Anforderungskatalog

Generalisierte Schnittstelle zur Personenregistrierung für Merge'n'Dice

Stand 19. Aug. 2025

Anmerkung: Dieses Dokument entstand im Projekt Merge´n´Dice gefördert durch VDI VDE IT im Rahmen der Förderung „Datentreuhandmodelle“ (Förderkennzeichen 16DTM231)

Inhalt

Einleitung 1

Funktionsanforderungen 2

Anforderungen an externe Schnittstellen 3

Nicht-funktionale Anforderungen 3

Glossar 4

Einleitung

Gegenstand dieses Anforderungskatalogs ist die Spezifikation einer generalisierten Schnittstelle zur Registrierung von Personen vorrangig über eine föderierte Treuhandstelle/Vertrauensstelle. Die Generalisierung kann jedoch auch in anderen Treuhandstellen-Modellen (z.B. lokal oder zentral) umgesetzt werden. Die Anforderungen werden anhand des föderierten Ansatzes überprüft.

Für die Beantwortung komplexer medizinischer Forschungsfragen sollen die dezentralen Datenbestände von Universitätskliniken, Krankenkassen, Registern etc. verfügbar gemacht werden. Ein übergreifendes Record Linkage stellt dabei sicher, dass Daten zu einer Person zusammengeführt werden, auch wenn diese an verschiedenen Standorten vorliegen. Dabei gilt es, unterschiedliche Rechtsgrundlagen und datenschutzrechtliche Regularien zu berücksichtigen. Für medizinische Forschungsvorhaben stand lange Zeit kein einheitlicher Personenidentifikator zur Verfügung, der für diesen Zweck verwendet werden durfte. Daher wurden Konzepte entwickelt, mit denen identifizierende Daten in codierter Form abgeglichen werden konnten. Diese für spezifische Projekte anwendbaren Schnittstellen sollen nun dahingehend generalisiert werden, dass sowohl codierte (Bloomfilter oder anderweitig verschlüsselte Daten) als auch im Klartext vorliegende identifizierende Daten und Identifikatoren abgeglichen werden können.

Diese breite, generalisierte Schnittstelle ist für diverse Forschungsvorhaben relevant, da diese eine Vereinheitlichung im Bereich von Treuhandstellen-/Vertrauensstellen-Anbindungen ermöglicht. So können Abgleiche diverser Datenquellen für verschiedene Vorhaben über einheitliche Schnittstellen hinweg durchgeführt werden. Dies ist nicht nur für instituts- und einrichtungs-übergreifende, sondern künftig auch für länderübergreifende Abgleiche relevant. Hierbei besteht das Potential, vorhandene Lücken in der Standardisierung zu schließen, insbesondere im Hinblick darauf, künftige Vorhaben wie den European Data Health Space (EHDS) oder nationale Rechtsgrundlagen wie das Gesundheitsdatennutzungsgesetz (GDNG) zu adressieren.

Die Schnittstelle gibt dabei nur die notwendigen Parameter vor. Das Projekt legt den jeweiligen Kontext fest, wodurch die konkrete Umsetzung von Prozessen und weiteren Anforderungen durch die

jeweils eingesetzten Werkzeuge erfolgt. Die Schnittstelle gibt nicht die einzusetzenden Werkzeuge vor. Der Kontext bestimmt, welche Daten letztendlich für den Abgleich verwendet werden.

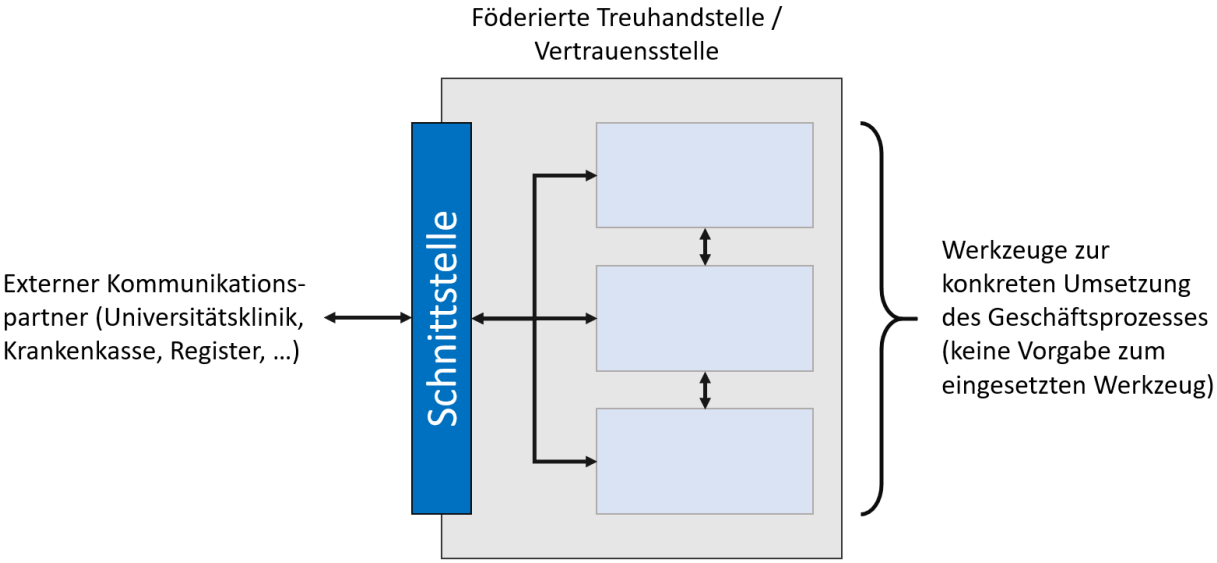

Abbildung 1: Die Schnittstelle ist Gegenstand dieses Anforderungskatalogs. Diese ist zentraler Bestandteil für eine einheitliche Anbindung an föderierte Treuhandstellen/Vertrauensstellen.

Eine Referenzimplementierung durch die Unabhängige Treuhandstelle der Universitätsmedizin Greifswald wird mit den dort eingesetzten Werkzeugen (E-PIX, gPAS, Dispatcher) angestrebt. Dadurch soll die Umsetzung der Schnittstelle praktisch gezeigt werden.

Funktionsanforderungen

FA-1: Die Schnittstelle ermöglicht die Personenregistrierung mittels eines eineindeutigen *Identifikators*, darunter auch den zehnstelligen, unveränderlichen Teil der Krankenversichertennummer nach § 290 Absatz 1 Satz 2 SGB V.

Im Fall von bereits vorhandenen *Identifikatoren* sollen diese für einen (in der Regel) exakten Abgleich verwendet werden können. Es sind beliebige *Identifikatoren* denkbar (z.B. *Pseudonyme*, die bereits in anderen Kontexten vergeben wurden, wie die Krankenversichertennummer).

FA-2: Die Schnittstelle ermöglicht die Personenregistrierung mittels codierter, identifizierender Daten in Form von *Bloomfiltern*. Es werden *Cryptographic Longterm Keys*, *Record-Level Bloomfilter* und *Field-Level Bloomfilter* berücksichtigt.

Die *Bloomfilter* werden bei den Datengebern erzeugt und an die abgleichende Treuhand-/Vertrauensstelle übermittelt. Der Abgleich von *Bloomfiltern* ist dabei unabhängig vom Verfahren (und ggf. weiteren Härtungsmaßnahmen). Die Schnittstelle berücksichtigt die Registrierung eines oder mehrerer *Bloomfilter*. *Field-Level Bloomfilter* werden pro Attribut erzeugt, sodass beliebig viele *Bloomfilter* pro Registrierung unterstützt werden müssen. Aktuelle Verfahren wie *Cryptographic Longterm Keys* und *Record-Level Bloomfilter* werden ebenfalls unterstützt. Diese erfordern jedoch nur die Übermittlung eines *Bloomfilters*, welcher mehrere codierte Attribute aufweist.

FA-3: Die Schnittstelle ermöglicht die Personenregistrierung mittels *identifizierender Daten/Attribute*.

Ein Abgleich von *identifizierenden Attributen* ist auch in einem föderierten System denkbar. Die Funktionalität, *identifizierende Attribute* entgegenzunehmen, wird im föderierten Modell zumindest bei der Implementierung einer Clearingstelle sowie darüber hinaus in lokalen und zentralen

Treuhandstellen benötigt. In der Clearingstelle werden *identifizierende Daten* nur temporär vorgehalten, was jedoch nicht von der Schnittstelle bestimmt wird.

FA-4: Die Schnittstelle ermöglicht die Angabe eines oder mehrerer Kontexte, in denen die Person pseudonymisiert wird.

Je nach Projekt soll es möglich sein, den jeweiligen Kontext anzugeben, in dem pseudonymisiert werden soll. Dabei kann es erforderlich sein, dass *Pseudonyme* verschiedener Kontexte angelegt und/oder geliefert (vgl. FA-5) werden sollen. Es ist daher erforderlich, dass der Schnittstelle ein oder mehrere Kontexte, z.B. unterschiedliche Leistungserbringer/Abteilungen, Bioproben und/oder Medizinprodukte, übermittelt werden können.

FA-5: Die Schnittstelle liefert im Erfolgsfall ein oder mehrere *Pseudonyme* für einen oder mehrere Kontexte zurück.

Je nach Projekt und Kontext werden üblicherweise *Pseudonyme* vergeben, wobei die tatsächliche Pseudonym-Hierarchie unabhängig der Schnittstelle ist. Meistens wird bei verschiedenen Kontexten (vgl. FA-4) je anfragenden bzw. datengebenden Standort ein *Pseudonym* zurückgeliefert. In bestimmten Kontexten kann die Vergabe und Rückgabe mehrerer *Pseudonyme* erforderlich sein (*Pseudonyme* für unterschiedliche Kontexte wie unterschiedliche Leistungserbringer/Abteilungen, Bioproben und/oder Medizinprodukte).

Anforderungen an externe Schnittstellen

AES-1: Der anfragende bzw. datengebende Standort muss projektspezifische Rückgaben entgegennehmen und interpretieren können.

Da die Rückgabe projektspezifisch erfolgt, kann der Umfang der Rückgabe variieren (z.B. Anzahl gelieferter Pseudonyme, verschiedene Kontexte). Diese Rückgabe muss vom Standort interpretiert und korrekt verarbeitet werden können (inkl. Anbindung der jeweils eingesetzten Werkzeuge).

AES-2: Der anfragende bzw. datengebende Standort unterstützt die projektspezifische Anbindung an die Schnittstelle.

Die Schnittstelle definiert nicht die konkrete Anbindung (NFA-6). Der Standort muss die projektspezifischen Vorgaben zur Absicherung des Transportwegs und die Anforderungen zur Authentifizierung umsetzen und dahingehend die Schnittstelle ansprechen.

AES-3: Der anfragende bzw. datengebende Standort muss die Schnittstelle aufrufen können.

Die Schnittstelle wird als Referenzimplementierung im *FHIR*-Standard bereitgestellt. Davon abweichende Implementierungen müssen vom Standort in seiner Rolle als Client aufrufbar sein. Die konkrete Implementierung findet dann ggf. am Standort als externe Schnittstelle statt.

Nicht-funktionale Anforderungen

NFA-1: Ein *Bloomfilter* wird als *Base64*-codierte Zeichenkette übermittelt.

Bloomfilter sollen als binäres Datum im Speziellem berücksichtigt werden. Die einheitliche Codierung im *Base64*-Format ermöglicht die Übermittlung binärer Daten und ggf. die einheitliche Decodierung.

NFA-2: *Identifikatoren* oder Referenzen (*Pseudonyme*) werden textbasiert übermittelt.

Identifikatoren oder anderweitig vergebene Referenzen wie *Pseudonyme* können textbasiert übertragen werden. Üblicherweise enthalten die Alphabete Buchstaben (Groß- und Kleinbuchstaben), Zahlen und Sonderzeichen wie Binde- oder Unterstriche.

NFA-3: Identifizierende Daten/Attribute werden textbasiert übermittelt.

Identifizierende Daten wie Vorname, Nachname etc. werden textbasiert übermittelt.

NFA-4: Die Schnittstelle wird u.a. in *FHIR* als Referenzimplementierung abgebildet.

Die Schnittstelle wird unabhängig der Notation entwickelt. Da *FHIR* im Gesundheitswesen ein vielgenutzter Standard ist, erfolgt eine Referenzimplementierung in *FHIR*. Da je nach Kontext auch andere Bereiche diese Schnittstelle verwenden können, kann eine Implementierung losgelöst von der jeweiligen Notation erfolgen.

NFA-5: Die Authentifizierung und Autorisierung finden unabhängig von der Schnittstelle statt.

Die Schnittstelle gibt keine Vorgaben zur Authentifizierung oder Autorisierung vor. Dies obliegt der konkreten Umsetzung im Projekt, inkl. der dort eingesetzten Werkzeuge. Die Schnittstelle soll dies möglichst nicht einschränken. So kann die Authentifizierung beispielsweise mittels *API-Key* über den HTTP-Header erfolgen.

NFA-6: Die Absicherung der Transportwege findet unabhängig der Schnittstelle statt.

Die Schnittstelle gibt keine Vorgaben zur Absicherung von Transportwegen vor und soll möglichst keine Einschränkungen machen. Übliche Verschlüsselungstechniken wie HTTPS zur Absicherung der Datenübertragung werden berücksichtigt.

NFA-7: Die Absicherung der Datenhaltung findet unabhängig der Schnittstelle statt.

Die Absicherung der Datenhaltung obliegt der konkreten projektspezifischen Umsetzung und wird durch die Schnittstelle möglichst nicht eingeschränkt. Übliche Absicherungsmechanismen wie die Verschlüsselung von Datenträgern bleiben durch die Schnittstelle unberührt.

NFA-8: Die Schnittstelle berücksichtigt vorhergehende Arbeiten zu Pseudonymisierungsschnittstellen im Rahmen der Medizininformatik-Initiative.

In der Medizininformatik-Initiative wurden bereits Konzepte zur Pseudonymisierung erarbeitet, welche mitunter die Anbindung von Treuhand-/Vertrauensstellen bzw. dort eingesetzten Werkzeugen (zur Pseudonymisierung) vorsehen. Diese Vorarbeiten[1] (insbesondere die Use Cases 1, 3a und 3b) werden berücksichtigt.

Glossar

API-Key

Ein API-Key ist eine meist anwenderspezifische (z.B. je Benutzer oder Standort) Kennung, welche die Nutzung einer Schnittstelle authentifiziert. Diese Kennung ist nur dem Anwender bekannt.

[1] Implementation Guide MII-Schnittstelle Pseudonymisierung Version 1.0.2 (2025-04-08), https://medizininformatik-initiative.github.io/mii-interface-module-pseudonymization/, Zugriff am 2025-04-29

Base64

Base64 ist ein Format, welches zur textuellen Übertragung von Binärinformationen verwendet wird. Dabei umfasst das Alphabet von Base64 entsprechend 64 Zeichen, bestehend aus A-Z, a-z, 0-9, + und / (zusätzlich = als Füllzeichen am Ende).

Bloomfilter

Ein Bloomfilter ist ein initial mit Nullen besetzter Bit-Vektor. Mittels Umrechnungsfunktionen werden einzelne oder mehrere Datenattribute derart umgerechnet, dass diese im Bloomfilter i.d.R. mehrere Positionen auf Eins setzen. Bloomfilter können miteinander verglichen werden, wobei hohe Ähnlichkeiten in den Ausgangsdaten auch zu hohen Ähnlichkeiten im Bloomfilter führen. Der Bloomfilter selbst lässt kaum oder keine Rückschlüsse auf die Ausgangsdaten zu.

Cryptographic Longterm Key

Ein Cryptographic Longterm Key ist ein Bloomfilter, bei dem mehrere Datenattribute (z.B. Vorname, Nachname) im selben Bloomfilter codiert wurden.

Field-Level Bloomfilter

Ein Field-Level Bloomfilter ist ein Bloomfilter, in dem nur ein Attribut codiert wurde. Üblicherweise werden für einen Abgleich mehrere solcher Bloomfilter verwendet, welche in Summe die Entität widerspiegeln. So können z.B. einzelne Bloomfilter erzeugt werden, die dann als Ausgangsdaten den Vornamen, den Nachnamen und das Geburtsdatum codieren.

FHIR (Fast Healthcare Interoperability Resource)

FHIR ist ein Standard zum Datenaustausch, welcher Elemente als sogenannte Ressourcen beschreibt. Die Repräsentation findet im XML oder JSON Format statt.

Identifikator / Personenidentifikator

Ein Personenidentifikator ist ein eineindeutiges Merkmal oder Kennung, das/die eineindeutig eine Person referenziert. Beispiele hierfür sind die Krankenversichertennummer, projektspezifisch eindeutig vergebene Pseudonyme oder anderweitig eineindeutige Kennungen.

Identifizierende Daten / Attribute

Identifizierende Daten oder Attribute sind Eigenschaften, die eine Person identifizieren. Dies umfasst oft sogenannte Quasi-Identifikatoren, welche allein für sich nicht identifizierend sind, jedoch in Summe identifizierend sein können. Üblicherweise sind dies Vornamen, Nachnamen, Geburtsdatum, Geschlecht, Adressangaben und ähnliche persönliche Daten.

Pseudonym

Ein Pseudonym ist eine eindeutige Kennung, die die Identität einer Person verschleiert. Das Pseudonym dient als Referenzierung einer Person. Ein Pseudonym kann dabei eine kontextfreie, zufällig gewählte Kennung sein, die nur anhand von Zuordnungstabellen wieder aufgelöst werden kann. Alternativ können Pseudonyme das Ergebnis von verschlüsselten Daten sein.

Record-Level Bloomfilter

Ein Record-Level Bloomfilter ist ein Bloomfilter, in dem mehrere Datenattribute codiert wurden (vgl. Cryptographic Longterm Key). Dabei werden die einzelnen Datenattribute jedoch unterschiedlich gewichtet, sodass Datenattribute unterschiedlich viele Positionen im Bloomfilter setzen. Durch diese

unterschiedliche Gewichtung besteht das Resultat beim Abgleich aus unterschiedlich großen Anteilen der Datenattribute.