

Der OSSE-Brückenkopf

Autoren Dr. Marita Muscholl und Dr. Martin Lablans
Universitätsmedizin der Johannes Gutenberg-Universität Mainz
Institut für Medizinische Biometrie, Epidemiologie und Informatik
55101 Mainz

08.Juni 2016

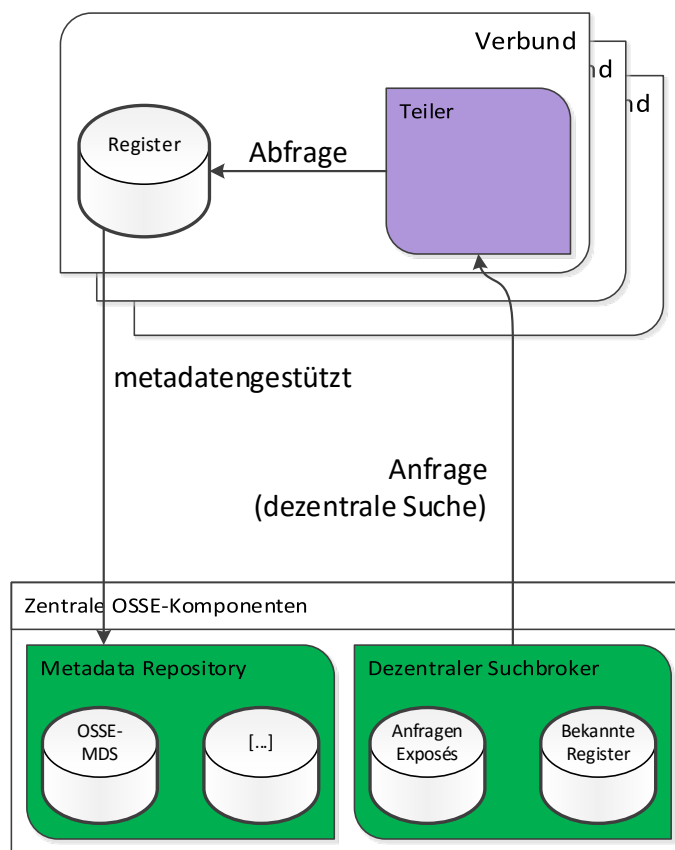
Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages

1 DAS PRINZIP DER DEZENTRALEN SUCHE

Ein wesentliches Prinzip der OSSE-Architektur ist die Vernetzung der Register auf der Basis einer verteilten Suchfunktion, der „dezentralen Suche“¹. Hintergrund dieser Architekturentscheidung ist die Überzeugung, dass die erfassten Patientendaten im jeweiligen Register verbleiben und nicht ohne konkrete Forschungsfragestellung in große, auch länderübergreifende Datensammlungen dupliziert werden sollten. Die Akzeptanz für solche Datensammlungen schätzen wir sowohl bei Patienten als auch bei den Ärzten, die diese Daten zu Forschungszwecken erfasst haben, als eher gering ein. Anders als bei anderen föderierten Suchen erfahren Dritte nichts über den Inhalt teilnehmender Datenbesitzer. Diese genießen also verglichen mit bisherigen Vernetzungsstrategien einen höheren Grad an Datenhoheit und die Daten einen höheren Grad an Datenschutz.



In OSSE wird die dezentrale Suche als Infrastruktur verwendet, die es erlaubt spezifische Fälle in den teilnehmenden Registern aufzufinden und dabei sowohl den Datenschutz als auch die Datenhoheit der Datenbesitzer zu berücksichtigen. Die Infrastruktur besteht aus einem zentralen Dienst zur Formulierung von Suchanfragen, dem *Suchbroker*, und Schnittstellen-Komponenten, den sogenannten *Teilern*, die lokal für jede Datenquelle (Register, Biobank etc.) existieren.

¹ Eine genaue Beschreibung und eine Abgrenzung zu anderen föderierten Suchmethoden findet sich in Lablans M, Kadioglu D, Muscholl M, Ückert F. Exploiting distributed, heterogeneous and sensitive data stocks while maintaining the owner's data sovereignty. *Methods Inf Med.* (in print).

Die dezentrale Suche beinhaltet folgende Schritte:

1. (einmalig:) Der Teiler des jeweiligen Registers registriert sich am Suchbroker. Außer den Kontaktdaten des Teilers werden dabei keine Daten übermittelt.
2. Ein Forscher ruft die Webseite des Suchbrokers auf und definiert die Suchkriterien seiner Anfrage. Dazu verwendet er Datenelemente, die im Metadaten-Repository (MDR) hinterlegt sind. Er gibt außerdem seine Kontaktdaten an und fügt ein Exposé seines Forschungsvorhabens hinzu.
3. Die Anfrage wird an alle registrierten Teiler weitergeleitet.
4. Jeder registrierte Teiler führt die Abfrage auf dem Datenbestand seines Registers durch. Wenn passende Datensätze gefunden werden, wird die Anfrage (Suchkriterien, Exposé und Kontaktdaten) zusammen mit dem Ergebnis der Abfrage einem Verantwortlichen des Registers vorgelegt. Andernfalls bleibt er von der Anfrage unbehelligt.
5. Der Dateneigentümer hat nun die Möglichkeit zu entscheiden, ob er zu dem Forschungsvorhaben beitragen und seine Daten (je nach Bedarf anonymisiert, pseudonymisiert oder auch in aggregierter Form und unter Einholung ggf. erforderlicher Einwilligungen) für das Forschungsvorhaben zur Verfügung stellen möchte. Er nimmt dazu direkt mit dem Anfragenden Kontakt auf. Entscheidet er sich dagegen, wird für den Anfragenden nicht ersichtlich, ob keine passenden Datensätze gefunden oder eine Teilnahme abgelehnt wurde.

Auf diese Weise werden Daten nur zweckgebunden für spezifische Fragestellungen (entsprechend den Datenschutzvorgaben und etwaigen Besitzrechten an den Daten) weitergegeben.

Jedes Register, das mit dem OSSE-Registerbaukasten realisiert wurde, beinhaltet bereits den Teiler und alle notwendigen Funktionen zur Teilnahme an der dezentralen Suche. Um auch Registern, die mit einer beliebigen anderen Software betrieben werden, die Vernetzung auf diesem Wege zu ermöglichen, kann dort ein sogenannter *OSSE-Brückenkopf* zusätzlich zur bestehenden Software installiert werden.

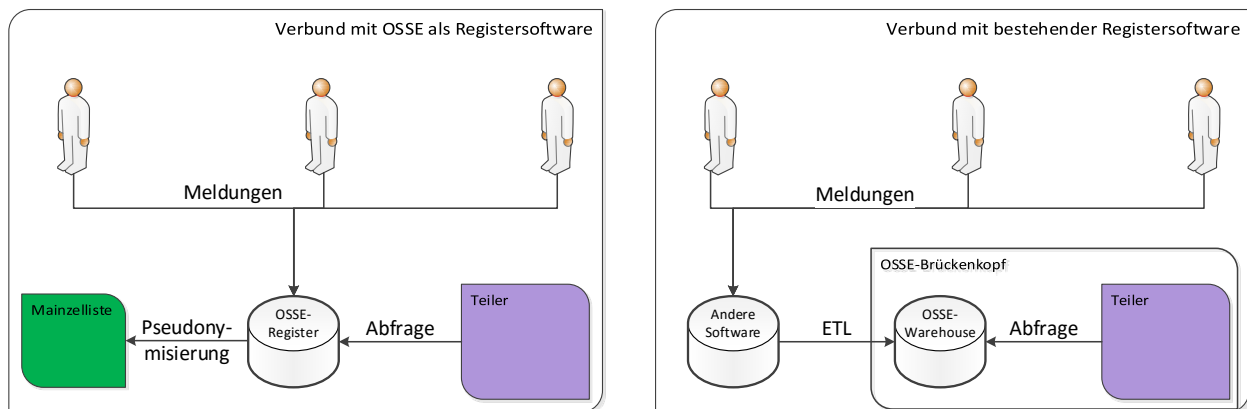


Abb. 2: Vernetzungsvarianten: mit OSSE-Register (links) oder OSSE-Brückenkopf (rechts)

2 WIE FUNKTIONIERT EIN OSSE-BRÜCKENKOPF?

Der OSSE-Brückenkopf besteht aus einer Art lokalem Data Warehouse, dem *OSSE-Warehouse*, und dem Teiler. Bei der Konfiguration des Warehouse werden Datenelemente im MDR ausgewählt oder neu angelegt und vergleichbar der Definition der Formulare für ein OSSE-Register in einem Stammdatenformular und einem Verlaufsdatenformular zusammengestellt. Der Formular-Editor erzeugt

das zugehörige XML-Schema (.xsd) samt XML-Beispiel, das als Vorlage für den Import dient. Für die regelmäßige Übertragung der Daten kann ein ETL-Prozess (Extraction-Transformation-Load) implementiert werden. Datenintegrations-Werkzeuge wie z.B. *Talend Open Studio* unterstützen die Datenübertragung mit einer Reihe von vorgefertigten und konfigurierbaren Komponenten in den folgenden Schritten:

1. **Extraktion** der Daten aus dem Register: Talend Open Studio ermöglicht die Extraktion von Daten durch SQL-Abfragen direkt aus der Register-Datenbank, aber auch die Verarbeitung unterschiedlicher dateibasierter Formate (CSV, XML), die von einer bereits vorhandenen Exportschnittstelle generiert werden können.
2. **Transformation** der Daten in die im Zielsystem benötigte Form: Im Transformationsschritt werden die exportierten Datenelemente aus dem Register in das oben generierte XML-Format transformiert und dabei die MDR-Keys des im Brückenkopf hinterlegten Datenschemas zugewiesen. Dazu kann das XML-Schema aus dem Konfigurationsschritt verwendet werden. Außerdem können in diesem Schritt bei Bedarf auch die Daten aus dem Register pseudonymisiert (oder ggf. Pseudonyme ausgetauscht) werden. Wir stellen dazu eine eigenentwickelte Komponente für Talend Open Studio zur Verfügung, die die Pseudonymisierung der Daten über unsere Identitätsmanagement-Software Mainzliste (oder eine schnittstellenkompatible Lösung²) ermöglicht.
3. **Laden** der Daten in den Brückenkopf: Der OSSE-Brückenkopf verfügt über eine Update-Schnittstelle, über die Daten erfasst, geändert oder gelöscht werden können. Die Daten werden in einem XML-basierten Format an eine REST-Schnittstelle übergeben.

Wird das vorhandene Register mit einer Schnittstelle ausgerüstet, die die notwendigen Voraussetzungen erfüllt (Extrahieren und Transformieren der Daten sowie Aufruf der Rest-Schnittstelle), entfällt die Verwendung eines Datenintegrationswerkzeugs.

Der Teiler des Brückenkopfes wird bei einem Suchbroker registriert und holt sich regelmäßig die vorhandenen Anfragen ab.

3 WAS BRAUCHE ICH UM EINEN OSSE-BRÜCKENKOPF ZU BETREIBEN?

3.1 HARDWARE

Für den OSSE-Brückenkopf wird ein (virtualisierter) Server benötigt. Der Speicherbedarf hängt vom Umfang des Inhalts ab; unsere Entwicklungsserver laufen schon ab 2GB RAM zuverlässig.

3.2 SOFTWARE

- OSSE-Brückenkopf-Software als Download unter <http://osse-register.de>.
- ggf. Talend Open Studio: Download unter <https://de.talend.com/products/data-integration>.

² Zur Schnittstelle vgl. a) zur Implementierung: die Dokumentation unter <http://www.mainzliste.de> und b) zur Rationale: die zugehörige Publikation Lablans M, Borg A, Ückert F. A RESTful interface to pseudonymization services in modern web applications. BMC Med Inform Decis Mak. 2015 Feb 7;15(1):2.

- Talend-ETL-Module für OSSE. als Download unter <http://osse-register.de>.

3.3 VORAUSSETZUNGEN AN DEN DATENEXPORT AUS DEM EXISTIERENDEN REGISTER

3.3.1 Updatesystematik

Für den Datentransfer an den Brückenkopf muss die Möglichkeit bestehen, regelmäßig Daten aus dem existierenden Register zu extrahieren. Dabei ist zu beachten, dass die Import-Schnittstelle des Brückenkopfes die Daten patientenweise verarbeitet, wobei grundsätzlich vollständige Patientendatensätze erwartet werden. Jeder Export bzw. Extrakt enthält und überschreibt also vollständige Datensätze von Patienten, die sich seit dem letzten Update geändert haben. Außerdem können Informationen über gelöschte Patienten verarbeitet werden. Werden im bestehenden Register synonyme Patientendatensätze zusammengeführt, sollte das nächste Update den Datensatz mit dem aktuellen Pseudonym aktualisieren und das stillgelegte Pseudonym löschen. Der Brückenkopf speichert keine Informationen über die Zuordnungshistorie. Exportiert das bestehende Register Datensätze, die identifizierende Daten enthalten, tritt dieses Problem nicht auf, da das Rekord-Linkage in der mitinstallierten Mainzelle stattfindet.

3.3.2 Struktur der Inhalte

Folgende inhaltliche Struktur wird im OSSE-Warehouse umgesetzt:

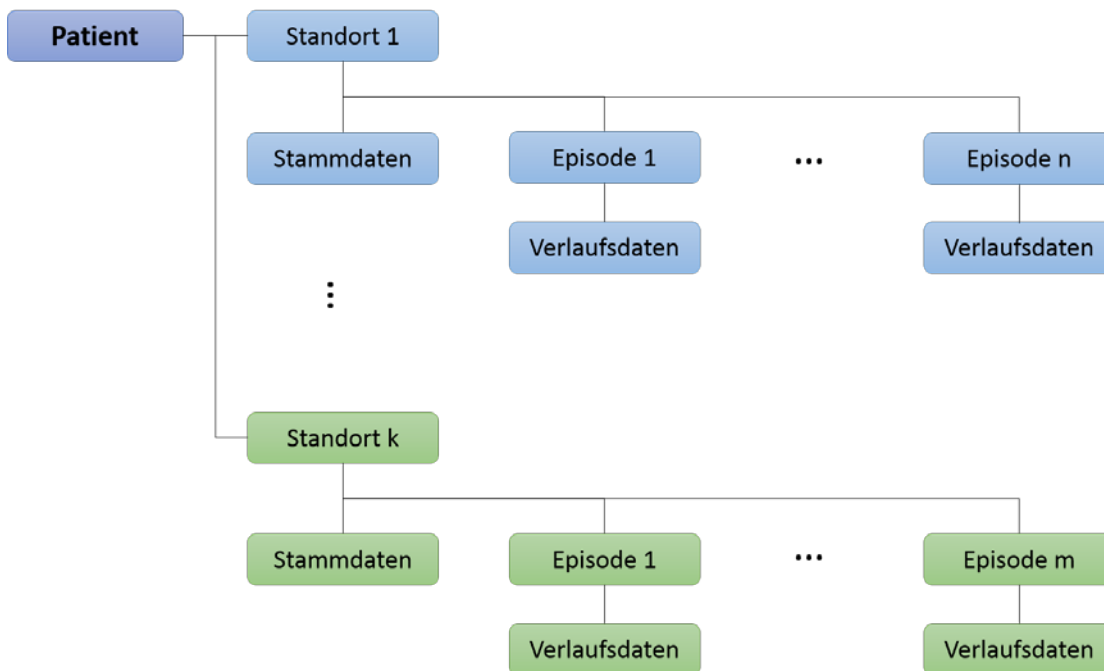


Abb. 3: Struktur des OSSE-Warehouse

Jeder Patientendatensatz kann Stammdaten und/oder Verlaufsdaten beinhalten. Zusätzlich werden Daten unterschieden, die von verschiedenen Einrichtungen (Standorten) erfasst wurden (z.B. falls der

Patient in verschiedenen der beteiligten Zentren behandelt wurde). Deshalb erwartet die OSSE-Importschnittstelle Daten eines Patienten sortiert nach Standorten (<Location>) und darunter aufgeteilt in Stammdaten (<Data type="basic">), die nur einmal pro Standort existieren, und Verlaufsdaten (<Data type="longitudinal">), die in jeder Episode (<Episode>) wiederholt werden.

Folgende Datenstrukturen sind möglich:

Elementare Datenelemente (<Field>) bestehen aus der eindeutigen ID des Datenelements im MDR (MDR-Key) und seinem Wert.

Wiederholbare Datenelemente (<RepeatableField>) bestehen aus dem MDR-Key des Datenelements und einer Liste von Werten.

Zusammengesetzte Datenelemente (<Record>) gruppieren mehrere zusammengehörige elementare Datenelemente.

Wiederholbare zusammengesetzte Datenelemente (<RepeatableRecord>) bestehen aus einer Liste zusammengesetzter Datenelemente.

3.4 DATENSCHUTZKONZEPT

Eine Schablone, mit der Sie das Datenschutzkonzept Ihres bestehenden Registers um die Brückenkopf-spezifischen Beschreibungen ergänzen können, steht unter <http://osse-register.de> zum Download bereit.

4 DIE IMPORTSCHNITTSTELLE

Für den Datenimport in das OSSE-Warehouse ist eine REST-Schnittstelle vorgesehen, die sowohl den Upload und das Löschen einzelner Patientendatensätze als auch eine Batch-Verarbeitung unterstützt.

4.1 UPLOAD VON PATIENTENDATENSÄTZEN

Ein Upload wird durchgeführt, wenn der Datensatz neu angelegt **oder** aktualisiert werden soll. Der Datensatz muss alle Daten des Patienten im existierenden Register enthalten, weil kein Delta-Update auf Teilen des Datensatzes (z. B. nur Episoden) vorgesehen ist. Werden Daten aus weiteren Systemen importiert, so wird dafür eine eigene Schnittstelle konfiguriert, die davon unabhängig ist, und deren Daten nur über das Pseudonym verknüpft werden.

Der Aufruf für den Upload von Patientendatensatzes lautet:

PUT /api/bridgehead/patients/

- Content Type = "application/xml"
- Im Header des Aufrufs werden die Credentials (i.a. ein Passwort – für die Authentisierung des Registers am Brückenkopf) mit dem Schlüssel "api key" übermittelt.
- Im Body werden ein oder mehrere Patientendatensätze im XML-Format übermittelt (Definitionen und Beispiele siehe Abschnitte 5 und 6).

4.2 DELETE EINES PATIENTENDATENSATZES

Aufruf: `DELETE /api/bridgehead/patients/{identifier}`

Parameter wie oben. Es wird kein Datensatz mitgeschickt.

5 SCHEMA-DEFINITIONEN

Das für Ihren Brückenkopf gültige XML-Schema können Sie jederzeit über die Konfigurationsoberfläche (unter Forms) mittels eines Klicks auf den Button Download XSD herunterladen. Dafür müssen jedoch die zu verwendenden Formulare aus dem Form-Repository importiert worden sein.