



Design eines Mappings und Implementierung eines Converters von LimeSurvey Umfragen und Antworten in CDISC ODM

Bachelorarbeit

vorgelegt von:

Antonius Johannes Mende

Matrikelnummer: 461 328

Studiengang: Informatik 1.FB

Thema gestellt von:

Dr. Ludger Becker

Arbeit betreut durch:

Dr. Ludger Becker und Tobias Brix

Münster, 14. Mai 2021

Inhaltsverzeichnis

1	Einleitung	1
1.1	Problem	1
1.2	Ziel	1
2	Methodik	3
2.1	Akronyme	3
2.2	LimeSurvey	3
2.2.1	Fragegruppen	4
2.2.2	Fragetypen	4
2.2.3	Export	4
2.3	ODM	5
2.4	dom4j	5
3	Ergebnisse	7
3.1	Mapping	7
3.1.1	Dummy Elemente in ODM	7
3.1.2	Umfrage-Eigenschaften	7
3.1.3	Fragegruppen	7
3.1.4	Fragen	7
3.2	Implementierung	7
4	Diskussion	9
4.1	Weglassen von Fragetypen	9
4.1.1	Datei-Upload	9
5	Fazit	11

1 Einleitung

Das Thema dieser Arbeit ist die Konvertierung von LimeSurvey Archiven in das Operational Data Model. LimeSurvey ist ein Werkzeug, mit dem man einfach Umfragen erstellen kann, welche dann wiederum von beliebig vielen Teilnehmern beantwortet werden können. Umfragewerkzeuge lassen sich sehr vielseitig einsetzen, um die Meinungen von Kunden, Patienten oder jedem anderen Menschen einzuholen. Gerade in den heutigen Zeiten, wo immer mehr Daten gesammelt und verarbeitet werden können, ist es wichtig, Kompatibilität zwischen verschiedenen Werkzeugen und Formaten herstellen zu können. Das Institut für Medizinische Informatik an der WWU ist eine der vielen Institutionen, welche LimeSurvey nutzt, um die Meinungen und Erfahrungen von Patienten einzuholen.

1.1 Problem

Der XML-Standard lässt dem Programmierer sehr viel Freiheit, was das Design eines XML-Formates angeht. Dadurch sind verschiedene XML-Formate im Regelfall nicht kompatibel beziehungsweise untereinander austauschbar. Das macht es notwendig, XML-Dokumente zu konvertieren, um ein Dokument in einem Format mit einem Werkzeug nutzen zu können, welches nur ein anderes XML-Format unterstützt. Als Zielformat wird ODM-XML von CDISC verwendet. ODM wurde mit dem Ziel entwickelt, den Austausch und die Archivierung von Forschungsdaten und anderen damit verbundenen Daten zu ermöglichen. Durch die Unabhängigkeit des Formates von spezifischen Plattformen oder Firmen wird es durch weit mehr Werkzeuge unterstützt als ein Format wie LSA, welches von LimeSurvey selbst und zu deren eigenen Zwecken entwickelt wurde. Dies sieht man auch an dem Institut für Medizinische Informatik, welche dieses Projekt betreuen. Dort wird ODM täglich zur Verwaltung klinischer Daten genutzt, gleichzeitig ist aber auch LimeSurvey das Umfragewerkzeug der Wahl, um Daten von Patienten zu sammeln. Diese gesammelten Daten müssen dann exportiert und konvertiert werden, um sie in bestehenden Systemen einpflegen zu können.

1.2 Ziel

Ziel der Arbeit ist es, genau darzustellen, wie man aus einem LimeSurvey Archiv in ein ODM Dokument umwandeln kann und einen Konverter zu imple-

1 Einleitung

mentieren, welcher das bewerkstelligt. Es sollen so viele der Forschungsdaten wie möglich übertragen werden, zusätzliche Daten, welche z.B. die Darstellung der Daten betreffen, sollen nicht mit übertragen werden.

2 Methodik

2.1 Akronyme

XML die Extensible Markup Language

CDISC Clinical Data Interchange Standards Consortium

CDISC ODM von CDISC entwickeltes XML-Format „Operational Data Model“

lsa Dateiendung des LimeSurvey Archives (siehe Abschnitt 2.2.3)

lsr Dateiendung der LimeSurvey Response Datei (siehe Abschnitt 2.2.3)

lss Dateiendung der LimeSurvey Struktur Datei (siehe Abschnitt 2.2.3)

EDC Electronic Data Capture, das Sammeln und Verarbeiten von Daten

DOM Document Object Model. Bietet die Möglichkeit, die Hierarchie der XML-Knoten in Baumform darzustellen und so zu navigieren/ den Baum zu bearbeiten

SAX Simple API for XML. Ein Standard, welcher beschreibt, wie man ein XML-Dokument parsen kann. Dieses wird sequentiell eingelesen und für definierte Ereignisse wird eine vorgegebene Rückruf-Funktion aufgerufen. Ein Programm kann eigene Funktionen registrieren und so das Dokument verarbeiten.

2.2 LimeSurvey

- Umfragewerkzeug, sammeln von Meinungen/Interessen/Entscheidungsgrundlagen - Firma gleichen Namens aus D - Wird von den Entwicklern in der Cloud-Edition angeboten - 4 Pläne - Free - Basic (34 Euro/Monat) - Expert (29Euro/Monat) - Enterprise (74 Euro/Monat) - Code auch Open-Source

2.2.1 Fragegruppen

- Name/Beschreibung/Randomisierungsgruppe/Relevanz - Kann beliebig viele Fragen enthalten

2.2.2 Fragetypen

- 36 Fragetypen (manche sind keine "Fragen" per se, davon gibt es 28), 5 Gruppen - Bedingungen

- Single Choice
- Skala 1-5 - Bootstrap/Dropdown/Radio sind untersch. Darstellungen - Liste mit Kommentar hat nur Kommentar - Image Select List Radio hat Bild dazu
- Arrays
- Man kann aus einer Festen Liste an Möglichkeiten eine pro Subfrage Auswählen - 5 Punkte, 10 Punkte, Erhöhung/Gleich/Erniedrigung, Ja/Nein/Unsicher, Selbst definierte Möglichkeiten - Nach Spalte: Gleiches wie Freitext, nur Achsen getauscht - Dual Scale: Zwei mal selbst definierte Möglichkeiten, aus beiden Skalen eine Antwort pro Subfrage auswählbar. - Matrizen - Freitext oder freie Zahl
- Multiple Choice
- Darstellungen: Bootstrap, simples Multiple Choice - Mit Bild oder mit Kommentar (Ein Kommentar pro Antwortfeld)
- Textfragen
- Browser Detect - Kurz/Normal/Lang-er Freitext - Mehrere kurze Freitexte
- Input on Demand
- Maskenfragen
- Tag/Zeit - Ja/Nein - Gleichung - Dateiupload - Geschlecht - Sprachumschaltung - Zahleneingabe (Auch mehrfach, mit Limitierbaren Maximal/Minimal-Werten/Summen, Genauen Gesamtergebnissen und nur Ganzzahl) - Ranking (Advanced) - Textanzeige

2.2.3 Export

LimeSurvey Archiv

Das LimeSurvey Archiv ist eine von sieben Möglichkeiten, Daten von einer LimeSurvey Umfrage zu exportieren. Ein solches Archiv ist eine komprimierte Datei im .lsa-Format, welche mehrere extrahierbare Dateien enthält. Die Zahl und Art der Dateien ist dabei abhängig von den Einstellungen. Zwei Dateien sind immer enthalten:

Die erste Datei enthält die Umfrage-Struktur sowie Informationen über die Art und Weise, wie die Fragen dargestellt werden sollen (die .lss-Datei), die zweite Datei enthält die Antworten der Teilnehmer (.lsr-Datei). Ausdrücklich erwähnt wird dabei, dass Dateien, die als Antwort auf eine Frage hochgeladen wurden,

nicht Teil des Archivs sind. Weitere optionale Dateien sind eine Token-Datei, welche Daten über die

Weitere Exportmöglichkeiten

LSS Es ist möglich, nur die im LSA enthaltene LSS-Datei zu exportieren, das wird mittels dieser Option gemacht.

Excel/.csv Hier sind weitere Einstellungen möglich, wie das Exportieren eines Teils der Antworten oder die Wahl eines bestimmten Formates (Word, Excel, CSV, HTML, PDF).

SPSS SPSS ist ein Software-Paket, welches zur statistischen Analyse von Daten genutzt wird. Auch hier kann ausgewählt werden, welche Antworten exportiert werden sollen. Die Nutzung der Open-Source Version PSPP ist auch möglich

R R ist eine Alternative zu SPSS, hier werden allerdings alle Daten exportiert

STATA-xml Auch STATA ist eine kommerzielle Lösung für Datenanalyse wie SPSS. Hierfür werden die Daten von LimeSurvey direkt in das proprietäre STATA-Format umgewandelt.

VV Durch "vertical verification" ist es möglich, die Antworten zu modifizieren und die modifizierte Datei dann wieder zu importieren

2.3 ODM

- Ein Datenstandard im XML-Format - entwickelt von der CDISC - Ziel: Ein klares, einheitliches Format für Klinische Daten - Zusätzliche Daten wie Metadaten oder admin. Daten werden auch gespeichert - Plattform-unabhängig - Unabhängig von spezifischen Firmen - Mittlerweile viel genutzt in EDC Tools

2.4 dom4j

- API für den Zugriff und die Verarbeitung von XML-Dokumenten - Bietet simple Wege, um existierende Funktionen wie XPath oder Parsing mittels DOM/SAX zu nutzen - Auch Dokumente erstellen möglich

3 Ergebnisse

3.1 Mapping

3.1.1 Dummy Elemente in ODM

- ODM kann mehr Beziehungen in Studien darstellen - Es müssen Dummy-Elemente angelegt werden, weil LimeSurvey diese Elemente nicht hat, sie aber in ODM vorkommen müssen - Betroffen: GlobalVariables, StudyEvent

3.1.2 Umfrage-Eigenschaften

- Eine Umfrage wird als ein „*Form*“ dargestellt - Titel, Beschreibung und ID der Umfrage werden zu Titel, Beschreibung und ID des „*Forms*“

3.1.3 Fragegruppen

- Aus jeder LS Gruppe wird eine ItemGroup in ODM - Eigenschaften werden 1:1 abgebildet

3.1.4 Fragen

- Aus einer Frage in LS werden pot. mehrere ItemDef's in ODM (fragetypabhängig) - Elemente aus „*Questions*“ und „*Subquestions*“ und „*Question_l10ns*“ werden genutzt

3.2 Implementierung

4 Diskussion

4.1 Weglassen von Fragetypen

Auch wenn mit diesem Konverter eine vollständige Umsetzung der LimeSurvey Archiv-Daten angestrebt wird, so wurden doch einige Fragetypen bewusst nicht umgesetzt. Im Folgenden soll erläutert werden, wie diese Fragetypen hätten umgesetzt werden können und warum die Entscheidung getroffen wurde, dies nicht zu tun.

4.1.1 Datei-Upload

Der Fragetyp *'Datei-Upload'* kann genutzt werden, um den Nutzer auf eine Frage mit einer Datei antworten zu lassen. In ODM hätte man diese Datei einbinden können, indem man sie in `hexBinary` umwandelt, ein Datentyp in ODM, welcher Stream-Daten in einem hexacodierten Binärformat sammelt. Trotzdem wurde dieser Fragentyp im Konverter nicht umgesetzt. Das liegt zum einen daran, dass die hochgeladenen Dateien nicht Teil des Archives sind (s. Abschnitt 2.2.3) und andererseits daran, dass die Praktikabilität dieses Vorgehens eher fragwürdig ist. Unter anderem wird eine Rekonstruktion zur Originaldatei schwer, da es zum Beispiel keine Informationen über den ursprünglichen Dateityp gibt, auch wird die XML-Datei nur noch sehr unangenehm von Hand lesbar, wenn man diese Stream-Daten in Antworten einbinden würde. Auch wird der Fragentyp nicht häufig genutzt, was die Umsetzung noch unattraktiver macht.

5 Fazit

Dieses Kapitel bildet die abschließende Zusammenfassung der Arbeit. Dazu können die folgende Punkte behandelt werden:

- Reflexion: wurden die Ziele der Arbeit erreicht?
- mögliche Erweiterungen und Verbesserungen („future work“)

Eidesstattliche Erklärung

Hiermit versichere ich, dass die vorliegende Arbeit über „*Titel*“ selbstständig verfasst worden ist, dass keine anderen Quellen und Hilfsmittel als die angegebenen benutzt worden sind und dass die Stellen der Arbeit, die anderen Werken – auch elektronischen Medien – dem Wortlaut oder Sinn nach entnommen wurden, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht worden sind.

Vorname Nachname, Münster, 14. Mai 2021

Ich erkläre mich mit einem Abgleich der Arbeit mit anderen Texten zwecks Auffindung von Übereinstimmungen sowie mit einer zu diesem Zweck vorzunehmenden Speicherung der Arbeit in eine Datenbank einverstanden.

Vorname Nachname, Münster, 14. Mai 2021