

# Entwicklung von ETL-Verfahren für die interoperable Nutzung von Gesundheitsdaten – Grundprinzipien und Lessons learned

E Thomas<sup>a,1</sup>, HM Kruse<sup>a</sup>, C Schubert<sup>a</sup>, M Hoffmann<sup>a</sup>, K Saleh<sup>a</sup>, D Ammon<sup>a</sup>

<sup>a</sup> Datenintegrationszentrum, Universitätsklinikum Jena

## Einleitung

Die Verwendung klinischer Routinedaten in der Forschung ist das wesentliche Ziel der Medizininformatik-Initiative[1]. Hierzu wurden Konzepte und technische Lösungen entwickelt, mit denen Daten in standardisierte Formate übertragen werden können. Daraus entstanden Data-Engineering-Aufgaben[2], da der ordnungsgemäße Einsatz von Standards und der vollständige Umfang der medizinischen Daten in bereits bestehenden klinischen Systemen nicht garantiert werden konnte. Hieraus wurden einige Erkenntnisse gewonnen, welche bei zukünftigen Entwicklungen als grundlegende Vorgehensweisen berücksichtigt werden können.

## Methodik

Transformationen klinischer Routinedaten in standardisierte Datenformate sind eine Herausforderung, welche sich nicht nur durch ein Mapping verschiedener Attribute oder Datentypen lösen lässt. Die gewonnenen Daten sind sowohl für den Austausch zwischen Systemen als auch für übergreifende wissenschaftliche Auswertungen notwendig. Der Transformationsprozess muss sich bereits bei der Konzeption verschiedener ETL-Strecken an den häufig proprietären Datenformaten aus den medizinischen Primärsystemen orientieren. Der Einsatz offener Standards ermöglicht hingegen größere Freiheiten. Durch die Transformation kann zudem ein Umgang mit unvollständigen oder fehlerhaften Daten ermöglicht werden. Eingesetzte Werkzeuge für den Transformationsprozess müssen flexibel

---

<sup>1</sup> Korrespondenz-Autor, Eric Thomas, Universitätsklinikum Jena, Datenintegrationszentrum, Stoystraße 3, 07743 Jena; E-Mail: eric.thomas@med.uni-jena.de

genug sein, um die zur Verfügung stehenden Daten von einem gegebenen Datenformat in ein anderes zu übertragen und gegebenenfalls zu ergänzen. Dabei ist der korrekte Umgang mit Datentypen, Einheiten und standardabhängigen Konventionen wichtig, um Verfälschung der Daten zu vermeiden.

## Ergebnisse

Für die Entwicklung verschiedener ETL-Strecken am DIZ Jena konnten folgende Grundprinzipien herausgearbeitet und an unterschiedlichen Beispielen dargestellt werden:

- Aufeinander aufbauende ETL-Verfahren
- Modulare Ansätze bei Verarbeitung unterschiedlicher Datenquellen
- Sicherstellung der Datenerhaltung aus den Quellsystemen
- Toleranz gegenüber fehlenden und fehlerhaften Informationen
- Nachträgliche Korrektur und Ergänzung der Informationen ermöglichen
- Rückführung der Erkenntnisse in den klinischen Betrieb
- Informationsanreicherung durch verschiedene Informationsquellen

Insbesondere die Auseinandersetzung mit den klinischen Quellsystemen ist notwendig, um Kompetenzen im Umgang mit den Datenformaten auszubilden. Hierdurch wurden auch Kenntnisse, die denen eines technischen Data Stewards entsprechen[3], erworben.

## Diskussion

Ein wichtiges Ziel der Datentransformation sollte die Beibehaltung oder Verbesserung der Datenqualität aus dem klinischen Primärsystem sein. Durch die große Vielfalt an klinischen Primärdaten und unterschiedlichen Vorgehensweisen innerhalb der ETL-Strecken, muss für jede Transformation eine individuelle Transformationslösung erarbeitet werden. Für den Umgang mit Patientendaten ist hierbei jedoch die Beachtung grundlegender Vorgehensweisen für die Entwicklung notwendiger Transformationen, die sich aus den geschilderten Erkenntnissen ergeben, essenziell.

## Referenzen

[1] Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. *Methods Inf Med.* 2018 Jul;57(S 01):e50-e56.

[2] Reis J, Housley M. Fundamentals of Data Engineering. Sebastopol, CA: O'Reilly Media; 2022

[3] Katzensteiner M, Müller A, Witte ML, Schewe N, Bott O (Leitungen der GMDS-Arbeitsgruppen "Datenmanagement in klinischen und epidemiologischen Studien" und "Curricula der Medizinischen Informatik"): Data-Steward(-ship) in der Digitalisierung des Gesundheitswesens. mdi: Forum der Medizin\_Dokumentation und Medizin\_Informatik 2022(4):123–125.