

EIN VORGEHENSMODELL ZUR ERHÖHUNG DER DATENQUALITÄT KLINISCHER ROUTINEDATEN IM KONTEXT DER SEKUNDÄRNUTZUNG

Holzer K¹, Dorda W¹, Duftschmid G¹, Nachbagauer A¹,
Strasser N¹, Wrba T¹, Gall W¹

Kurzfassung

Inhalte von Electronic Health Records werden für die Forschung immer bedeutender. Klinische Routine und Forschung stellen wegen unterschiedlicher Datenqualitätsanforderungen schwer vereinbare Anwendungsgebiete dar. Es wird ein Vorgehensmodell präsentiert, das zur Erhöhung der Datenqualität für die Sekundärnutzung beitragen soll, um die Anforderungen beider Gebiete zu berücksichtigen. Dabei wird auf bestehende Modelle und Standards (z.B. Health Level 7 Clinical Research Functional Profile, Clinical Data Acquisition Standards Harmonization etc.) zurückgegriffen.

Abstract

Electronic Health Records gain importance for clinical research. Routine medical care and clinical research are two very different areas concerning their requirements on data quality and data structure. Within this paper a model of procedures is proposed, which should improve the quality of data regarding secondary use. Therefore, established standards and models (e.g. Health Level 7 Clinical Research Functional Profile, Clinical Data Acquisition Standards Harmonization etc.) are used to consider the requirements of both areas.

Keywords – *Electronic Health Record, Datenqualität, Sekundärnutzung, Vorgehensmodell*

1. Einleitung

Klinische Routine und Forschung sind Bereiche, die beide im klinischen Umfeld stattfinden, aber sehr unterschiedliche Anforderungen hinsichtlich Beschaffenheit und Struktur der verwendeten Daten aufweisen. Somit werden im Regelfall unterschiedliche IT-Systeme zur Datenerfassung und Datenauswertung im jeweiligen Bereich verwendet. Durch die vermehrte Implementierung von Electronic Health Records (EHRs) und die stetig steigende Menge an routinemäßig erfassten und elektronisch verfügbaren Gesundheitsdaten, gewinnen diese Daten im Sinne der Sekundärnutzung auch vermehrt das Interesse von Wissenschaft und Forschung. Eine wachsende Anzahl an Publikationen (unter anderem [1-5]) beschreiben sekundäre Anwendungen routinemäßig erfasster Gesundheitsdaten mit teils beachtlichen Ergebnissen, die das große Potential der Sekundärnutzung bereits erahnen lassen.

¹ Institut für Medizinisches Informationsmanagement und Bildverarbeitung, Zentrum für Medizinische Statistik, Informatik und Intelligente Systeme, Medizinische Universität Wien

In den genannten Publikationen werden aber immer wieder Unzulänglichkeiten der untersuchten Daten beschrieben, die Limitationen für deren Nutzung ergeben. Es wird häufig von Inkonsistenzen, Eingabefehlern (z.B. Missing Values) oder fehlender Codierung bestimmter Merkmalsarten (z.B. ICD für Diagnosen) berichtet. Zusammenfassend können diese Unzulänglichkeiten großteils unter dem Aspekt der Datenqualität subsumiert werden. Orr beschreibt in [6] auch ein allgemeines Problem hinsichtlich der Datenqualität indem er im Zuge seiner systemtheoretisch betrachteten sechs Regeln der Datenqualität wiederholt indirekt den Zeitaspekt bei der Datenerfassung bzw. Datenverwaltung und dessen Einfluss auf die Qualität der Daten adressiert.

2. Methoden

Im Zuge der vorliegenden Arbeit wird ein Vorgehensmodell präsentiert, das zu einer Verbesserung der Datenqualität in EHRs in Hinblick auf eine Sekundärnutzung beitragen soll.

Um die Thematik der Sekundärnutzung umfassend betreffend der Datenqualität zu beleuchten, muss eine Betrachtung aus zwei Blickwinkeln hinsichtlich des zuvor erwähnten Zeitaspekts erfolgen: Einerseits aus dem Blickwinkel einer zukunftsorientierten Betrachtung zur Erstellung neuer Datenquellen mit adäquater Datenqualität zur Sekundärnutzung („Ex-ante“) und andererseits aus dem vergangenheitsorientierten Blickwinkel der Datenqualitätsverbesserung bestehender Datenquellen („Ex-post“).

Zur Erstellung des Vorgehensmodells sollen möglichst existierende Modelle und Standards verwendet werden, die bereits eine initiale Annäherung der Bereiche klinische Routine und Forschung unterstützen und zwar unter Berücksichtigung der individuellen Anforderungen.

3. Ergebnisse

Es wurden zunächst bestehende Modelle und Standards ausgewählt, analysiert und beurteilt. Zur Abdeckung der „Ex-ante“-Sichtweise wurden Standards von CDISC (Clinical Data Interchange Standards Consortium) und HL7 (Health Level 7) betrachtet, die laut [7] und [8] auch im paneuropäischen Projekt EHR4CR (Electronic Health Records for Clinical Research) verwendet werden sollen. Zusätzlich wurde ein möglicher Ansatz von IHE (Integrating the Healthcare Enterprise) betrachtet, der einen Mechanismus zur Befüllung von Datenquellen berücksichtigt. Zur partiellen Abbildung der „Ex-post“-Sichtweise wird ein Ansatz zur Evaluierung bestehender Datenquellen im Kontext der Sekundärnutzung betrachtet.

3.1. Zukunftsorientierte Sichtweise – „Ex-ante“

3.1.1. HL7 EHR4CR FP (Electronic Health Record Clinical Research Functional Profile)

Die Organisation HL7 hat mit der Erstellung des so genannten HL7 EHR Clinical Research Functional Profile (HL7 EHR4CR FP) auf die zunehmende Relevanz der Sekundärnutzung von routinemäßig erfassten Gesundheitsdaten in EHRs reagiert. Dieses Functional Profile [9] baut auf dem HL7 EHR Functional Model (HL7 EHR FM) auf und erweitert Letzteres um die notwendigen Funktionen zur Nutzung, der in einem EHR gespeicherten Daten für die Forschung. Es soll damit eine Integration der Abläufe klinischer Forschung in die Abläufe der klinischen Routine ermöglicht werden, und zwar möglichst ohne erhöhten Aufwand im Zuge der Datenerfassung zu erzeugen (vgl. [9]).

Eine im Zuge der Analyse erhobene Differenzmenge der im HL7 EHR FM gegenüber dem HL7 EHR FM in ihrer Gesamtheit hinzugefügten bzw. geänderten Funktionen, zeigt genau jene Unzulänglichkeiten, die eine Verwendung von Standard-EHRs für die Forschung erschweren (beispielsweise fehlende Dokumentationsmöglichkeiten für Adverse Events oder Studienprotokollverletzungen, fehlende Identifier zur Durchführung von Studien).

3.1. 2. CDASH (Clinical Data Acquisition Standards Harmonization)

CDASH (Clinical Data Acquisition Standards Harmonization) stellt einen Teil der technischen Roadmap von CDISC (Clinical Data Interchange Standards Consortium) dar [10]. Die Verwendung nicht-standardisierter Lösungen wird von CDISC als Beschränkung für den Datenaustausch zwischen Klinik und Forschung gesehen. Eine Harmonisierung der Standards in der klinischen Forschung mit den Standards der routinemäßigen klinischen Gesundheitsversorgung wird daher angestrebt.

Im Rahmen des CDASH Standards wird neben den umfangreichen Domämentabellen in [10] auch ein Prozessmodell beschrieben. Dieses Modell ist zwar auf die Erstellung von eCRFs (elektronische Case Report Forms) für klinische Studien ausgelegt, es kann allerdings ohne umfassende Änderung auch auf die Erstellung von Datenerfassungsinstrumenten (z.B. Formulare in einem KIS) im Kontext der Routinedatenerfassung angewendet werden. Der wichtigste Punkt liegt dabei im Review des Entwurfs dieser Instrumente durch ein interdisziplinäres Team (Forschung/Klinik/IT), um die Abdeckung der individuellen Anforderungen dieser Anspruchsgruppen zu gewährleisten.

3.1. 3. IHE Profile RFD (Retrieve Form for Data Capture) und CRD (Clinical Research Document)

Die Initiative IHE entwickelt das Profil RFD (Retrieve Form for Data Capture) [11], welches aktuell im Status Trial-Implementation und demnach als Draft Version verfügbar ist. Ziel dieses Profils ist es, eine Methode zur Verfügung zu stellen, um innerhalb einer Anwendung (z.B. EHR-System) bestimmte Anforderungen eines externen Systems (z.B. EDC (Electronic Data Capture)-System zur Unterstützung klinischer Forschung) erfüllen zu können. Es soll mittels RFD möglich sein, direkt aus einer laufenden Applikation heraus ein Formular einer entfernten Formular-Quelle zu beziehen, dieses Formular mit Daten aus der laufenden Applikation vorzubelegen (Single Source Ansatz), es zu ergänzen und anschließend wieder an eine empfangende Applikation zurückzusenden.

Das Profil RFD macht keine Vorgaben bezüglich des Inhalts der Formulare. Als mögliche Empfehlung wurde das IHE Profil CRD (Clinical Research Document) [12] entwickelt. Es befindet sich aktuell ebenfalls im Trial Status und stellt ein Content Profil dar, in welchem auf die Parameter der unterschiedlichen Domämentabellen des zuvor bereits erwähnten CDASH Standards zurückgegriffen wird.

3. 2. Vergangenheitsorientierte Sichtweise – „Ex-post“

Zur Evaluierung der Datenqualität ist es notwendig vergleichbare Werte zu schaffen, auf deren Basis eine Aussage über das Niveau der Datenqualität in einer bestimmten Datenquelle getätigt werden kann. Nonnemacher et al. [13] empfehlen zu diesem Zweck die so genannte Source Data Verification (SDV). Dabei handelt es sich um ein (prinzipiell) quantitatives Verfahren, das es erlaubt, die vorliegende Datenqualität in Forschungsregistern zu bestimmen. Um eine repräsentative Maßzahl bzw. einen Gesamtscore zu erhalten, werden für die SDV die Ebenen Plausibilität, Organisation und Richtigkeit gebildet, die jeweils eine bestimmte Anzahl an Qualitätsindikatoren beinhalten. Diese Indikatoren wurden aus der Literatur bzw. aus Expertenbefragungen abgeleitet,

definiert und kategorisiert. Sie stellen jeweils Prozentzahlen dar, aus welchen in weiterer Folge unter Einbezug von Schwellwerten und Gewichtungsfaktoren ein Gesamtscore zur Bestimmung der Datenqualität einer betrachteten Datenquelle errechnet werden kann.

Die Methode der SDV zeigt somit eine Möglichkeit, die inhärente Datenqualität einer Datenquelle zu messen, vergisst jedoch einen wichtigen Punkt in Hinblick auf Standardisierung, da keine Maßzahl für die Evaluierung des Standardisierungsgrades einer Datenquelle berücksichtigt wird (z.B. Verwendung von Codesystemen und Klassifikationen). Im Sinne der notwendigen Interoperabilität wäre dies jedoch dringend erforderlich und müsste ergänzt werden.

3.3. Entwicklung des Vorgehensmodells

Abbildung 1 stellt das Schema eines möglichen Vorgehensmodells dar, welches die in Kapitel 2 beschriebene Zweiteilung hinsichtlich des zeitlichen Aspekts berücksichtigt und die zuvor gezeigten und existierenden Standards und Modelle strukturiert verbindet.

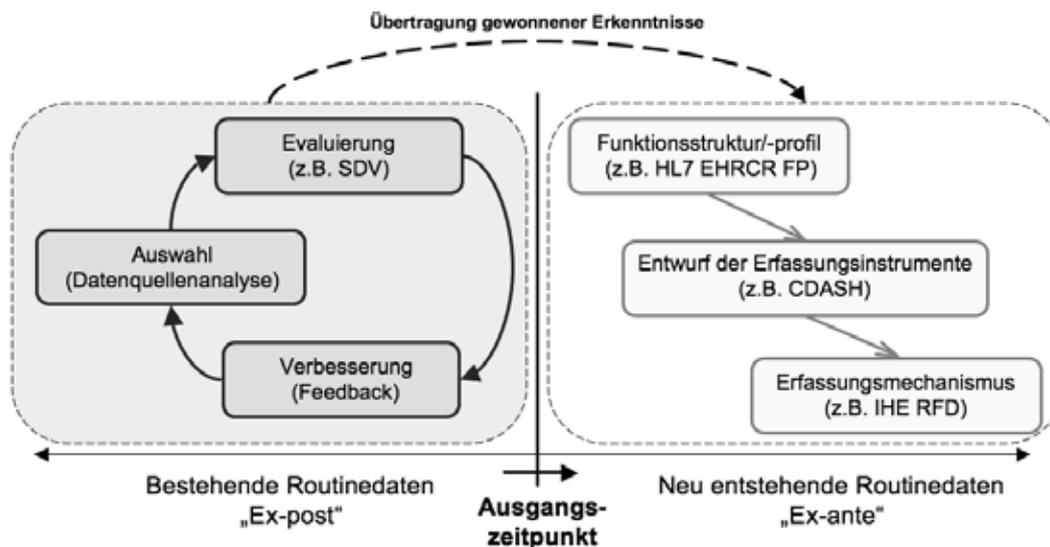


Abbildung 1: Vorgehensmodell zur Erhöhung der Datenqualität hinsichtlich Sekundärnutzung

So lässt sich für die vergangenheitsorientierte Betrachtung ein Kreisprozess und für die zukunftsorientierte Betrachtung eine Prozesskette abbilden.

Ausgehend von einem bestimmten Anwendungsfall beginnt der Kreisprozess („Ex-post“) mit der Auswahl einer bestimmten Datenquelle als Ergebnis einer Datenquellenanalyse (z.B. Auswahl aus einem Datenquellenkatalog anhand bestimmter Kriterien). Diese Datenquelle wird anschließend hinsichtlich ihrer Datenqualität evaluiert (z.B. mittels adaptierter SDV wie in Kapitel 3.2 angemerkt). Aus den Ergebnissen der Evaluierung werden Verbesserungsmöglichkeiten abgeleitet und mittels eines Feedbackmechanismus an den Betreiber der Datenquelle übermittelt. Zu einem späteren Zeitpunkt kann der Kreisprozess erneut beginnen. Das bedeutet, dass eine Datenquelle, die nach einem ersten Durchlauf aufgrund von Unzulänglichkeiten hinsichtlich der Evaluierungskriterien nicht als geeignet erschien, nach Umsetzung der Verbesserungsmöglichkeiten die Evaluierungskriterien in einem späteren Durchlauf des Kreisprozesses erfüllen und somit verwendet werden kann.

Die Prozesskette („Ex-ante“) zur Neuerstellung einer Datenquelle beginnt mit der Auswahl bestimmter, zum Ausgangszeitpunkt vorliegender Funktionsstrukturen/-profile (z.B. basierend auf

dem HL7 EHR4CR FP [9]). Auf Basis dieser Strukturen können anschließend Erfassungsinstrumente (Formulare etc.) umgesetzt werden, die zusätzlich auf bestimmten Standards (z.B. CDISC CDASH [10]) beruhen und interdisziplinäre Anforderungen erfüllen. Die Verwendung dieser Erfassungsinstrumente wird anschließend in Form eines bestimmten Erfassungsmechanismus (z.B. Erfassungsprozess und Akteure des IHE Profils RFD [11]) abgebildet.

Führt man diese Betrachtung aus der Gegenwart weiter über den Zeitstrahl nach rechts in die Zukunft, so lässt sich diese Kombination aus Kreisprozess und Prozesskette fortführen. D.h., ein zum Ausgangszeitpunkt t erstellter Erfassungsmechanismus für Routinedaten soll in Zukunft zum Ausgangszeitpunkt $t+x$ ebenfalls den Kreisprozess der vergangenheitsorientierten Betrachtung durchlaufen, um eine kontinuierliche Verbesserung der Datenqualität in der betrachteten Datenquelle zu erreichen. Die im Zuge des vergangenheitsorientierten Kreisprozesses aufgedeckten Verbesserungspotentiale müssen Einfluss auf die Prozessschritte bzw. die zugrunde liegenden Vorgaben der zukunftsorientierten Prozesskette aufweisen (Übertragung gewonnener Erkenntnisse), um jeweils den aktuellsten Erkenntnisstand bei der Erstellung neuer Erfassungsmechanismen zu berücksichtigen.

4. Diskussion & Ausblick

Generell greifen die beschriebenen Standards und Modelle im Hinblick auf Sekundärnutzung nativ nicht ausreichend ineinander, da der Zeitaspekt zu wenig Beachtung findet. Durch das gezeigte Vorgehensmodell kann unter Berücksichtigung dieser Integration und unter Einbezug der Zeitkomponente eine prozessorientierte Betrachtung zur Erreichung einer hochqualitativen Datenentstehung und Datenverarbeitung hinsichtlich Sekundärnutzung gewährleistet werden. Die Praktikabilität des gezeigten Vorgehensmodells sollte im Rahmen weiterer Forschungsarbeiten validiert werden.

Eine Alternative zur zuvor erwähnten Kombination der IHE Profile RFD und CRD liefern Fadly et al. [8], indem sie ein neues IHE Profil Cross Enterprise Template Sharing (XTS) vorschlagen, das zusätzlich einen Mappingmechanismus zwischen den Datenelementen des Studiensystems und des EHRs vorsieht, was einem Single Source Ansatz entspricht.

5. Literaturangaben

[1] Botsis T, Hartvigsen G, Chen F, Weng C. Secondary Use of EHR: Data Quality Issues and Informatics Opportunities. AMIA Summits Transl Sci Proc. 2010;2010:1-5.

[2] Breil B, Semjonow A, Müller-Tidow C, Fritz F, Dugas M. HIS-based Kaplan-Meier plots--a single source approach for documenting and reusing routine survival information. BMC Med Inform Decis Mak. 2011;11:11.

[3] Weber GM, Murphy SN, McMurry AJ, Macfadden D, Nigrin DJ, Churchill S, et al. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. J Am Med Inform Assoc. 2009;16(5):624-30.

[4] Hanauer D, Rhodes D, Chinnaiyan A. Exploring clinical associations using '-omics' based enrichment analyses. PLoS One. 2009;4(4):e5203.

[5] Roque FS, Jensen PB, Schmock H, Dalgaard M, Andreatta M, Hansen T, et al. Using electronic patient records to discover disease correlations and stratify patient cohorts. PLoS Comput Biol. 2011;7(8):e1002141.

[6] Orr K. Data quality and systems theory. Communications of the Acm. 1998;41(2):66-71.

[7] Lastic P-Y. Overview of the EHR4CR Project. 2011 [cited 12.01.2012]; Available from: <http://www.hl7.de/download/veranstaltungen/jahrestagungen/2011/Lastic.pdf>.

[8] El Fadly A, Rance B, Lucas N, Mead C, Chatellier G, Lastic PY, et al. Integrating clinical research with the Healthcare Enterprise: From the RE-USE project to the EHR4CR platform. J Biomed Inform. 2011;44 Suppl 1:S94-S102.

[9] Health Level 7. HL7 Clinical Research Functional Profile, Release 1. 2009.

[10] CDISC. Clinical Data Acquisition Standards Harmonization (CDASH). Austin, TX: Clinical Data Interchange Standards Consortium; 2011.

[11] IHE. IHE IT Infrastructure Technical Framework Supplement - Retrieve Form for Data Capture (RFD). 2011 [cited 04.11.2011]; Available from: http://www.ihe.net/Technical_Framework/upload/IHE_ITI_Suppl_RFD_Rev2-2_TI_2011-08-19.pdf.

[12] IHE. IHE Quality, Research and Public Health (QRPH) Technical Framework Supplement – Clinical Research Document (CRD). 2010 [cited 04.11.2011]; Available from: http://www.ihe.net/Technical_Framework/upload/IHE_QRPH_Suppl_CRD_Rev2-2_TI_2010-08-30.pdf.

[13] Nonnemacher M, Weiland D, Stausberg J. Datenqualität in der medizinischen Forschung - Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern. Berlin: MWV Medizinisch Wissenschaftliche Verlagsgesellschaft; 2007.

Corresponding Author

Karl Holzer

Institut für Medizinisches Informationsmanagement und Bildverarbeitung

Zentrum für Medizinische Statistik, Informatik und Intelligente Systeme

Medizinische Universität Wien

Spitalgasse 23, A-1090 Wien

E-Mail: karl.holzer@gmx.at