronischen Gesundheitsakte (ELGA) eine „hohe gesundheitspolitische Priorität“ beigemessen. Die im Rahmen der Machbarkeitsstudie für ELGA 2006 [2] analysierten Informations-Architekturen, namentlich EN/ISO13606 [3] und HL7 v3 CDA R2 [4] (CDA), beruhen alle auf dem Zweimodell-Ansatz, der Information und Wissen trennt. Beide Standards bauen auf statischen Referenzmodellen auf. Ausgehend von der ELGA-Machbarkeitsstudie fokussiert diese Arbeit auf die Standards EN/ISO13606 und HL7 v3 CDA R2, auf ähnliche Konzepte von z.B. openEHR² wird im Rahmen dieses Artikels nicht eingegangen.

EHR-Extrakte (EN/ISO13606) und CDA-Dokumente (HL7) stellen Instanzen dieser Referenzmodelle dar und werden im Folgenden unter dem Oberbegriff „EHR-Daten“ subsummiert. Repräsentiert werden EHR-Daten in der Regel als XML-Dokumente³, die mittels Implementierungsleitfäden und/oder Templates (HL7) bzw. Archetypen (EN/ISO13606) semantisch beschrieben werden können.

Werden EHR-Daten kommuniziert, so empfiehlt sich empfängerseitig eine Validierung, d.h. eine Überprüfung der erhaltenen EHR-Daten auf Konformanz mit dem Referenzmodell und den vereinbarten semantischen Beschreibungen, da nur so deren fehlerfreie Integration in das EHR-System

¹ Institut für Medizinische Informations- und Auswertesysteme, Medizinische Universität Wien

² <http://www.openehr.org/>

³ <http://www.w3.org/XML/>

des Empfängers gesichert ist. Für die Validierung auf Konformität zu in Archetypen oder Templates spezifizierten semantischen Beschreibungen standen bis jetzt nur Schematron-Skripts¹ bei der CDA und eine Java API² bei der EN/ISO13606 zur Verfügung. Die Java API befindet sich noch unter Entwicklung und stellt nur einen 1. Schritt für eine Validierung dar. Für die Validierung von CDA-Dokumenten gegenüber dem RIM gibt es seit kurzem ein auf Eclipse basierendes CDA-Validierungstool³. Derzeit erlaubt dieses Tool jedoch noch keine Validierung von semantischen Beschreibungen. Weiters wurden existierende Schematron Scripts in Webservices verwendet (z.B. CDA Validator von Alschuler⁴ und NIST⁵), um CDA-Dokumente gegenüber den Implementierungsleitfäden zu validieren.

Einen anderen, naheliegenden Ansatz stellt die Verwendung von XML-Schemas zur Validierung der EHR-Daten dar. Auf Basis bestehender W3C-XML-Schemas⁶ (siehe www.linkehr.org für EN/ISO13606 bzw. www.hl7.org für CDA) lassen sich EHR-Daten auf Übereinstimmung mit den zugrunde liegenden Referenzmodellen überprüfen. Eine direkte Transformation der semantischen Beschreibungen in ein XML-Schema ist jedoch aufgrund der in [5] beschriebenen *Unique Particle Attribution Rule* nicht möglich.

In diesem Beitrag präsentieren wir eine Methode, die es erlaubt, die *Unique Particle Attribution Rule* zu umgehen und damit EHR-Daten mittels XML-Schemas auf Einhaltung semantischer Beschreibungen zu überprüfen. Gegenüber der im Falle von Implementierungsleitfäden tlw. für diesen Zweck eingesetzten Schematron-Skripts [6] hat dies den Vorteil, dass

- man auf ein breites Spektrum von bestehenden XML Tools zurückgreifen kann, die neben der Validierung auch Unterstützung in vielen anderen Bereichen, z.B. dem grafischen Editieren, bieten,
- keine neue Sprache erlernt werden muss, da EHR-Daten als XML-Dokumente vorliegen,
- die endgültige Struktur der EHR-Daten leicht erkennbar ist,
- die angesprochenen XML-Schemas auch für ein visuelles Mapping zum Zweck einer EHR-Daten Transformation verwendet werden können [7, 8].

2. Methoden

2.1. Erstellung des erweiterten XML-Schemas

Um die Komplexität des Validierungsprozesses zu reduzieren, führen wir die im Zweimodell-Ansatz getrennten Ebenen des Referenzmodells und der semantischen Beschreibungen von Archetypen/Templates/Implementierungsleitfäden in ein gemeinsames Modell über. Dazu verbinden wir die einzuhaltenden semantischen Beschreibungen und die bestehende XML-Schema-Repräsentation des Referenzmodells zu einem sogenannten „erweiterten XML-Schema“ (siehe *Abbildung 1*).

¹ <http://www.schematron.com/>

² <http://www.openehr.org> und <http://www.openehr.org/wiki/display/dev/Data+Validation>

³ http://hl7book.net/index.php?title=Eclipse_Instance_Editor

⁴ <http://www.alschulerassociates.com/validator/>

⁵ <http://xreg2.nist.gov/cda-validation/validation.html>

⁶ <http://www.w3.org/XML/Schema#dev>

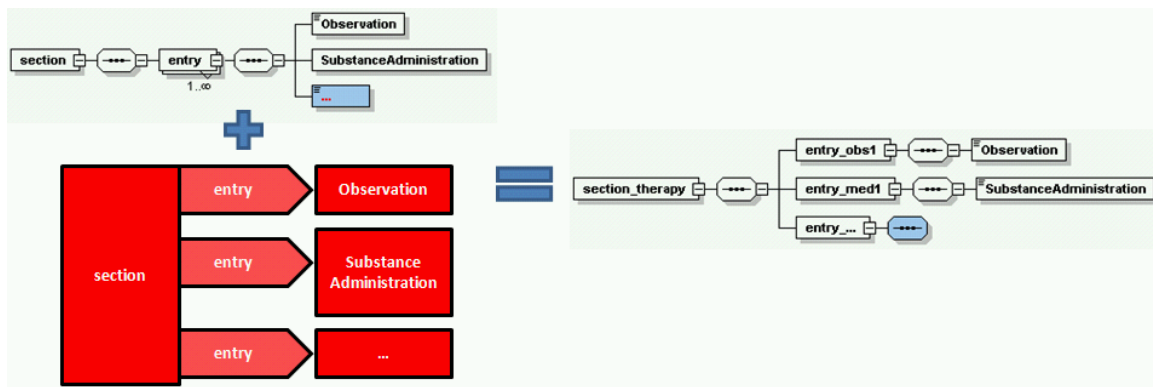


Abbildung 1: Referenzmodell als XML-Schema und CDA-Einschränkung laut Implementierungsleitfaden werden in ein erweitertes XML-Schema übergeführt

Die semantischen Beschreibungen können dabei unter anderem vorschreiben, dass Instanzen derselben Klasse des Referenzmodells mit unterschiedlicher Bedeutung und Substruktur auf gleicher hierarchischer Ebene in den EHR-Daten vorkommen müssen. In einem XML-Schema sind derartige Strukturen, bei denen zwei Elemente auf gleicher hierarchischer Ebene den gleichen Namen bei unterschiedlicher Substruktur haben, aufgrund der *Unique Particle Attribution Rule* jedoch nicht erlaubt. Wir umgehen dieses Problem, indem wir bei der Erstellung des XML-Schemas „virtuelle Subklassen“ des Referenzmodells einführen, um die wir Letzteres für den Zweck der Validierung fiktiv erweitern. Die virtuellen Subklassen manifestieren sich als Schema-Elemente, deren Namen sich aus der jeweils zu instanzierenden Klasse des Referenzmodells und der inhaltlichen Bedeutung der Instanz laut semantischer Beschreibung zusammensetzen. EHR-Daten, die eine Instanz eines um virtuelle Subklassen erweiterten XML-Schemas darstellen, bezeichnen wir im Folgenden als „erweiterte EHR-Daten“.

Da bei EN/ISO13606 alle relevanten Klassen (*COMPOSITION*, *ENTRY*, *CLUSTER*, *ELEMENT*) von der Wurzelklasse *RECORD_COMPONENT* abgeleitet werden, kann der Name der Schema-Elemente aus den laut semantischer Beschreibung zu instanzierenden Klassen und den darin vorgegebenen Werten des obligatorischen Attributs *name* zusammengesetzt werden. Schwerer gestaltet es sich bei der CDA. Zwar gibt es die Klasse *InfrastructureRoot* mit dem Attribut *templateId*, der alle relevanten Klassen zugrunde liegen, jedoch gibt es keine obligatorische Identifikation, die bei allen Klassen ident ist und unabhängig des Templates/Implementierungsleitfadens für eine Namenserverweiterung herangezogen werden könnte. Die Namenserverweiterung sollte hier wie in Kapitel 2.2 dargestellt so gewählt werden, dass sie auf EHR-Daten generisch angewandt werden kann.

2. 2. Die Validierung

Um EHR-Daten mit dem erweiterten XML-Schema zu validieren, werden diese in erweiterte EHR-Daten transformiert. Dazu verwenden wir ein generisches XSLT-Script¹, das in den EHR-Daten eine analoge Namenserverweiterung vornimmt, wie bei der Erstellung des erweiterten XML-Schemas in Kapitel 2.1 beschrieben wurde. Pro Anwendungsfall (z.B. Archetyp, Implementierungsleitfaden) muss ein erweitertes XML-Schema erzeugt werden, das die zu überprüfenden Einschränkungen enthält (siehe *Abbildung 2*).

¹ <http://www.w3.org/TR/xslt20/>

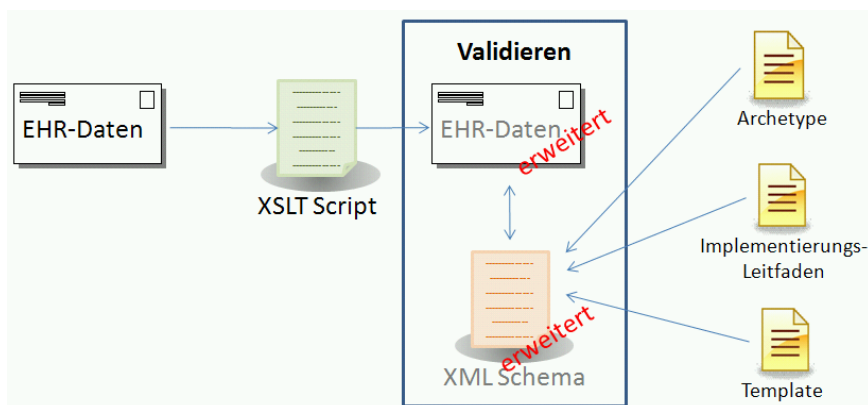


Abbildung 2: Validierungsprozess

Mittels des erweiterten XML-Schemas können so detaillierte strukturelle Einschränkungen wie zum Beispiel „Der Entry Blutdruckmessung muss ein Element mit dem systolischen Wert, gefolgt von einem Element mit dem diastolischen Wert enthalten“ überprüft werden. Weiters können mittels XML-Schema reguläre Ausdrücke (Regular Expressions), Ausprägungslisten und numerische Wertebereiche festgelegt werden, um z.B. Einschränkungen wie „Der systolische Wert sollte zwischen 80 und 140 liegen und die Einheit muss in *mmHg* angegeben sein“ zu validieren.

3. Ergebnisse

Die Validierungsmethode wurde mittels dreier EN/ISO13606-Archypen („dimension“, „heart rate“, „body weight“) und einem CDA-Implementierungsleitfaden für Diabetes-Therapiedaten getestet.

Die benötigten erweiterten XML-Schemas wurden manuell erzeugt. Der Zeitaufwand für ihre Erstellung betrug in Abhängigkeit von der Größe des Archypen bzw. Implementierungsleitfadens zwischen 4 und 8 Stunden.

Das XSLT-Script für die Transformation von EHR-Extrakten in erweiterte EHR-Extrakte besteht bei EN/ISO13606 aus 10 Zeilen Code. Da die Namenserverweiterung bei CDA-Dokumenten abhängig von den jeweiligen Templates bzw. Implementierungsleitfäden ist, benötigten wir ein umfangreicheres XSLT-Script (ca. 200 Zeilen Code). Das erzeugte XSLT-Script kann mit leichten Adaptierungen für weitere Validierungen verwendet werden. Die Granularität der Validierung kann durch Adaptieren des XSLT-Scripts und Editieren des erweiterten XML-Schemas reguliert werden. So kann von *Sections* bis zu den einzelnen Attributen alles validiert werden.

4. Diskussion

XML-Schema unterstützt zwei Arten von Vererbung (*Extension* und *Restriction*), da diese jedoch für eine effiziente Nutzung beliebig kombinierbar sein müssten, werden im erweiterten XML-Schema alle Klassen geklont. Dies hat den bekannten Nachteil, dass Adaptierungen in der Schema-Repräsentation einer Referenzmodellklasse an mehreren Stellen nachgezogen werden müssen. Um diese Problematik zu entschärfen arbeiten wir an einer automatischen Generierung des erweiterten XML-Schemas aus Archypen. Adaptierungen in der Schema-Repräsentation einer Referenzmodellklasse würden bei einer automatischen Generierung des Schemas im Algorithmus programmatisch realisiert und müssten nicht manuell an mehreren Stellen des Schemas nachgezogen werden.

Eine neue Version von XML-Schema (W3C XML-Schema 1.1) ist momentan im „Last Call Working Draft“. In der neuen Version soll es möglich sein, *Assertions* (z.B. systolischer Wert größer diastolischer Wert), wie in Schematron, zu überprüfen. Auch die *Unique Particle Attribution Rule* wird in der aktuellen Version des Drafts nicht mehr gefordert. Es ist daher künftig mit neuen Potentialen von XML-Schema für die Validierung von EHR-Daten gegenüber semantischen Beschreibungen zu rechnen.

5. Danksagung

Diese Arbeit wurde teilweise durch das OeNB-Projekt „Telediab: Ein telemedizinisches Tagebuch für Patienten mit Diabetes mellitus Typ 2“ finanziert.

6. Literaturangaben – Literaturhinweise

- [1] BUNDESKANZLERAMT ÖSTERREICH. Regierungsprogramm 2008-2013; 2008.
- [2] BUNDESMINISTERIUM FÜR GESUNDHEIT UND FRAUEN FAMILIE UND JUGEND. Machbarkeitsstudie ELGA. ; 2007.
- [3] EUROPEAN COMMITTEE FOR STANDARDIZATION, INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. EN/ISO 13606-1 Electronic healthcare record communication: Part 1: Reference model; 2006.
- [4] DOLIN RH, ALSCHULER L, BOYER S, BEEBE C, BEILEN FM, BIRON PV, et al. HL7 Clinical Document Architecture, Release 2. J Am Med Inform Assoc. 2006;13(1):30-9.
- [5] TUN Z, BIRD LJ, GOODCHILD A. Validating Electronic Health Records Using Archetypes and XML: CRC for Enterprise Distributed Systems: University of Queensland; 2002.
- [6] FERRANTI JM, MUSSER RC, KAWAMOTO K, HAMMOND WE. The clinical document architecture and the continuity of care record: a critical analysis. J Am Med Inform Assoc. 2006 May-Jun;13(3):245-52.
- [7] RINNER C, WRBA T, DUFTSCHMID G. Publishing relational medical data as prEN 13606 Archetype compliant EHR extracts using XML technologies. Tagungsband der eHealth 2007 – Medical Informatics meets eHealth. Wien 2007:35-9.
- [8] RINNER C, DUFTSCHMID G, WRBA T. CEN prEN 13606 konformer Export von medizinischen Daten aus einem Entity-Attribute-Value basierten Informationssystem Telemedizinführer Deutschland 2008 (Sonderkapitel "Best of Telemed Berlin"); 2008; Berlin; 2008. p. 34-8.