



CENTER FOR MEDICAL DATA SCIENCE  
MEDICAL UNIVERSITY OF VIENNA  
Institute of Medical Information Management

Topic for: Master's thesis (potentially also KfK-Praktikum, KfK-Seminar)

Contact: Florian Katsch, PhD [florian.katsch@meduniwien.ac.at](mailto:florian.katsch@meduniwien.ac.at)

## Anonymization of national EHR CDA documents

---

Electronic Health Record (EHR) documents, as used by ELGA, are mainly based on the HL7 Clinical Document Architecture (CDA). Those documents contain sensitive patient information that must be protected when data is used for methodological research, system development and testing, or data exchange. The objective of this topic is to investigate methods for identifying personally identifiable information (PII) within CDA-based EHR documents and to analyze how HL7 template definitions can support automated detection of sensitive data elements. Based on this analysis, a systematic anonymization approach shall be designed that leverages CDA structure and template constraints to locate and classify PII. The proposed method will then be implemented as a prototype capable of processing ELGA CDA documents and removing or redacting sensitive information while preserving the structural integrity of those documents. The prototype is then evaluated regarding correctness, completeness, and usability.

We offer a friendly and supportive research group and are interested in a timely completion of the thesis. Regular in-person or online meetups can be arranged, office equipment can be provided (but there is no need for fancy hardware).

We look for motivated, pro-active students, who are interested in medical information management and or health data standardization. Proficiency in Python and git is a must, ideally you are using a Unix-like operating system. Sorry, no Java!