Validating and Transforming
HL7 CDA Documents for i2b2 Import

Tingyan Xu1, Volker S. Thiemann2, Rainer Röhrig1 and Raphael W. MAJEED1

1) Department of Medical Informatics, Carl von Ossietzky University Oldenburg

Health Exploring Complexity 2016
28th August – 2nd September 2016 - Munich

The AKTIN Project
We are building a decentralized emergency care research IT-Infrastructure to ensure patient privacy and comply with the many and varied privacy laws in Germany. (Refer to [1] for more details.)

Structure
Each participating hospital will store their own data in a data warehouse located on its own premises. The data originates from the emergency room protocol (ER note) as designed in [2]. The protocol is documented in HL7 v3 Clinical Document Architecture (CDA). The data warehouse is based on the open source clinical data warehouse system i2b2. Customized configurations and additional tools are installed to validate and import the ER-CDA documents, organize the centralized research queries and manage the answers from each participating hospital.

Transmission and Validation
We offer two standard interfaces for transmitting the CDA documents: Via the IHE XDS.b document repository for SOAP API calls or with HL7 FHIR binary endpoint for RESTful data transfers. All incoming documents have to be validated before further transformation and usage. The automatically generated Schematron rules from the CDA-ER specification (available at http://aktin.art-decor.org/) are used for this purpose, ensuring only valid documents of this one specification are loaded into the DWH.

Transformation
After validation the complex CDA XML structure of the incoming documents have to be broken down into the key-value (also Entity-Attribute-Value, EAV) format to fit the i2b2's database structure. Although not a new function, we need a configurable function in this case, designed and implemented for the AKTIN environment, to transform from ER-CDA to i2b2: Some facts in our case may be empty, others may not be necessary in the researching context and can be removed. Documents can be retransmitted to update the data in i2b2.

Results and discussion
Although there are other ETL tools and methods to load data from CDA into a DWH [3], our approach is highly adaptive and we provide a specified interface and software to handle the CDA documents and manage the i2b2 DWH.

References
1. Abbrardt et al. Balancing the need for big data and patient data privacy—an IT infrastructure for a decentralized emergency care research database. PMID: 25160287.
2. Kulla M et al. Nationaler Datensatz "Notaufnahme": Entwicklung, Struktur und Konsentierung durch die Deutsche Interdisziplinäre Vereinigung für Intensivmedizin und Notfallmedizin. DOI: 10.1007/s00330-014-2295-x.