



Institut für Qualitätssicherung und
Transparenz im Gesundheitswesen

Verknüpfung der Module Geburtshilfe und Neonatologie des QS-Verfahrens Perinatalmedizin

Technische Dokumentation zur Umsetzung der
Pseudonymisierung der PID-Daten in der Vertrauensstelle

Erstellt im Auftrag des
Gemeinsamen Bundesausschusses

Stand: 30. Juni 2023

Impressum

Thema:

Technische Dokumentation zur Umsetzung der Pseudonymisierung der PID-Daten für die Module Geburtshilfe und Neonatologie des QS-Verfahrens Perinatalmedizin in der Vertrauensstelle

Autoren:

Dr. Jens Meier, Tobias Jakscha, Prof. Dr. Rainer Schnell, PD Dr. Günther Heller

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum der Veröffentlichung:

30. Juni 2023

Herausgeber:

IQTIG – Institut für Qualitätssicherung
und Transparenz im Gesundheitswesen

Katharina-Heinroth-Ufer 1
10787 Berlin

Telefon: (030) 58 58 26-0
Telefax: (030) 58 58 26-999

info@iqtig.org

<http://www.iqtig.org>

Inhaltsverzeichnis

Inhaltsverzeichnis.....	3
Abbildungsverzeichnis.....	5
Tabellenverzeichnis.....	5
1 Einleitung.....	6
1.1 Hintergrund.....	6
1.2 Zielsetzung und Zielgruppe	7
2 Grundlagen.....	8
2.1 Bloomfilterverfahren.....	8
2.2 Keyed-Hash Message Authentication Code (HMAC)	9
3 Pseudonymisierungsverfahren zur Verknüpfung der Datensätze Geburtshilfe und Neonatologie.....	10
3.1 Datengrundlage.....	10
3.2 Datenaufbereitung.....	10
3.2.1 Vorgaben der Spezifikation	11
3.2.2 Reduktion auf 3 Namensbestandteile.....	11
3.2.3 Kürzung der Namensbestandteile.....	11
3.2.4 Ergebnisse	11
3.3 Schnell-Bloomfilterverfahren.....	11
3.3.1 Bildung von Bigrammen	11
3.3.2 Pseudonymisierung der Namen	12
3.3.3 Pseudonymisierung des Geburtsdatums des Kindes	13
4 Prozesse.....	14
4.1 Leistungserbringer	14
4.1.1 XML-Struktur	15
4.1.2 Erläuterung.....	15
4.2 Datenannahmestellen auf Landesebene	16
4.3 Vertrauensstelle	16
4.3.1 Generierung der Hash-Schlüssel	16
4.3.2 Speicher- und Löschfristen.....	16
4.3.3 Entschlüsselung der PID-Informationen.....	17

4.3.4	Pseudonymisierung der PID-Daten	17
4.3.5	XML-Struktur zur Übermittlung der Pseudonyme an das IQTIG	17
4.3.6	Übermittlung der Pseudonyme an das IQTIG	18
4.4	Bundesauswertungsstelle	19
4.4.1	Speicher- und Löschrfristen	19
5	Literaturverzeichnis	20

Abbildungsverzeichnis

Abbildung 1: Beispiel der Anwendung eines 10-Bit-Bloomfilters.....	8
Abbildung 2: Prozesse beim Leistungserbringer.....	14
Abbildung 3: XML-Schema für das <perineo_pid> Element zur Übermittlung der PID-Daten zwischen LE und DAS	15
Abbildung 4: Prozesse in der DAS auf Landesebene.....	16
Abbildung 5: XML-Schema des <patient> Elementes nach Pseudonymisierung in der VST.....	17
Abbildung 6: XML-Schema des <bloomfilter> Elementes zur Übermittlung der bloomfilterspezifischen Pseudonyme.....	18

Tabellenverzeichnis

Tabelle 1: PID-Elemente zur Verknüpfung der Datensätze	10
--	----

1 Einleitung

1.1 Hintergrund

Die Weiterentwicklung der gesetzlichen medizinischen Qualitätssicherung hat unter anderem eine einrichtungsübergreifende Qualitätsanalyse zum Ziel¹. Dabei sollen qualitätsrelevante Dokumentationen (QS-Dokumentationen) eines Patienten aus unterschiedlichen Einrichtungen und Zeiten anhand eines Patientenpseudonyms miteinander verknüpft werden, ohne dass der Patient selbst identifiziert werden kann.

Dies gilt insbesondere für die seit langem angestrebte Verknüpfung der Perinatal- und Neonatalerhebung [1], die als ein unabdingbarer Baustein für die Weiterentwicklung der Qualitätssicherung in der Neugeborenenversorgung angesehen wird. Die Verknüpfung wird auch mit Blick auf die Ziele der Qualitätssicherungs-Richtlinie Früh- und Reifgeborene (QFR-RL)² dringend benötigt. Eine Verknüpfung der Module Geburtshilfe (Perinatalerhebung) und Neonatologie (Neonatalerhebung) ist sinnvoll, denn nur so kann der Behandlungsverlauf eines Kindes von der Geburt bis zur endgültigen Entlassung nach Hause nachvollzogen werden. Beispielsweise werden Kinder mit einem kritischen Outcome, welches in der Geburtshilfe erfasst wird, häufig auf die neonatologische Abteilung verlegt.

Für GKV-versicherte Patientinnen und Patienten wurde im Rahmen der Weiterentwicklung der Verfahren der externen stationären Qualitätssicherung ein Verfahren getestet, das eine Verknüpfung über eine pseudonymisierte Versichertennummer der Krankenversichertenkarte bzw. der lebenslangen Krankenversichertennummer ermöglicht. Dieses basiert auf den im Jahr 2009 mit dem BSI beratenen und für geeignet angesehenen Verfahren.

Die Evaluationen dieses auch für die Verknüpfung der Perinatal- und Neonatalerhebung erprobten Verfahrens mittels pseudonymisierter Versichertennummer [2] zeigten, dass analog zu Berichten aus der wissenschaftlichen Literatur, systematische Unterschiede zwischen verknüpften und nicht verknüpften Kindern bzw. Kind-Mutter-Paaren [3], [4], d. h. Unterschiede in Datensätzen aus der Perinatal- und Neonatalerhebung zum selben Kind, die derzeit noch nicht verknüpft werden, bestehen. Es ist demnach von einem „differential linkage error“ auszugehen, der die Ergebnisse von Qualitätsanalysen beeinflusst. Da statistische Qualitätsanalysen auf nichtlinearen Modellen basieren, können einzelne (seltene) Ereignisse die Ergebnisse in Hinsicht auf die Qualität einer Einrichtung gravierend verändern. Daher ist eine vollständige Verknüpfung aller Fälle für tragfähige Qualitätsanalysen und Qualitätsaussagen unabdingbar.

Darüber hinaus ergab sich, dass mit dem oben erwähnten Verfahren nur etwa 20% der Fälle verknüpft werden konnten. Die Ergebnisse der Auswertungen liefern auch deutliche Hinweise darauf, dass ein erheblicher Teil der potenziell verknüpfbaren Datensätze nicht über pseudonymisierte Versichertennummern zusammengeführt werden konnten. So fanden sich für 20% der Fälle, für die über weitere im Datensatz vorhandene Informationen eine Zuordnung möglich

¹ Vgl. z. B. §§ 136 ff. SGB V

² <https://www.g-ba.de/informationen/richtlinien/41/>

war, keine identischen PID-Felder [2]. Diese schlechten Ergebnisse lassen sich dadurch begründen, dass in einem relevanten Anteil der neonatologischen Aufenthalte die Versicherungsnummer der Mutter nicht verfügbar ist.

Die in Kapitel 3 beschriebenen zwei Verfahren schlagen zum einen eine Verknüpfung der Datensätze auf Basis von Bloom-Filtern und zum anderen die Anwendung des Verfahrens der epidemiologischen Krebsregister (Krebsregisterverfahren) vor, jeweils auf der Basis von sowohl in der Geburtshilfe als auch in der Neonatologie vorhandenen Datenfeldern zusammen mit dem Namen der Mutter.

1.2 Zielsetzung und Zielgruppe

Die Technische Dokumentation zur Verknüpfung der Peri/Neo-Daten beschreibt detailliert die Pseudonymisierungsverfahren sowie die Prozesse beim Leistungserbringer, in den Datenannahmestellen auf Landesebene, bei der Vertrauensstelle und beim IQTIG als Bundesauswertungsstelle (BAS). Das Dokument wurde basierend auf den Vorgaben des Gemeinsamen Bundesausschuss (G-BA), insbesondere des Fachausschuss QS-IT erstellt [5]. Die Pseudonymisierungsverfahren werden in einem technischen Detaillierungsgrad beschrieben, der es den jeweiligen Stellen erlaubt, diese in Software zu implementieren. Die Prozesse beschreiben, wie die Pseudonymisierungsverfahren angewendet werden und welche Datenflüsse zur Durchführung etabliert werden müssen.

Das Dokument richtet sich an die Leistungserbringer, Datenannahmestellen auf Landesebene und die unabhängige Vertrauensstelle (VST) nach § 299 SGB V.

2 Grundlagen

Der Abschnitt Grundlagen beschreibt detailliert die technischen Verfahren, die im Rahmen der Verknüpfung der Datensätze Perinatologie und Neonatologie Anwendung finden.

2.1 Bloomfilterverfahren

Bloom entwickelte ein Verfahren, welches es ermöglicht ressourcenschonend zu bestimmen, ob ein Testdatum in einer Menge von Datensätzen enthalten ist, und vergleicht dieses Verfahren mit konventionellen Hashing-Methoden [6]. Abhängig davon wie stark im Hinblick auf die Hashgröße optimiert wird, erhöht sich die erlaubte Fehlerrate (false-positive) des Verfahrens.

Bloom beschreibt in seinem Paper zwei unterschiedliche Methoden, wobei im Kontext der Verknüpfung Peri/Neo Methode 2 Anwendung findet und im Folgenden beschrieben wird. Bei dieser Methode wird der Hashing-Bereich als Array von N individuell adressierbaren Bits α angesehen. Alle Bits werden mit 0 initialisiert. Jede Nachricht, die in dem Array gespeichert werden soll, wird durch Hashing auf d unterschiedliche Bits ($\alpha_1, \alpha_2, \dots, \alpha_d$) abgebildet, wobei gilt: $d \in \{0, \dots, N - 1\}$. Der Wert der entsprechenden Bits α_d wird auf 1 gesetzt.

Beispiel:

Die Nachricht „Test“ soll in einem Bloomfilter gespeichert werden. Der Bloomfilter besteht aus einem 10-Bit-Array. Die Nachricht soll auf zwei Bit in dem Bloomfilter abgebildet werden. Es werden nun zwei Hashing-Funktionen $h_1(x)$ und $h_2(x)$ genutzt um die Nachricht auf zwei unterschiedliche Bits abzubilden. Die Abbildung ergibt folgendes Ergebnis:

$$h_1(\text{Test}) = \alpha_2, \quad h_2(\text{Test}) = \alpha_8$$

Der entsprechende Bloomfilter nach Anwendung der Hashfunktionen auf die Nachricht ist in Abbildung 1 dargestellt.

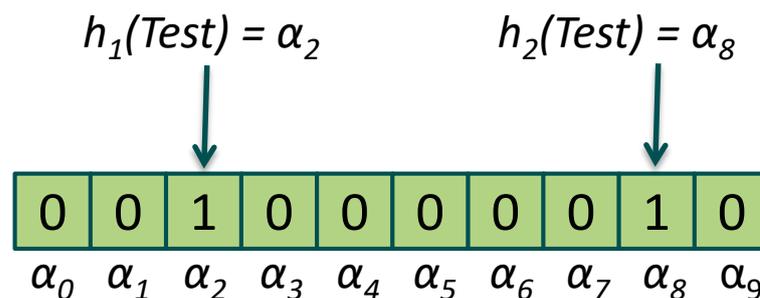


Abbildung 1: Beispiel der Anwendung eines 10-Bit-Bloomfilters

Um zu prüfen, ob eine neue Nachricht in einem Set von Nachrichten enthalten ist, muss diese mit den gleichen Hashfunktionen (in diesem Beispiel $h_1(x)$, $h_2(x)$) auf d Bits ($\alpha'_1, \alpha'_2, \dots, \alpha'_d$) abgebildet werden. Sind alle d Bits der neuen Nachricht im Array des Bloomfilters = 1 wird die Nachricht akzeptiert.

2.2 Keyed-Hash Message Authentication Code (HMAC)

Message Authentication Codes (MACs) [7] dienen dazu die Integrität einer Nachricht zu überprüfen. HMACs unterscheiden sich durch die Verwendung von kryptografischen Hash-Funktionen (MD5, SHA1, SHA2) und geheimen Passwörtern bei der Erzeugung der MACs. Der Sender einer Nachricht verschlüsselt diese mit Hilfe des HMAC—Algorithmus und des geheimen Schlüssels. Der Empfänger der Nachricht kann anschließend die Integrität der Nachricht durch die Erneute Anwendung des HMAC-Algorithmus und des geheimen Schlüssels auf die Nachricht selbst und den Vergleich mit dem erhaltenen HMAC prüfen. Die Identität des Senders wird implizit verifiziert, wenn der beim Empfänger erzeugte HMAC dem HMAC entspricht der vom Sender erzeugt wurde. Implementierungen des Algorithmus werden u. A. in OpenSSL³, Apache Commons⁴ oder BouncyCastle⁵ zur Verfügung gestellt.



Achtung

Im Rahmen der Peri/Neo-Verknüpfung ist der HMAC ausschließlich mit der Hashfunktion **SHA2** (SHA256) zu erstellen!

³ <https://www.openssl.org/>

⁴ <https://commons.apache.org/proper/commons-codec/apidocs/org/apache/commons/codecs/digest/HmacUtils.html>

⁵ <https://www.bouncycastle.org/java.html>

3 Pseudonymisierungsverfahren zur Verknüpfung der Datensätze Geburtshilfe und Neonatologie

3.1 Datengrundlage

Die Pseudonyme zur Verknüpfung der Datensätze werden von patientenidentifizierenden Daten aus den QS-Verfahren Geburtshilfe und Neonatologie generiert (siehe Tabelle 1).

Tabelle 1: PID-Elemente zur Verknüpfung der Datensätze

Datum	Technische Feldkennung in XML
Vorname der Mutter	vorname_mutter
Nachname der Mutter	nachname_mutter
Geburtsdatum des Kindes	GEBDATUMK

* wird nur bei den Datensätzen der Neonatologie erhoben

Die vorhandenen Informationen müssen anschließend in die für die Pseudonymisierung und das Record-Linkage notwendigen Datenfelder überführt werden:

- Vorname der Mutter
- Vorname der Mutter, Teil 2
- Vorname der Mutter, Teil 3
- Nachname der Mutter
- Nachname der Mutter, Teil 2
- Nachname der Mutter, Teil 3
- Geburtsdatum des Kindes

Die Datenfelder *phonetisch standardisierter Vorname* der Mutter und *phonetisch standardisierter Nachname der Mutter* werden ausschließlich zur Umsetzung des adaptierten Krebsregisterverfahrens benötigt, jedoch zur Gewährleistung einer Vergleichbarkeit beider Pseudonymisierungsverfahren während der Evaluationsphase auch im Schnell-Bloomfilterverfahren berücksichtigt. Im Fall der Nutzung des Schnell-Bloomfilterverfahrens im Regelbetrieb werden diese Datenfelder wieder entfernt.

3.2 Datenaufbereitung

Der Vorname der Mutter, der Nachname der Mutter sowie die Vorsatzworte liegen in jeweils einem Datenfeld vor. In diesen Datenfeldern können jeweils mehrere Vornamen bzw. mehrere Nachnamen und Vorsatzworte durch Leerzeichen getrennt in den Datenfeldern erfasst werden.

3.2.1 Vorgaben der Spezifikation

Die Spezifikation definiert bereits grundlegende Eigenschaften bei der Erfassung und dem Export des Vornamens sowie des Nachnamens der Mutter, die die Weiterverarbeitung der Daten in der VST erleichtert.

Die Informationen werden entsprechend der eGK-Versichertenkarte verarbeitet. Die eGK-Versichertenkarte enthält separate Datenfelder zu akademischen Titeln, Vorsatzworten und Namenszusätzen. Akademische Titel sowie Namenszusätze sollen laut Vorgabe nicht im Bloomfilter berücksichtigt werden, so dass diese Datenfelder der eGK nicht verarbeitet werden. Vorsatzworte werden aus dem entsprechenden Feld der eGK ausgelesen und im Exportfeld automatisch den Nachnamen der Mutter getrennt durch ein Leerzeichen angefügt. Die Daten liegen in den KIS-Systemen der Krankenhäuser codiert nach ISO 8859-15 vor. Vor dem Export der Felder für die Peri/Neo-Verknüpfung werden die Daten nach den Vorgaben in verarbeitet, so dass diese in der VST ohne weitere Bearbeitung pseudonymisiert werden können.

3.2.2 Reduktion auf 3 Namensbestandteile

Im Rahmen der Peri/Neo-Verknüpfung werden lediglich die ersten 3 Bestandteile des Vornamens bzw. des Nachnamens benötigt. Alle Namensbestandteile > 3 werden entfernt.

3.2.3 Kürzung der Namensbestandteile

Die noch vorhandenen Namensbestandteile werden aus datenschutzrechtlichen Gründen jeweils auf 10 Zeichen gekürzt (z. B. schnarrenberger → schnarrenb).

3.2.4 Ergebnisse

Nach der Datenaufbereitung liegt der Name (Vorname/Nachname) folgendermaßen vor:

Standardisiert mit maximal 3 Bestandteilen zu je maximal 10 Zeichen. Die Namensbestandteile sind jeweils durch ein Leerzeichen getrennt. Vorsatzworte befinden sich am Ende.

3.3 Schnell-Bloomfilterverfahren

Das Schnell-Bloomfilterverfahren wurde von Schnell et. al. als probabilistic record linkage Verfahren zur Verknüpfung von Datensätzen aus unterschiedlichen Datenbanken unter Berücksichtigung des Datenschutzes entwickelt [8]. Das Verfahren wurde anschließend durch den G-BA im Fachausschuss QS-IT (FA QS-IT) zur Verknüpfung der Peri/Neo-Datensätze angepasst [5]. Im Folgenden werden die Ausführungen aus dem FA QS-IT in Bezug auf die technische Umsetzung in der Vertrauensstelle detailliert.

3.3.1 Bildung von Bigrammen

Der standardisierte Name (jeweils Vor-/Nachname) aus der Datenaufbereitung (siehe Abschnitt 3.2.4) wird zur Durchführung der Pseudonymisierung in sogenannte Bigramme zerlegt. Ein Bigramm besteht immer aus zwei aufeinanderfolgenden Buchstaben des Namens. Es ist zu beachten, dass das erste Bigramm aus einem Unterstrich () gefolgt von dem ersten Buchstaben des

Namens bzw. Namensbestandteiles und das letzte Bigramm aus dem letzten Buchstaben und einem Unterstrich besteht.

Beispiel: Bigramm-Repräsentation eines Doppelnamens

maier schmidt → m ma ai ie er r_ s sc ch hm mi id dt t_

3.3.2 Pseudonymisierung der Namen

Die Pseudonymisierung im Schnell-Bloomfilterverfahren wird durch die Abbildung jedes einzelnen Bigramms des Vor- und Nachnamens in einen Bloomfilter unter Nutzung unterschiedlicher Hashfunktionen erreicht. Die Nutzung mehrerer verschiedener Hashfunktionen f_i zur Pseudonymisierung jedes einzelnen Bigramms stellt eine grundlegende Notwendigkeit dar, um mögliche Attacken zur Depseudonymisierung der Namensbestandteile zu verhindern [9]. Folgende Informationen sind für die Pseudonymisierung relevant:

- 4 x Hash-Schlüssel jeweils für 1 Erfassungsjahr key_{ID_EJ}
- Geburtsdatum des Kindes T in der Form dd.MM.YYYY
- Feld-ID (vorname_mutter bzw. nachname_mutter) ID
- Bigramm b
- Länge des Bloomfilters n
- Anzahl der zu nutzenden Hashfunktionen k



Achtung

Folgende Parameter sind bei der Umsetzung zu nutzen:

- Länge des Bloomfilters n : 1000
 - Anzahl Hashfunktionen k : 10
-

Hashfunktion

$\forall_{i=0}^k: f_i(b, key_{ID_EJ}, T, ID) = HMAC(key_{ID_EJ}, i + T + ID + b) \bmod n$ bildet ein Bigramm b mit Hilfe k unterschiedlicher Hashfunktionen auf k Positionen im Bloomfilter ab. Das Ergebnis von f_i ist ein positiver Integer-Wert zwischen 0 und $n-1$.

Durchführung

1. Konkatenation der geheimen, in der VST vorliegenden Hash-Schlüssel für 4 Erfassungsjahre jeweils mit dem aktuellen technischen Feldnamen (ID) des Vornamens bzw. des Nachnamens, das Ergebnis ist der key_{ID_EJ}
 - Beispiel 1: $ID = \text{vorname_mutter}$, Hash-Schlüssel für EJ 1 = 35DB7, $key_{ID_EJ} = \text{vorname_mutter35DB7}$
 - Beispiel 2: $ID = \text{nachname_mutter}$, Hash-Schlüssel für EJ 2 = XR79T, $key_{ID_EJ} = \text{nachname_mutterXR79T}$
2. Erstellung eines Bloomfilters der Länge n bei dem jedes Bit mit 0 bzw. $false$ initialisiert wird

3. Konkatenation der Zählvariablen der aktuellen Hashfunktion (i), des Geburtstages des Kindes (T), des Feldnamens (ID) und des aktuellen Bigramms (b)
4. Erstellung des Pseudonyms mit Hilfe des HMAC-Algorithmus
5. Durchführung einer Modulardivision mit der Länge des Bloomfilters (n)
6. Das Bit im Bloomfilter entsprechend des Ergebnisses der Modulardivision auf 1/true setzen
7. Punkte 3-6 für jede Hashfunktion $f_i | i \in 0, k$ wiederholen
8. Punkte 3-7 für jedes Bigramm b wiederholen
9. Punkte 2-7 für jeden Schlüssel key_{ID_EJ} (4 Erfassungsjahre) wiederholen

**Hinweis**

Der Vorname der Mutter und der Nachname der Mutter sind jeweils in einem separaten Bloomfilter abzubilden!

Im Falle, dass einer der beiden oder beide Angabe nicht vorliegen enthält der Bloomfilter für das jeweilige Pseudonym einen leeren String.

3.3.3 Pseudonymisierung des Geburtsdatums des Kindes

Das Geburtsdatum des Kindes wird nicht in einem Bloomfilter abgebildet. Das Geburtsdatum wird lediglich mit dem HMAC-Algorithmus pseudonymisiert. Folgende Schritte sind hierfür durchzuführen:

1. Für jedes der 4 Erfassungsjahre Konkatenation der geheimen, in der VST vorliegenden Hash-Schlüssel mit dem aktuellen Feldnamen (ID) *Geburtsdatum des Kindes* („*GEBDATUMK*“):
 key_{ID_EJ}
2. Bildung des Pseudonyms für 4 Erfassungsjahre: $\forall_{key_{ID_EJ}}: HMAC(key_{ID_EJ}, GEBDATUMK)$

4 Prozesse

4.1 Leistungserbringer

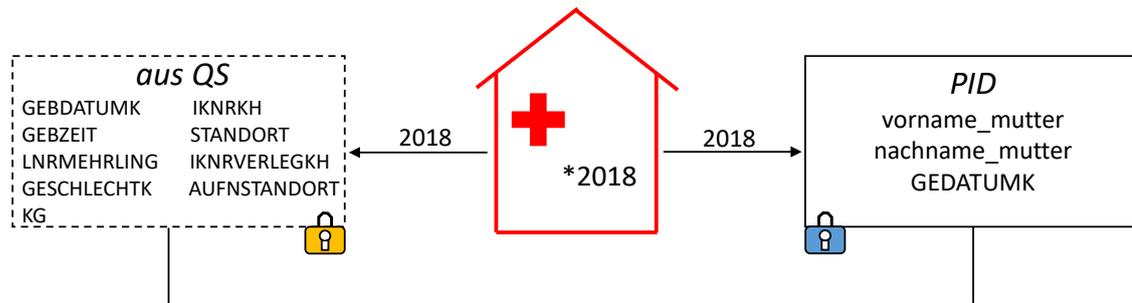


Abbildung 2: Prozesse beim Leistungserbringer

Beim Leistungserbringer müssen die QS-Daten (Abbildung 2 linker Bereich) und die patientenidentifizierenden Daten (PID-Daten, Abbildung 2, rechter Bereich) separat verarbeitet werden. Die QS-Daten werden mit dem öffentlichen Schlüssel der zuständigen DAS XML-verschlüsselt an die DAS (LQS/LKG) übermittelt⁶. Die PID-Daten dürfen von der DAS nicht eingesehen werden und müssen mit dem öffentlichen Schlüssel der VST XML-verschlüsselt werden. Der Datencontainer für die PID-Daten wird nur erstellt, wenn es sich bei dem Patienten um GKV-Versicherte handelt.

⁶ Die XML-Verschlüsselung der QS-Daten erfolgt aus prozessualen Gründen und nach Vorgaben der Richtlinien des G-BA. Einerseits sind Datenflüsse indirekter Verfahren in anderen Leistungsbereichen vorgesehen, in denen die zuständige DAS keine Einsicht in die QS-Daten haben darf. Andererseits werden QS-Daten stets XML-verschlüsselt an die BAS übermittelt, unabhängig von Verfahren und Datenfluss. Somit werden die QS-Daten stets von der an die BAS versendende bzw. weiterleitende Stelle XML-verschlüsselt.

4.1.1 XML-Struktur

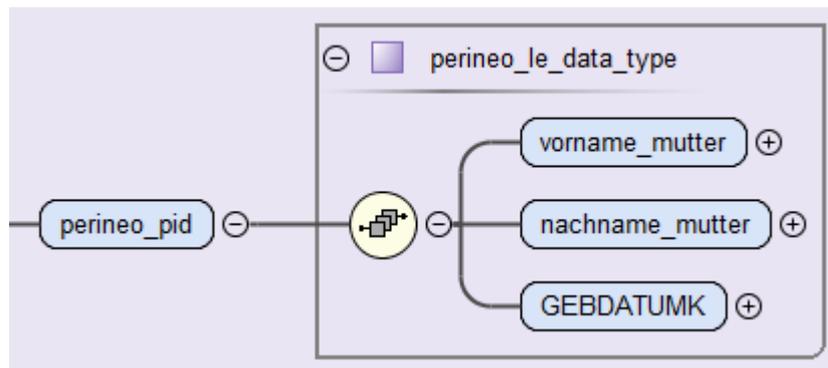


Abbildung 3: XML-Schema für das `<perineo_pid>` Element zur Übermittlung der PID-Daten zwischen LE und DAS

4.1.2 Erläuterung

Das `<patient>` Element wird im Rahmen der Basisspezifikation genutzt um patientenspezifische Informationen GKV-Versicherter in einem separaten Container zu transportieren. Dieser Container enthält außerhalb der QS-Verfahren Perinatalogie/Neonatologie lediglich die eGK Versichertennummer des Patienten im `<pid>` Element. Das `<patient>` Element wird beim Leistungserbringer mit dem öffentlichen Schlüssel der VST verschlüsselt und ist anschließend nur im Rahmen der Erstellung des Patientenpseudonyms bei der VST einsehbar.

Die PID-Daten der Verknüpfung Peri/Neo werden genau wie die eGK-Versichertennummer des Patienten bei der VST pseudonymisiert und dürfen nicht in den DAS auf Landesebene einsehbar sein. Aus diesem Grund wird bei Datensätzen der Perinatalogie und der Neonatologie das `<pid>` Element durch ein `<perineo_pid>` Element ersetzt. Das `<perineo_pid>` Element enthält in seinen Kindelementen den Vornamen der Mutter `<vorname_mutter>`, den Nachnamen der Mutter `<nachname_mutter>`, das Geburtsdatum des Kindes `<GEBDATUMK>` sowie bei Datensätzen der Neonatologie die eGK Versichertennummer des Kindes `<VERSICHERTENIDNEUK>`.

Im Zuge der Verschlüsselung des `<patient>` Elementes mit dem öffentlichen Schlüssel der VST beim LE werden somit alle für die VST relevanten Informationen in einem Container hinterlegt.

4.2 Datenannahmestellen auf Landesebene



Abbildung 4: Prozesse in der DAS auf Landesebene

Die DAS auf Landesebene führen eine Datenprüfung der QS-Daten durch, verschlüsseln diese mit dem öffentlichen Schlüssel des IQTIG und leiten die gesamte Datenlieferung an die VST weiter. Die PID-Daten werden in der DAS nicht verarbeitet.

4.3 Vertrauensstelle

Die VST ist für die Durchführung der Pseudonymisierung der PID-Daten nach dem Schnell-Bloomfilterverfahren sowie nach dem adaptierten Krebsregisterverfahren verantwortlich. Sie empfängt die zu pseudonymisierenden PID-Daten im Rahmen der regulären Datenübermittlung der QS-Datensätze Perinatalogie/Neonatologie von den DAS auf Landesebene. Die entsprechenden Informationen sind mit dem öffentlichen Schlüssel der VST XML-verschlüsselt. Die VST benötigt zur Durchführung der Pseudonymisierung Hash-Schlüssel die Sie nach den Vorgaben in diesem Dokument erstellen, aufbewahren und löschen muss.

4.3.1 Generierung der Hash-Schlüssel

Die für die Pseudonymisierung benötigten Geheimnisse (Hash-Schlüssel) werden von der VST nach den gleichen Vorgaben erstellt wie die Geheimnisse, die zur Pseudonymisierung der eGK-Versichertennummer des Patienten Anwendung finden. Die Sicherheit dieses Verfahrens ist höher als die durch den G-BA bei der Pseudonymisierung der PID-Daten geforderte Sicherheitsstufe von 128 Bit.

4.3.2 Speicher- und Löschfristen

Die Verknüpfung der Peri/Neo-Daten erfolgt über maximal 4 aufeinanderfolgende Jahre. Aus diesem Grund werden in der VST zu jedem Zeitpunkt 4 Hash-Schlüssel (siehe Abschnitt 4.3.1) vorgehalten. Mit Ablauf des Erfassungsjahres wird der älteste Schlüssel gelöscht und ein neuer Schlüssel erzeugt.

Eine Ausnahme bildet das Erfassungsjahr der erstmaligen Durchführung der Verknüpfung der Peri/Neo-Daten. Die VST generiert in dem des Erfassungsjahres vorangegangenen Jahr 4 Hash-Schlüssel zur Nutzung für die folgenden 4 Erfassungsjahre.

4.3.3 Entschlüsselung der PID-Informationen

Die PID-Daten befinden sich XML-verschlüsselt im `<patient>` Element, sofern es sich bei der Mutter und dem Kind um GKV-Versicherte handelt. In allen anderen Fällen ist kein `<patient>` Element vorhanden, so dass in diesen Fällen keine Pseudonymisierung durch die VST durchgeführt werden muss. Die VST entschlüsselt die PID-Daten zur anschließenden Pseudonymisierung mittels Schnell-Bloomfilterverfahren sowie dem adaptierten Krebsregisterverfahren.

4.3.4 Pseudonymisierung der PID-Daten

Die Pseudonymisierung der PID-Daten erfolgt in der VST folgendermaßen:

1. Entschlüsselung der PID-Daten mit dem privaten Schlüssel der VST
2. Standardisierung des Vornamens und des Nachnamens der Mutter gemäß den Vorgaben aus Abschnitt 3.2
3. Pseudonymisierung des Vornamens, Nachnamens, Geburtsdatums des Kindes sowie der eGK-Nummer des Kindes gemäß den Vorgaben aus Abschnitt 3.3

4.3.5 XML-Struktur zur Übermittlung der Pseudonyme an das IQTIG

Die VST entfernt nach der Pseudonymisierung alle Kindelemente aus dem Element `<perineo_pid>` und fügt die Elemente `<bloomfilter>` ein (siehe Abbildung 5). Die Inhalte der einzelnen Elemente werden in den folgenden Abschnitten detailliert beschrieben.

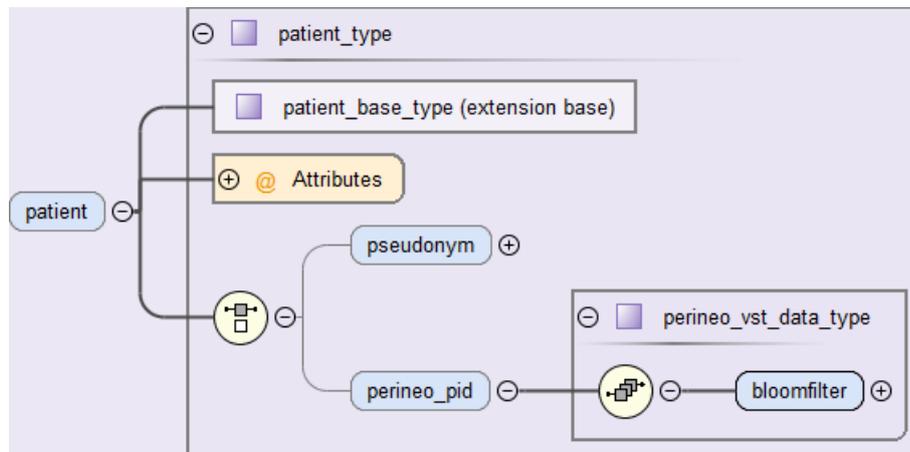


Abbildung 5: XML-Schema des `<patient>` Elementes nach Pseudonymisierung in der VST

Bloomfilter-Spezifische Pseudonyme

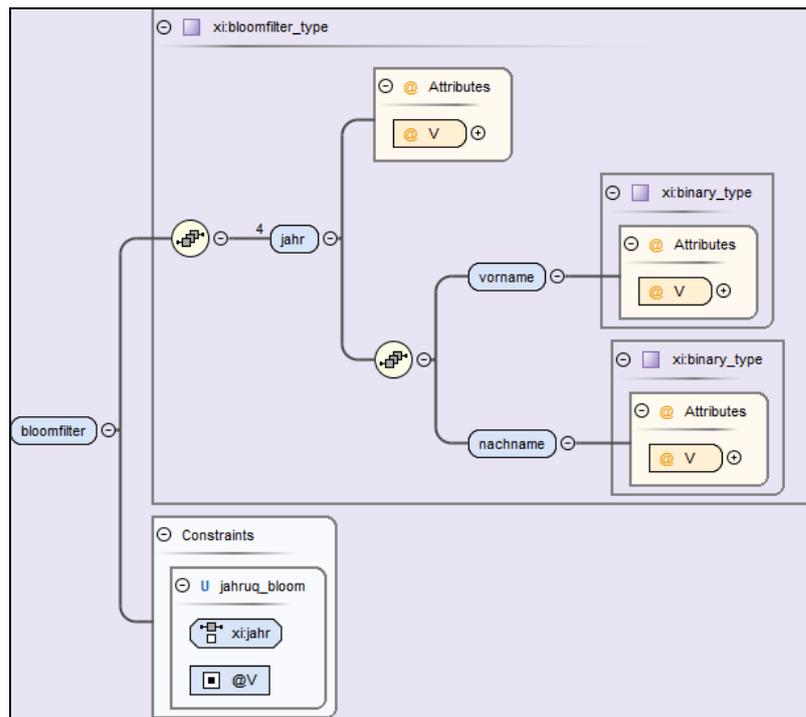


Abbildung 6: XML-Schema des `<bloomfilter>` Elementes zur Übermittlung der bloomfilterspezifischen Pseudonyme

In Abbildung 6: XML-Schema des `<bloomfilter>` Elementes zur Übermittlung der bloomfilterspezifischen Pseudonyme. Abbildung 6 ist die XML-Struktur für die Pseudonyme des Vornamens und des Nachnamens der Mutter beim Bloomfilterverfahren dargestellt. Ein `<bloomfilter>` Element enthält verpflichtend vier `<jahr>` Elemente mit fortlaufenden Jahresangaben im Attribut V. In der Spezifikation sind jeweils vier verschiedene Jahresangaben zulässig. Jedes `<jahr>` Element muss einen eindeutigen Jahreswert beinhalten. Dies wird per Constraint in der XML-Schemadefinition sichergestellt. Jedes `<jahr>` Element muss je ein Element `<vorname>` sowie `<nachname>` enthalten. Die Werte (Attribut V = Value) enthalten die jeweilige Bloomfilterrepräsentation des Vor- bzw. Nachnamens als 1000-stelliger String der Zeichen 0 und 1.

4.3.6 Übermittlung der Pseudonyme an das IQTIG

Die Pseudonyme werden entsprechend den Regelungen zur Übermittlung des Patientenpseudonyms in den in Abschnitt 4.3.5 definierten Datenstrukturen an das IQTIG übermittelt.

4.4 Bundesauswertungsstelle

4.4.1 Speicher- und Löschfristen

Das IQTIG erhält jeden PID-Datensatz in vierfacher Ausfertigung von der Vertrauensstelle jeweils pseudonymisiert mit dem Hash-Schlüssel des aktuellen Erfassungsjahres sowie den Hash-Schlüsseln der drei folgenden Erfassungsjahre. Die jeweiligen Datensätze werden entsprechend im IQTIG gepflegt und zu den gegebenen Löschfristen gelöscht.

5 Literaturverzeichnis

- [1] G. Heller, S. Konheiser, und T. Thomas, „Qualitätsreport 2011“, AQUA-Institut für Qualitätsförderung und Forschung im Gesundheitswesen GmbH, Hrsg. Göttingen, S. 122–127.
- [2] AQUA-Institut, „Bericht zum Sonderexport 2013“. 2012. Zugegriffen: 5. Januar 2017. [Online]. Verfügbar unter: http://www.sqg.de/downloads/Entwicklung/Follow_up/QSKH-Follow-up-Bericht-2013_01_16.pdf
- [3] J. B. Ford, C. L. Roberts, und L. K. Taylor, „Characteristics of unmatched maternal and baby records in linked birth records and hospital discharge data“, *Paediatr. Perinat. Epidemiol.*, Bd. 20, Nr. 4, S. 329–337, Juli 2006, doi: 10.1111/j.1365-3016.2006.00715.x.
- [4] J. P. Bentley, J. B. Ford, L. K. Taylor, K. A. Irvine, und C. L. Roberts, „Investigating linkage rates among probabilistically linked birth and hospitalization records“, *BMC Med. Res. Methodol.*, Bd. 12, S. 149, 2012, doi: 10.1186/1471-2288-12-149.
- [5] R. Schnell und G. Heller, „Alternative Verfahren zur pseudonymisierten Verknüpfung der Daten: Hier für die Leistungsbereiche Geburtshilfe und Neonatologie“, Gemeinsamer Bundesausschuss, Hrsg. Berlin, 2015.
- [6] B. H. Bloom, „Space/Time Trade-offs in Hash Coding with Allowable Errors“, *Commun. ACM*, Bd. 13, S. 422–426, 1970.
- [7] C. M. Gutierrez, J. M. Turner, D. D. Foreword, und C. F. Director, *The Keyed-Hash Message Authentication Code (HMAC)*. 2008.
- [8] R. Schnell, T. Bachteler, und J. Reiher, „Privacy-preserving record linkage using Bloom filters“, *BMC Med. Inform. Decis. Mak.*, Bd. 9, S. 41, 2009, doi: 10.1186/1472-6947-9-41.
- [9] F. Niedermeyer, S. Steinmetzer, M. Kroll, und R. Schnell, „Cryptanalysis of Basic Bloom Filters Used for Privacy Preserving Record Linkage“, *J. Priv. Confidentiality*, Bd. 6, Nr. 2, Dez. 2014, [Online]. Verfügbar unter: <http://repository.cmu.edu/jpc/vol6/iss2/3>