

Einsatzmöglichkeiten von CDISC ODM in der klinischen Forschung

Matthias Löbe^{1,2,3}, Christoph Aßmann¹, Richard Beyer¹, Jan Gaebel¹, Sebastian Kiunke¹, Edgar Lensing¹, Markus Nentwig¹, Christian Petzold¹, Antonia Siebert¹, Robin Sommerweiß¹, Lars Voitel^{2,4}

¹ Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE)

² Integriertes Forschungs- und Behandlungszentrum Sepsis und Sepsisfolgen (CSCC)

³ Integriertes Forschungs- und Behandlungszentrum AdipositasErkrankungen

⁴ Zentrum für Klinische Studien Leipzig (ZKSL)

matthias.loebe@imise.uni-leipzig.de

Abstract: Das CDISC Operational Data Model (ODM) ist ein populärer Standard in klinischen Datenmanagementsystemen (CDMS). Er beschreibt sowohl die Struktur einer klinischen Prüfung inklusive der Visiten, Formulare, Datenelemente und Codelisten als auch administrative Informationen wie gültige Nutzeraccounts. Ferner enthält er alle erhobenen klinischen Fakten über die Probanden. Sein originärer Einsatzzweck liegt in der Archivierung von Studiendatenbanken und dem Austausch klinischer Daten zwischen verschiedenen CDMS. Aufgrund der reichhaltigen Struktur eignet er sich aber auch für weiterführende Anwendungsfälle. Im Rahmen studentischer Praktika wurden verschiedene Szenarien für funktionale Ergänzungen des freien CDMS OpenClinica untersucht und implementiert, darunter die Generierung eines Annotated CRF, der Import von Studiendaten per Web-Service, das semiautomatisierte Anlegen von Studien sowie der Export von Studiendaten in einen relationalen Data Mart und in ein Forschungs-Data-Warehouse auf Basis von i2b2.

1 Einleitung

Seit 2008 fördert das Bundesministerium für Bildung und Forschung (BMBF) „Integrierte Forschungs- und Behandlungszentren“ (IFB)¹, welche die Translation zwischen grundlagenbezogener und patientenbezogener klinischer Forschung stärken sollen. Das Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE) und das Zentrum für Klinische Studien Leipzig (ZKSL) unterstützen die IFBs AdipositasErkrankungen (Leipzig) und Sepsis und Sepsisfolgen (Jena) methodisch und infrastrukturtechnisch in einem der zentralen Arbeitsgebiete, der Durchführung klinischer Studien. Eine zentrale Aufgabe ist hierbei die Bereitstellung eines leistungsfähigen Clinical Data Management Systems vor dem Hintergrund regulatorischer Vorschriften wie GCP oder FDA 21 CFR Part 11 bzw. Richtlinien aus ECRIN [Oh11].

¹ <http://www.ifb-portal.de/>

OpenClinica² ist ein quelloffenes CDMS mit einer modernen Architektur. Es setzt auf einem Java EE Applikationsserver auf (z.B. Apache Tomcat) und verwendet populäre Frameworks wie JavaServer Pages, Spring, Hibernate und Liquibase, die selbst wieder als Open-Source-Software zur Verfügung stehen. Somit lässt sich, im Zusammenspiel mit PostgreSQL als Datenbank, eine Infrastruktur mit beliebig vielen lizenzkostenfreien Instanzen zusammenstellen.

Case Report Forms (CRF) zur Dokumentation von Probandendaten werden mit Hilfe einer Excelvorlage spezifiziert. Beim Upload werden aus dieser Vorlage sowohl die Datenbank als auch die webbasierten Eingabemasken und die Validierungsprüfungen erstellt. Probandendaten aus unterschiedlichen Zentren oder Studien sind logisch getrennt. OpenClinica verfügt über eine komplexe Rechtestruktur mit typischen Rollen aus dem Bereich Datenmanagement (Dokumentar, Studienarzt, Hauptprüfer etc.). Änderungen werden in einem Audit Trail festgehalten. OpenClinica kennt leistungsfähige Funktionen zur Sicherstellung der Datenqualität, z.B. komplexe feldübergreifende Regeln, und verwaltet Plausibilitätsfragen in einem anspruchsvollen Querymanagementsystem [Lö11]. Es wird seit Januar 2011 produktiv in derzeit 10 Studien mit 7.500 Probanden eingesetzt [Me11].

OpenClinicas Datenmodell lehnt sich stark an den CDISC ODM-Standard³ an, weshalb das Datenbankschema intuitiv verständlich ist. Wie in vielen Systemen, die potenziell sehr große Datenmengen verwalten müssen, wurde ein Entity-Attribute-Value-Ansatz mit einer zentralen Faktentabelle ITEM_DATA gewählt. Ferner bietet OpenClinica einen ODM-Export als XML-Serialisierung in den Versionen ODM 1.2 und ODM 1.3 an, optional angereichert durch „Vendor Extensions“, die bestimmte, OpenClinica-spezifische Zusatzinformationen in einem eigenen XML-Schema abbilden. ODM wurde mit dem Fokus entworfen, klinische Studiendaten und -metadaten zwischen CDMS austauschen zu können [Ku06] bzw. rechtssicher zu archivieren [Ku09]. Darüber hinaus unterstützt ODM als internationaler, XML-basierter, strukturierter und metadatenreicher Standard [Na11, S. 376] die Entwicklung moderner, metadatengetriebener Applikationen.

2 Methoden

Im Rahmen des Mastermoduls „Informationsmanagement in der klinischen Forschung“⁴ bearbeiten die Studenten neben der Vorlesung auch ein umfangreiches Praktikum allein oder in kleinen Gruppen. Dabei wird besonderer Wert auf einen möglichst engen Bezug zur Praxis gelegt. Im Sommersemester 2010 und 2011 lag der Schwerpunkt in der Evaluation des CDMS OpenClinica und im Besonderen auf Möglichkeiten, zusätzliche Funktionalitäten auf Basis des ODM-Exports prototypisch zu implementieren. Daraus ergaben sich klare Anhaltspunkte für die Eignung von ODM als Austauschformat für die Anbindung externer Systeme zu einer integrierten Forschungsinfrastruktur.

² <https://www.openclinica.com/>

³ <http://www.cdisc.org/odm>

⁴ <http://www.imise.uni-leipzig.de/Lehre/Semester/2012/IMKlinischeForschung/>

2.1 Generierung eines Annotated CRF

Im ZKS Leipzig ist der annotierte CRF ein Hilfsmittel für die Kommunikation zwischen Biometriker, Datenbankentwickler und Datenmanagement. Optisch entspricht er dem CRF in Papierform, zusätzlich sind aber wichtige statistische Größen wie Variablenamen und Codierungen farblich angefügt. Er wurde bislang manuell erstellt, layoutet und gepflegt. Der im Rahmen des Praktikums implementierte annotierte CRF extrahiert eine solche Maske automatisch aus dem ODM und umfasst eine grafische Darstellung der Protokollmetadaten, d.h. Visiten, Formulare und Itemgruppen sowie eine Darstellung aller Datenelemente (Items), ihrer Bezeichnungen, Einheiten und Kodierungen in den jeweiligen Containern der Studienhierarchie.

Basis der informatischen Umsetzung war die Programmierung eines Parsers für ODM-Dateien in ein Java-Objektmodell. Für die Überführung des normativen XML-Schemas in das Objektmodell wird JAXB (Java Architecture for XML Binding) genutzt. Unter Anwendung des Binding Compilers (xjc) werden für jede ODM-Schemaversion Java-Klassen für die Datenbindung erstellt. Jedem im ODM-Schema definierten Element ist eine Java-Klasse zugeordnet. Aus dem erstellten Objektmodell ist anschließend das webbasierte Formular zu erstellen, um bestehende ODM-Dateien im Browser darstellen zu können. Hierzu wird das JEE Framework für Webanwendungen Apache Click⁵ verwendet.

2.2 Automatisierter Import von Studiendaten

Das Ausfüllen von elektronischen Formularen und noch mehr das Abtippen von Daten aus Papier-CRFs sind aufwendig und fehlerträchtig. Vorteilhafter wäre die direkte Datenübernahme aus den primären Informationssystemen, solange eine ausreichende Datenqualität sichergestellt werden kann. Dafür wurden unter dem Schlagwort Single Source oder Secondary Use of Electronic Health Records verschiedene Anwendungsszenarien erarbeitet [Fa07], [Fa10].

Die aktuelle Version 3.1 von OpenClinica wurde in ihrer Architektur grundlegend überarbeitet und modularisiert. Neben einer Kernkomponente existieren nun Komponenten für den WWW-Zugriff und für eine Webservice-Schnittstelle. Letztere nutzt Spring WS⁶ und basiert auf den W3C-Standards SOAP⁷ und WSDL⁸. Klinische Daten aus Krankenhausinformationssystemen können als ODM automatisch importiert werden.

Alle verfügbaren Webservices wurden unter Zuhilfenahme des Werkzeugs soapUI⁹ erfolgreich getestet. Des Weiteren wurde ein eigener Webservice mit JAXB implementiert, der einen OpenClinica-Nutzeraccount anlegt, was bei potenziell hunderten Prüfarzten pro Studie eine deutliche Arbeitserleichterung darstellt.

⁵ <http://click.apache.org/>

⁶ <http://static.springsource.org/spring-ws/sites/2.0/>

⁷ <http://www.w3.org/TR/soap/>

⁸ <http://www.w3.org/TR/wSDL/>

⁹ <http://www.soapui.org/>

2.3 Transfer einzelner Studien zwischen OpenClinica-Instanzen

Während OpenClinica einen vollständigen Export einer Studie inklusiver der klinischen Daten, der administrativen Daten und der Metadaten (Struktur der Studie) im CDISC ODM-Format anbietet, ist ein Import nur für die klinischen Daten eines Probanden möglich. Studienmetadaten müssen entweder manuell über die Webmaske eingetragen (Studie, Zentren, Events, Nutzerkonten) bzw. als Excel-Tabelle spezifiziert werden (CRFs, Items). Um Studien zwischen verschiedenen Instanzen austauschen zu können, ist eine Transformation des ODM-Exports in solche Komponenten nötig, die von OpenClinica wieder geladen werden können. Ein Export/Import einer Studie kann in vielen Fällen nützlich sein:

- beim Testen unterschiedlicher Softwarestacks, z.B. verschiedener Datenbanken
- beim Überspielen aus Test- in Produktivsysteme
- beim Update auf eine neue Version oder der Migration auf ein anderes CDMS
- bei der lokalen Dateneingabe mit Notebooks für Zentren ohne Internetzugang
- beim Export von Teilstrukturen, z.B. der Nutzerzugänge oder der CRF-Bibliothek
- bei der Verteilung einzelner Studien aus Performanzgründen

Im vorliegenden Fall wurde dreiteilig vorgegangen: Das eigentliche Anlegen einer Studie erfordert die Eingabe relativ vieler Metadaten (Registriernummer, erwartete Patientenzahl, Principal Investigator, ...). Dies lässt sich aktuell nicht automatisieren, zumal viele der Werte nicht im ODM-Export enthalten sind. Für die vorhandenen Werte wurde eine Checkliste erzeugt. Der zweite Teil bestand in Erstellung der Excel-CRF-Spezifikation. Dafür wurde das ODM-XML in ein Java-Objektmodell geparkt und dann mit Hilfe von Apache POI¹⁰ valide Exceldokumente erstellt. Der letzte Teil bestand im Import der Probandendaten, die bereits als ODM vorliegen. Mithilfe der Webservice-API ist dies recht trivial.

2.4 Export von Studiendaten in einen Data Mart

OpenClinica bietet für Auswertungszwecke den Export in verschiedene Statistiksoftware (z.B. SPSS) an. Viele Biometriker sind es jedoch gewöhnt, auf SQL-Datenbanken zu arbeiten. Das OpenClinica-Datenbankschema benutzt aber einen EAV-Ansatz, der für direkten analytischen Zugriff ungeeignet ist. Für ein solches Szenario müssen die Daten in ein übliches relationales Schema übersetzt werden, wo in den Fakten-Tabellen die einzelnen Items als Spalten und die Probanden als Zeilen repräsentiert werden. Eine solche Transformation kann sehr komplex und fehleranfällig sein [Br02].

OpenClinica verwendet in Version 3.1 ODM als primäres Exportformat, aus dem alle anderen Exportformate wie SPSS, Excel oder HTML erzeugt werden. Dies geschieht über einen XSLT-Postprozessor. Dieser ist über Konfigurationsdateien anpassbar, daher war es möglich, eine zusätzliche Exportoption für SQL-Queries hinzuzufügen. Diesem Konfigurationseintrag wird ein selbsterstelltes XSLT-Skript übergeben, welches für jede ItemGroupDef ein CREATE TABLE Statement erzeugt und für jedes ItemRef eine

¹⁰ <http://poi.apache.org/>

werden. Dies erfolgte unter Einsatz des Open-Source-ETL-Tools Talend Open Studio¹², welches die Verarbeitung einer Vielzahl von Datenquellen und –senken über eine grafische Oberfläche unterstützt (siehe Abbildung 1).

3 Ergebnisse

3.1 Generierung eines Annotated CRF

Die automatische Erzeugung des annotierten CRF spart viel Zeit, wenn in der Entwurfsphase und eventuell auch später während der Durchführungsphase einer klinischen Studie viele Versionen erstellt und mit der Datenbankumsetzung konsistent gehalten werden müssen (siehe Abbildung 2).

MEDUSA
Follow-up - V1.0

F_FOLLOWUP_V10-Ungrouped

Annotationen ausblenden

Label	Annotation
Entlassung ITS ME-ITSEntlass_D	<input type="text"/>
Status bei Entlassung ITS ME-ITSEntlassStatus	<input type="radio"/> überlebt <input type="radio"/> verstorben
Entlassung Krankenhaus ME-KKHEntlass_D	<input type="text"/>
Status bei Entlassung Krankenhaus ME-KKHEntlassStatus	<input type="radio"/> überlebt <input type="radio"/> verstorben

Abbildung 2: Annotierter CRF der MEDUSA-Studie (Ausschnitt)

Problematisch waren gewisse Eigenarten der ODM-Interpretation von OpenClinica, z.B. die Benutzung von ItemGroups nicht als Untersektionen, sondern um komplexe, potenziell mehrfach auftretende Itemtupel wie mikrobiologische Untersuchungen abzubilden. Des Weiteren trifft ODM keine Annahmen über das CRF-Layout, so wird z.B. nicht zwischen Optionsfeldern und Auswahllisten unterschieden (eigentlich ist nicht einmal die Unterscheidung zwischen Einfach- und Mehrfachauswahlfeldern möglich).

Der gewählte Ansatz lässt sich für eine ganze Reihe weiterer Szenarien einsetzen, z.B. für Studienregister, CRF-Mahnwesen oder zur Visitenplanung im Patientenkalender [Ae11].

3.2 Automatisierter Import von Studiendaten

Zum Zeitpunkt des Praktikums verfügte leider keines der relevanten Systeme (Laborinformationssystem, Patientendatenmanagementsystem, Intensivstation) über die

¹² <http://de.talend.com>

Fähigkeit zum Export von Daten im ODM-Format. Daher konnte der Import nur prototypisch getestet werden (siehe Abbildung 3). Inzwischen werden Daten aus jenen Studienzentren im Routinebetrieb elektronisch per ODM übertragen, die das „Sepsis Informationssystem zur Qualitätssicherung“¹³ (SIQ) verwenden, für das ein ODM-Export extern entwickelt wurde.

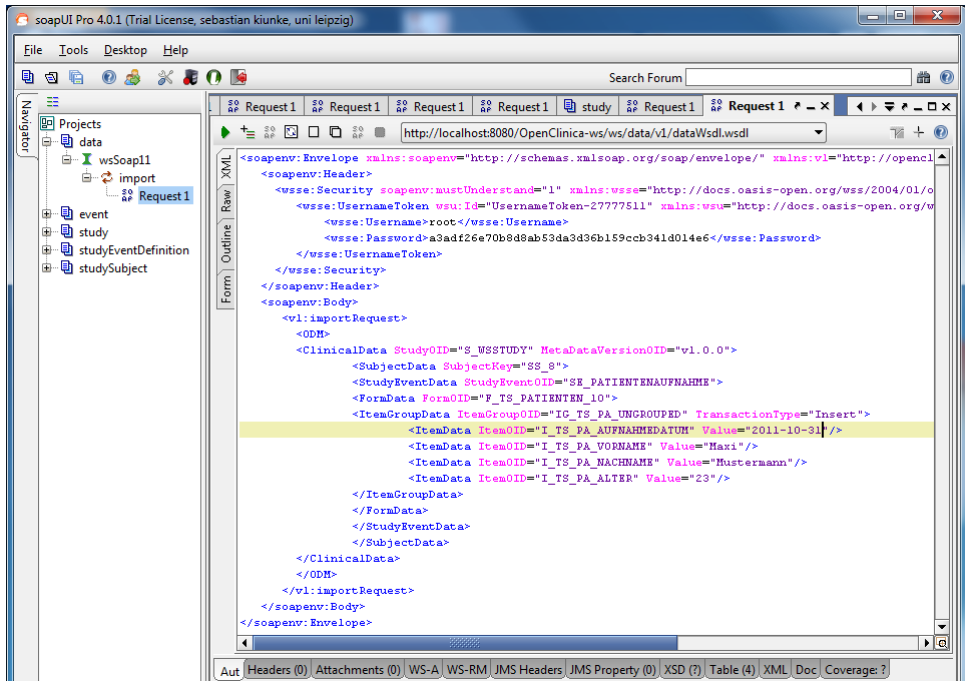


Abbildung 3: Import von klinischen Daten per Webservice im ODM-Format

Der Import von klinischen Daten ist ein nativer Anwendungsfall für ODM und wahrscheinlich der zukünftig am meisten genutzte. Die Geschwindigkeit, Aktualität und Qualität der Dateneingabe ist manuell nicht erreichbar. Dies gilt jedoch nur soweit, wie bei der Erstellung von ODM-Exporten sorgfältig gearbeitet wurde. Beim automatisierten Import von ODM-Dateien muss OpenClinica (wie jedes andere System auch) entscheiden, wie bei ungültigen OIDs, Datentypen, Formaten, Codelistenoptionen, bei fehlenden Pflichtfeldern, falschen CRF-Versionen oder Normbereichsverletzungen verfahren werden soll. Entweder führt dies zu einer strikten Ablehnung (wobei die Fehlergründe aufgrund der maschinellen Erstellung der ODM-Datei eventuell für den Dokumentar gar nicht korrigierbar sind) oder zu vielen Queries, deren Behebung aufwendiger ist als die manuelle Eingabe.

¹³ <http://siqmed.de/>

3.3 Transfer einzelner Studien zwischen OpenClinica-Instanzen

Die Entwicklung einer produktionsreifen Migrationslösung kann enorm aufwendig sein [Ye11]. In unserem Fall erschwerten verschiedene Faktoren die Entwicklung eines umfassenden Export/Import-Moduls. So enthält der ODM-Export aktuell bei weitem nicht alle notwendigen Komponenten, z.B. keinen vollständigen Auditlog, keine Queryhistorie, keine Nutzerberechtigungen und nicht alle CRF-Spalten. Ferner stellte sich heraus, dass es größere Unterschiede zwischen den Version 3.0 und 3.1 gab, die getrennte Parser nötig machten. Auch die Behandlung interner Datenbank-IDs ist problematisch.

3.4 Export von Studiendaten in einen Data Mart

Die angebotene Lösung als Data Mart eignet sich gut für Auswertungszwecke. Der SQL-Export lässt sich gegebenenfalls noch verfeinern, um beispielsweise bestehende SQL-Makros ablaufen lassen zu können. Vorteilhaft ist auch die leichtere Zusammenführbarkeit mit externen Daten, die häufig in tabellarischen Formaten vorliegen, entweder aus Primärsystemen der Versorgung oder aus anderen Studienprojekten.

	subjectkey character va	l_testc_date character va	l_testc_singleselect character varying(25)	l_testc_multiselect character varying(25)	l_testc_checkbox character varying(2)	l_testc_textarea character varying(l_testc_radiobutton2 character varying(25)	l_testc_radiobutton character varying(25)
1	SS_03	2012-01-12	0,2	0,1	test	1		
2	SS_04	2012-01-10	2			1		

Abbildung 4: Ausschnitt des Data Marts als PostgreSQL-Datenbank

Das entstandene XSLT-Skript zeigt nur die prinzipielle Möglichkeit für den Export der klinischen Daten auf. Für eine Routinenutzung müssten eine Reihe von Anpassungen vorgenommen werden, z.B. für die Erkennung der korrekten Datentypen und verschiedene Formatierungsprobleme als auch im Hinblick auf die Performanz sowie die Unterstützung von REPEATED_GROUPS und mehrfach auftretenden Itemtupeln.

3.5 Export von Studiendaten in ein Data Warehouse

Der leichtere Zugriff für klinische Forscher über die Oberfläche des Data Warehouse (siehe Abbildung 5) erlaubt diesen vielfältige Abfragealternativen spielerisch in kurzer Zeit durchzuführen. Er kann aber nicht ohne teilweise erheblichen Aufwand realisiert werden. Einerseits ist die Einarbeitungszeit in ein funktionsreiches ETL-Tool wie Talend Open Studio nicht zu unterschätzen, andererseits ist die Transformation von CDISC ODM in das i2b2 Datenbankschema komplex, weniger generisch als erwartet und im Routinebetrieb zu einem gewissen Teil fehlerträchtig, besonders wenn Modifikationen an der Studienstruktur vorgenommen werden.

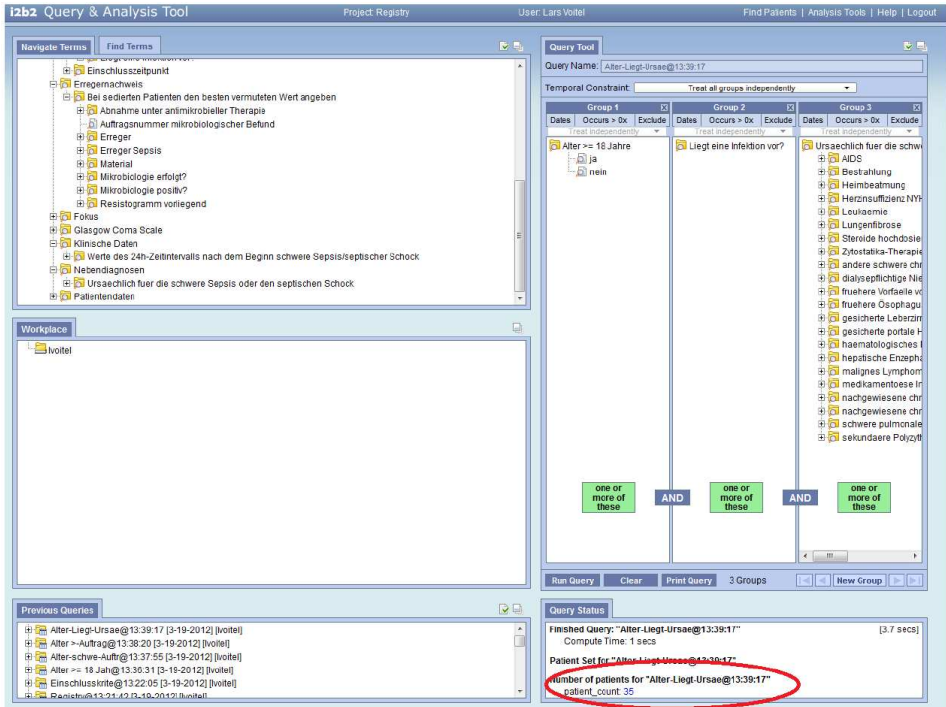


Abbildung 5: Selektion einer Teilkohorte des zentralen Sepsisregisters in i2b2

4 Diskussion und Ausblick

CDISC ODM eignet sich sehr gut, zusätzliche Funktionen an bestehende CMDS anzubinden. Dabei kann es sich sowohl um funktionale Ergänzungen für das System selbst als auch um die Unterstützung der Integration in eine größere Forschungsinfrastruktur handeln. Es zeigt sich, dass Ansätze, die auf offene, standardbasierte Software setzen, leicht wart- und erweiterbar sind.

Die hier vorgestellten Beispiele haben nicht den Anspruch, produktionsreife Lösungen anzubieten, sondern sind paradigmatisch zu sehen. Die Praktikumsarbeiten und Ergebnisse lassen sich über das Instituts-Wiki abrufen^{14,15}.

Danksagung

Die vorliegende Arbeit wurde vom BMBF als Teil der Verbundvorhaben IFB Sepsis und Sepsisfolgen (CSCC) und IFB AdipositasErkrankungen sowie des Zentrums für

¹⁴ <https://wiki.imise.uni-leipzig.de/Lehre/2010SS/VIKF/Praktikum>

¹⁵ <https://wiki.imise.uni-leipzig.de/Lehre/2011SS/VIKF/Praktikum>

Klinische Studien Leipzig (ZKSL) gefördert (Förderkennzeichen: 01E01002, 01E01001, 01KN1102).

Literaturverzeichnis

- [Ae11] Aerts, J. Generating a caBIG Patient Study Calendar from a Study Design in ODM with Study Design Model Extension. CDISC Journal, October 2011.
- [Br02] Brandt CA, Morse R, Matthews K, Sun K, Deshpande AM, Gadagkar R, et al. Metadata-driven creation of data marts from an EAV-modeled clinical research database. International journal of medical informatics. 2002 Nov;65(3):225-241.
- [Fa07] El Fadly A, Daniel C, Bousquet C, Dart T, Lastic PYY, Degoulet P. Electronic Healthcare Record and clinical research in cardiovascular radiology. HL7 CDA and CDISC ODM interoperability. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2007;p. 216-220.
- [Fa10] El Fadly A, Lucas N, Rance B, Verplancke P, Lastic PYY, Daniel C. The REUSE project: EHR as single datasource for biomedical research. Studies in health technology and informatics. 2010;160(Pt 2):1324-1328.
- [Ga11] Ganslandt T, Mate S, Helbing K, Sax U, Prokosch HU. Unlocking Data for Clinical Research – The German i2b2 Experience. Applied Clinical Informatics. 2011 Mar;2(1):116-127.
- [Ku06] Kuchinke W, Wiegelmann S, Verplancke P, Ohmann C. Extended cooperation in clinical studies through exchange of CDISC metadata between different study software solutions. Methods Inf Med. 2006;45(4):441–6.
- [Ku09] Kuchinke W, Aerts J, Semler SC, Ohmann C. CDISC standard-based electronic archiving of clinical trials. Methods of information in medicine. 2009 Jul;48(5):408-413.
- [Lö11] Löbe M, Meineke F: Open Source Clinical Data Management mit OpenClinica, 56. GMDS-Jahrestagung und 6. DGEpi-Jahrestagung, Mainz, 2011.
- [Me11] Meineke FA, Löbe M, Stäubert S, Kropf S, Engel C, Hagel S, Bahr V, Rissner F, Hagel S, Bloos, Bauer, Ludewig K, Brunkhorst FM: Integration of the clinical data management system OpenClinica into the research infrastructure of the Center for Sepsis Control and Care (CSCC), Weimar Sepsis Update 2011
- [Na11] Nadkarni, PM: Metadata-driven Software Systems in Biomedicine - Designing Systems that can adapt to Changing Knowledge. Springer, Heidelberg, 2011.
- [Oh11] Ohmann, C. et al. (2011). Standard requirements for GCP-compliant data management in multinational clinical trials. Trials 12(1):85+.
- [Ye11] Yeomans, A. Using CDISC ODM to Migrate Data. CDISC Journal, October 2011.