

Christian Dominik Kohl
Dr. sc. hum.

Patientenübergreifende, multiple Verwendung von Patientendaten für die klinische Forschung unter Nutzung von Archetypen

Promotionsfach: Medizinische Biometrie u. Informatik
Doktormutter: Prof. Dr. Petra Knaup-Gregori

Sowohl in der Routineversorgung als auch in klinischen Studien werden immer mehr Daten elektronisch verarbeitet. Trotzdem ist ein Austausch von Daten zwischen beiden Bereichen häufig noch nicht etabliert. Dies führt dazu, dass Daten mehrfach erfasst werden müssen. Die redundante Datenerfassung ist zeitaufwändig und kann zu Inkonsistenzen zwischen Krankenhausinformationssystem (KIS) und Studiendatenmanagementsystem (SDMS) führen. Obwohl ein Datenaustausch zwischen Forschung und Versorgung oft technisch möglich wäre, scheitert er meist noch an mangelnder semantischer Interoperabilität.

Archetypen sind ein innovatives Konzept zur Gestaltung von flexiblen und leicht erweiterbaren elektronischen Gesundheitsakten. Sie ermöglichen semantische Interoperabilität zwischen Systemen, welche dieselben Archetypen nutzen. Das Archetypen-Konzept hat mittlerweile auch Eingang in internationale Standards gefunden (ISO 13606). Die *openEHR*-Spezifikationen definieren ein mit ISO 13606 kompatibles jedoch weitergehendes Modell für elektronische Gesundheitsakten. Bisher wurden Archetypen hauptsächlich für Informationssysteme in der Routineversorgung und weniger für die klinische Forschung entwickelt und genutzt. Ziel dieser Arbeit war es daher, basierend auf den *openEHR*-Spezifikationen und Archetypen generische Ansätze zu erarbeiten, die eine multiple Verwendung von Daten aus der Versorgung in der Forschung ermöglichen und deren Umsetzbarkeit zu prüfen.

In einer Voruntersuchung wurde ermittelt, dass 35 % der in der betrachteten Studie zu erhebenden Merkmalsarten aus dem untersuchten KIS übernommen werden könnten, wenn die Daten dort elektronisch und ausreichend strukturiert vorlägen. In einem zweiten Schritt wurde mit *openSDMS* der Prototyp eines auf Archetypen basierenden integrierten elektronischen Gesundheitsakten- und Studiendatenmanagementsystems zur Verfügung gestellt. Aus der Voruntersuchung und der Implementierung von *openSDMS* wurden Anforderungen abgeleitet und eine auf *openEHR*-Archetypen basierende Referenzarchitektur entwickelt, welche die Nutzung von Daten aus KIS in klinischen Studien unterstützt. Dabei wird sowohl die Integration von KIS beschrieben, die auf Archetypen basieren, als auch von klassischen KIS. Kernkomponenten dieser Architektur sind auf Archetypen basierende semantische Annotationen von Studiendaten sowie Import- und Exportmodule, welche die Archetype Query Language nutzen. Die vorgestellte Referenzarchitektur ermöglicht den Übergang von der multiplen Erfassung hin zur multiplen Verwendung von Daten in Forschung und Versorgung.

Um die entwickelte Referenzarchitektur realisieren zu können, werden geeignete Archetypen auch für Forschungsdaten benötigt. Daher wurden Archetypen zur Dokumentation aller Datenelemente der vier CDASH Domänen ‚Common Identifier Variables‘, ‚Common Timing Variables‘, ‚Adverse Events‘ sowie ‚Prior and Concomitant Medications‘ spezifiziert (Studiendaten). Hierzu wurden insgesamt 23 Merkmalsarten basierend auf Archetypen neu definiert, wozu drei bestehende Archetypen spezialisiert und zwei neu entwickelt wurden. Zur Definition von CDASH-konformen elektronischen Datenerhebungsbogen für die betrachteten Domänen wurden, basierend auf den spezifizierten Archetypen, vier *openEHR*-Templates entworfen. Ferner wurden 71 Merkmalsarten in 16 Archetypen zur Dokumentation von Studien-Metadaten definiert. Alle neu entworfenen Archetypen wurden jeweils in englischer

und deutscher Sprache beschrieben und können nun als Referenzinformationsmodell für Forschungsdaten genutzt werden. Ergänzend wurden alle von den bereitgestellten Archetypen definierten Merkmalsarten auf die im Bereich der klinischen Forschung etablierten Modelle BRIDG, CDASH und ODM abgebildet.

More and more data is collected electronically in patient care and in clinical trials. However, the exchange of data between these areas is not well-established. This often causes a multiple collection of the same data. Redundant data collection is time consuming and may lead to inconsistencies between hospital information system (HIS) and clinical trial data management system. Although the exchange of data between research and patient care is often possible on a technical level, it is hard to achieve semantic interoperability.

Archetypes are an innovative approach to build flexible and easily adaptable electronic health records. Archetypes facilitate semantic interoperability among all systems which use the same archetypes. The archetype approach strongly influenced the international ISO 13606 standard on electronic health record exchange. The openEHR-specifications define a model for electronic health records which is compatible with ISO 13606 but more comprehensive. Up to now, archetypes have been developed and used mainly for regular clinical care but not for clinical research. The aim of this work was to develop generic approaches based on the *openEHR*-specifications and archetypes which facilitate a multiple use of data collected in patient care for research and to verify the feasibility of this approach.

In a preparatory study it was found that 35 % of the data items required in the analyzed trial could be reused from the HIS if all data were highly structured and stored electronically. In a second step openSDMS as a prototype of a combined electronic health record and clinical trial data management system has been implemented. Based on the preparatory study and the implementation of openSDMS requirements have been analyzed and a reference architecture based on *openEHR*-archetypes has been developed. This architecture includes the integration of both, archetype based and conventional HIS. Core elements of this architecture are import and export modules which make use of the archetype query language and the semantic annotation of study data based on archetypes. This reference architecture facilitates the transition from a multiple collection to a multiple use of data in clinical research and care.

Archetypes for research data are required to be able to implement the proposed architecture. Therefore, archetypes have been developed to represent all data elements of the CDASH domains 'Common Identifier Variables', 'Common Timing Variables', 'Adverse Events' and 'Prior and Concomitant Medications'. For 23 attributes for which no archetype based representation existed, two archetypes had to be newly defined and three had to be specialized. An *openEHR*-template has been generated for each of the four investigated CDASH domains to be able to define CDASH-compliant electronic case report forms based on archetypes. To represent study meta data 71 attributes in 16 archetypes have been defined. All newly built archetypes have been provided in English and German and may be used as reference information model for clinical research data. In addition, all attributes defined by these archetypes have been mapped to BRIDG, CDASH and ODM as established models in clinical research.