Validation of Standardized and Structured EHR Data in the Dual Model Approach

Christoph RINNER, Georg DUFTSCHMID
Core Unit for Medical Statistics and Informatics, Medical University of Vienna, Austria

Abstract. We present a W3C XML Schema-based method to validate standardized EHR data against semantic constraints that build the knowledge layer within the dual model approach. The approach was tested with three EN/ISO13606 archetypes and an HL7 CDA implementation guide for diabetes therapies.

Keywords. dual model approach, validation, electronic health record

Standardized electronic health record (EHR) data should be validated by its receiver before being integrating into an EHR system. The currently most cited EHR standards EN/ISO13606 and HL7 CDA are based on the dual model approach which separates information instantiated from a reference model (RM) from knowledge given by Archetypes (ATs), Templates (TPs) or Implementation Guides (IGs). Currently no tools exist to validate EHR data for conformance with ATs or TPs. Some IGs were represented as Schematron scripts for this purpose. W3C XML Schema would provide a better manageable format as it is well supported by off-the-shelf tools and is a more common schema language than Schematron. Due to XML Schema’s “unique particle attribution rule”, however, ATs, TGs and IGs cannot be directly translated to XML Schemas [1].

For the purpose of validation we use a so called extended XML Schema that unites the separate layers of the dual model approach by introducing virtual subclasses of RM classes [2]. EHR data is transformed into extended EHR data, validated against the extended XML Schema and finally retransformed into HER data. This transformation (adding the subclasses) is solved generically using simple XSLT Scripts and specific attributes within the EHR data (i.e., ENTRY becomes ENTRY_bloodpressure and vice versa). We tested the method with three EN/ISO13606 ATs and a CDA IG for diabetes therapies. Up to now we created the extended XML Schemas manually which took approximately four to eight hours per schema. Currently we are working on a method to automatically derive them from Archetypes.
