Bundesministerium
des Innern, für Bau
und Heimat

Digital Gipfel
Nürnberg 2018

Rolf Schwartmann / Steffen Weiß (Hrsg.)

Anforderungen an den datenschutzkonformen Einsatz von Pseudonymisierungslösungen

Ein Arbeitspapier der Fokusgruppe Datenschutz der Plattform Sicherheit, Schutz und Vertrauen für Gesellschaft und Wirtschaft im Rahmen des Digital-Gipfels 2018

Digital Gipfel
Nürnberg 2018

Anforderungen an den datenschutzkonformen Einsatz von Pseudonymisierungslösungen

Ein Arbeitspapier der Fokusgruppe Datenschutz der Plattform Sicherheit, Schutz und Vertrauen für Gesellschaft und Wirtschaft im Rahmen des Digital-Gipfels 2018

Leitung:

Prof. Dr. Rolf Schwartmann
Kölner Forschungsstelle für Medienrecht – Technische Hochschule Köln – Mitglied der Datenethikkommission der Bundesregierung

Koordination:

Steffen Weiß, LL.M.
Gesellschaft für Datenschutz und Datensicherheit e.V.

Mitglieder:

Prof. Dr. Christoph Bauer
ePrivacy GmbH

Patrick von Braunmühl
Bundesdruckerei GmbH

Dr. Guido Brinkel
Microsoft Deutschland GmbH

Susanne Dehmel
Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V.

Philipp Ehmann
eco – Verband der Internetwirtschaft e.V.

Walter Ernestus
Die Bundesbeauftragte für den Datenschutz und die Informationsfreiheit

Nicolas Goß
eco – Verband der Internetwirtschaft e.V.

Michael Herfert
Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.

Maximilian Hermann
Kölner Forschungsstelle für Medienrecht – Technische Hochschule Köln

Dr. Detlef Houdeau
Infineon Technologies AG

Angelika Hüsch-Schneider
Deutsche Telekom AG

Clemens John
United Internet AG

Annette Karstedt-Meierrieks
Deutscher Industrie- und Handelskammertag e.V.

Robin L. Mühlenbeck
Kölner Forschungsstelle für Medienrecht – Technische Hochschule Köln

Daniel Krupka
Gesellschaft für Informatik e.V.

Johannes Landvogt
Die Bundesbeauftragte für den Datenschutz und die Informationsfreiheit

Prof. Dr. Michael Meier
Universität Bonn/Gesellschaft für Informatik e.V.

Dr. Frank Niedermeyer
Bundesamt für Sicherheit in der Informationstechnik

Jonas Postneek
Bundesamt für Sicherheit in der Informationstechnik

Frederick Richter, LL.M.
Stiftung Datenschutz

Dr. Sachiko Scheuing
Acxiom Deutschland GmbH

Irene Schlünder
Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.

Sebastian Schulz
Bundesverband E-Commerce und Versandhandel Deutschland e.V.

Dr. Claus D. Ulmer
Deutsche Telekom AG

Dr. Winfried Veil
Bundesministerium des Innern

Dr. Martina Vomhof
Gesamtverband der Deutschen Versicherungswirtschaft e.V.

Benjamin Walczak
Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein

Version 1.01, 2019

Urheber:
Fokusgruppe Datenschutz des Digital-Gipfels

Leitung:
Prof. Dr. Rolf Schwartmann

Kölner Forschungsstelle für Medienrecht

Technology
Arts Sciences
TH Köln

Kontakt:
Steffen Weiß

Gesellschaft für Datenschutz und Datensicherheit e.V.
Heinrich-Böll-Ring 10
53119 Bonn
Tel.: +49 228 96 96 75 00
E-Mail: info@gdd.de
Internet: www.gdd.de

GDD
Gesellschaft für Datenschutz und Datensicherheit e.V.

Vorwort

Pseudonymisierung als Brücke zwischen informationeller und unternehmerischer Selbstbestimmung

Maschinen sollen das Leben sicherer, leichter, angenehmer und länger machen. Der Mensch bzw. dessen Intelligenz ist der Ausgangspunkt der KI. Die Technik soll menschliches Verhalten durch maschinelles Arbeiten und Verstehen nachahmen, um sie auf dieser Grundlage gegebenenfalls selbständig anzuwenden. Hierzu werden riesige Datenmengen mit dem Ziel verarbeitet, aus den Daten Muster zu erkennen, sie auszuwerten und Schlüsse daraus zu ziehen.

„Künstliche Intelligenz – ein Schlüssel für Wachstum und Wohlstand". Damit ist der Digital-Gipfel der Bundesregierung 2018 überschrieben. Deutschland ist in der Digitalwirtschaft stark und soll in der KI führend werden. Die Strategie ist richtig. Sie ist aber an rechtliche Leitlinien gebunden. „Die Datenschutz-Grundverordnung (DSGVO) bildet einen verlässlichen gesetzlichen Rahmen für innovative Technologien und Anwendungen auch im Bereich der KI. Sie enthält Vorschriften zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Verkehr solcher Daten. Die Überarbeitung der E-Privacy-Verordnung soll dieses Schutzkonzept abrunden." So lautet das klare Bekenntnis der Bundesregierung im Eckpunktepapier zur Digitalstrategie.

Um personenbezogene Daten wirtschaftlich nutzbar zu machen setzt die DS-GVO auf die Pseudonymisierung. Sie hat eine Doppelfunktion, indem sie personenbezogene Daten zugleich schützen und deren wirtschaftliche Nutzung ermöglichen soll. Der Kern der Pseudonymisierung besteht darin, Identitätsdaten einer Person wie bei einem KfZ-Kennzeichen durch eine Zeichenkette zu ersetzen. Der Rückschluss vom Pseudonym auf die Person erfolgt nach festen Regeln.

Die Fokusgruppe Datenschutz der Plattform Sicherheit, Schutz und Vertrauen hat 2017 ein Whitepaper* vorgelegt. Es legt Leitlinien für eine rechtssichere Nutzung von Pseudonymisierungslösungen unter Berücksichtigung der DS-GVO vor.

Im Jahr 2018 hat die Fokusgruppe ihre Arbeit fortgesetzt und legt dieses Arbeitspapier vor. Es formuliert Anforderungen an den datenschutzkonformen Einsatz von Pseudonymisierungslösungen. Zugleich stellt es einen notwendigen Zwischenschritt vom Whitepaper auf dem Weg zu einem Vorschlag für einen Pseudonymisierungsstandard dar, mit dem sich die Fokusgruppe Datenschutz im Jahr 2019 befassen will. Er soll auf dem Digital-Gipfel 2019 vorgelegt werden und der Wirtschaft zu mehr Investitionssicherheit verhelfen.

Allen Mitwirkenden an der Fokusgruppe gilt herzlicher Dank für ihren intensiven, konstruktiven und effizienten Einsatz in diesem Gremium. Herrn Assessor Steffen Weiß, gilt besonderer Dank, dafür dass er die Arbeit der Gruppe so fachkundig und umsichtig koordiniert.

Köln, im November 2018

Professor Dr. Rolf Schwartmann

Leiter der Fokusgruppe Datenschutz der Plattform Sicherheit, Schutz und Vertrauen für Gesellschaft und Wirtschaft im Rahmen des Digital-Gipfels 2018 und Mitglied der Datenethikkommission der Bundesregierung

* Abrufbar unter: https://www.gdd.de/downloads/whitepaper-zur-pseudonymisierung (deutsche Fassung); https://www.telekom.com/resource/blob/503396/a358f4551a46a542c1c918756996f771/dl-170912-white-paper-pseudonymisiation-data.pdf (englische Fassung).

Inhalt

A. Einleitung

Ziel des Leitfadens ist es, die für die Datenverarbeitung Verantwortlichen bei der rechtssicheren Umsetzung von Pseudonymisierungsmaßnahmen durch entsprechende Vorgaben zu unterstützen. Bereits im Whitepaper für den Digital-Gipfel 2017[1] hat die „Fokusgruppe Datenschutz der Plattform Sicherheit, Schutz und Vertrauen für Gesellschaft und Wirtschaft“ die Bedeutung der Pseudonymisierung herausgearbeitet.

Wesentliches Kennzeichen einer Pseudonymisierung ist, dass Pseudonyme ohne Hinzuziehung zusätzlicher Informationen nicht mehr der spezifisch betroffenen Person zugeordnet werden können. Insofern schützt die Pseudonymisierung Bürgerinnen und Bürger, deren personenbezogene Daten verarbeitet werden, vor einer ungewollten Identifikation. Im Gegensatz zur Anonymisierung[2] ist bei pseudonymisierten Daten eine Rückführung auf die Einzelperson (Re-Identifizierung) möglich. Die Stärke einer Pseudonymisierung hängt davon ab, wie hoch das Risiko, die Kosten sowie der Zeitaufwand für eine direkte oder indirekte Identifizierung durch Dritte einzuschätzen sind.

Die Pseudonymisierung von Daten ist ein geeignetes Mittel zur datenschutzkonformen Erstellung von Statistiken, zur Durchführung von Forschungsvorhaben, sowie zur Durchführung von Werbemaßnahmen. Spezielle Anwendungsszenarien für Pseudonymisierungen sind in dem o.g. Whitepaper zu Pseudonymisierung aufgeführt.

In diesem Leitfaden werden im Anschluss an die Darstellung der rechtlichen Einordnung der Pseudonymisierung (Abschnitt B.), die Voraussetzungen für einen rechtssicheren Pseudonymisierungsprozess herausgearbeitet (Abschnitt C.) sowie Anforderungen aufgezeigt, die eine Pseudonymisierung typischerweise erfüllen muss (Abschnitt D.).

B. Rechtliche Einordnung der Pseudonymisierung

Die Pseudonymisierung allein macht eine Datenverarbeitung nicht rechtmäßig. Sie ist lediglich ein Baustein, um eine Datenverarbeitung im Einklang mit der EU-Datenschutz-Grundverordnung (DSG-VO) zu gewährleisten. Erforderlich ist daher immer, dass eine Erlaubnisgrundlage für die Datenverarbeitung vorliegt. Es müssen die Anforderungen des Art. 6 DSGVO („Rechtmäßigkeit der Verarbeitung“) sowie bei besonderen Kategorien personenbezogener Daten die des Art. 9 DSGVO („Verarbeitung besonderer Kategorien personenbezogener Daten“) erfüllt sein.

Bei der Pseudonymisierung sind daher zwei wesentliche Anwendungsfälle zu unterscheiden:

1. Pseudonymisierung als technische Schutzmaßnahme

Pseudonyme sind z.B. erforderlich, wenn kritische Datenverarbeitungen gegenüber unzulässigem Zugriff besonders geschützt werden müssen. In diesem Fall dient die Pseudonymisierung vorwiegend der Risikoreduktion im Sinne des Art. 32 DSGVO. Bei den gesetzlichen Anforderungen gem. Art. 25 DSGVO an „Privacy by Design“ sorgt eine Pseudonymisierung dafür, dass bereits in einem frühen Stadium eine Entkoppelung persönlicher Informationen von anderen Daten erfolgen kann.

2. Pseudonymisierung als Ermöglichung einer Verarbeitung bzw. Weiterverarbeitung

Nach der DSGVO können Pseudonyme wegen der damit einhergehenden Risikoreduktion bestimmte Datenverarbeitungen aber auch ermöglichen. Das sind insbesondere die Fälle der sogenannten kompatiblen Weiterverarbeitung nach Art. 6 Abs. 4 DSGVO, d.h. der Zweck für die Ersterhebung von Daten und eine darauf folgende Zweckänderung sind als miteinander kompatibel einzustufen, was im Ergebnis zu einer zulässigen Datenverarbeitung führt.

Ob ein neuer Verarbeitungszweck mit dem ursprünglichen Zweck vereinbar ist und die Weiterverarbeitung daher auf die ursprüngliche Rechtsgrundlage gestützt werden kann, ist das Ergebnis einer Abwägung verschiedener in Art. 6 Abs. 4 DSGVO genannter Kriterien. Ein Kriterium, das für eine Kompatibilität der Zwecke spricht, ist das Vorhandensein geeigneter Garantien, wozu auch Verschlüsselung oder Pseudonymisierung gehören können (Art. 6 Abs. 4 lit. e DSGVO). Auch im Rahmen einer Interessenabwägung nach Art. 6 Abs. 1 lit. f DSGVO kann sich eine Pseudonymisierung zugunsten eines Verarbeiters auswirken und eine Verarbeitung legitimieren.

C. Voraussetzungen für Pseudonymisierungen

Jede Pseudonymisierung muss bestimmte Voraussetzungen einhalten, um als solche rechtssicher im Sinne der DSGVO gestaltet zu sein.

1. Zuweisung von Fachverantwortlichkeiten

Für die Überwachung des Pseudonymisierungsprozesses ist durch die verantwortliche Stelle eine Person, z.B. ein Fachverantwortlicher zu bestimmen. Diese Person sollte das notwendige techni-

1 https://www.gdd.de/downloads/whitepaper-zur-pseudonymisierung.
2 Die Anonymisierung ist ein eigenständiges Verfahren, für das spezielle Anforderungen gelten.

sche und rechtliche Verständnis zur Steuerung des Prozesses vorweisen können. Die Rolle dieser Person besteht darin, die Verantwortung für wichtige Entscheidungen zu übernehmen. Er/Sie sollte in die Lage versetzt werden, einen einheitlichen Ansatz beim Datenverarbeiter bezüglich der Pseudonymisierung zu koordinieren und die Möglichkeit erhalten, auf Know-how aus dem eigenen Haus oder außerhalb der Organisation zurückgreifen zu können. Die verantwortliche Stelle bleibt weiterhin in der Verantwortung, die Aufgaben und Pflichten der DSGVO zu erfüllen.

2. Anwendbarkeit und rechtliche Zulässigkeit

Je nach Anwendungsfall der Pseudonymisierung sind unterschiedliche Zulässigkeitsvoraussetzungen zu beachten.

a. Bei der Pseudonymisierung als Maßnahme zur Realisierung eines angemessenen Schutzes personenbezogener Daten im Sinne des Art. 32 DSGVO wird sich die Anforderung an die Pseudonymisierung in aller Regel aus einer Risikobetrachtung unter Einbeziehung der in Art. 32 DSGVO genannten Kriterien (Stand der Technik, Implementierungskosten, Art, Umfang, Umstände und Zweck der Verarbeitung sowie Höhe des Risikos für die Rechte der betroffenen Person) ergeben. Besonders kritische Datenverarbeitungen sind besonders zu schützen und mit einer entsprechend starken Pseudonymisierung zu versehen (zu den technisch-organisatorischen Anforderungen vgl. D.).

b. Soll die Pseudonymisierung zur Ermöglichung einer Verarbeitung oder Weiterverarbeitung von personenbezogenen Daten eingesetzt werden, kann dies entweder über eine kompatible Weiterverarbeitung gem. Art. 6 Abs. 4 DSGVO bzw. über eine Interessensabwägung nach Art. 6 Abs. 1 lit. f DSGVO erfolgen. Bei der Weiterverarbeitung gem. Art. 6 Abs. 4 DSGVO sind die gesetzlichen Voraussetzungen dieser Norm kumulativ zu prüfen und abzuwägen. Je nach Ausprägung der einzelnen Prüfpunkte (lit. a) bis e)) kann sich die Zulässigkeit der Weiterverarbeitung ergeben oder nicht. Bei der Interessensabwägung im Sinne des Art. 6 Abs. 1 lit. f DSGVO ist insbesondere das berechtigte Interesse des Verantwortlichen zu prüfen.

Die Art und Qualität der Pseudonymisierung ist bei beiden Fallvarianten von besonderer Bedeutung.

3. Betroffeneninformation – Transparenz – und Widerspruchsmöglichkeiten

Der Betroffene ist auch bei der Verarbeitung pseudonymisierter Daten nach den allgemeinen Grundsätzen über die Datenverarbeitung ausreichend zu informieren. Dies kann z.B. über Datenschutzhinweise erfolgen. Die gesetzlichen Handlungsmöglichkeiten wie z.B. Widerspruchsrechte oder Auskunftsrechte sind ihm einzuräumen.

a. Bei der Pseudonymisierung zu Schutzzwecken werden die allgemeinen Informationen bei der Datenerhebung ausreichend sein. Widerspruchsrechte oder Einwilligungserfordernisse bestimmen sich hier nach den gesetzlichen Erlaubnistatbeständen für die ursprünglich geplante Verarbeitung.

b. Erfolgt eine pseudonyme Weiterverarbeitung, beispielsweise zu kompatiblen Zwecken, ist der Betroffene hierüber grundsätzlich zu informieren. Der andere Zweck, Inhalt und Umfang der Weiterverarbeitung sind ihm darzulegen. Hierbei bietet es sich an, den Betroffenen auch über die vorgenommene Pseudonymisierung zu informieren. Zudem ist der Betroffene in diesem Kontext ggf. auf sein Widerspruchsrecht hinzuweisen mit dem er verhindern kann, dass seine ursprünglich erhobenen Daten Teil einer kompatiblen Weiterverarbeitung werden (Art. 6 Abs. 4 lit. d)).

c. Hat eine verantwortliche Stelle pseudonymisierte Daten von einer dritten Stelle erhalten und kann die verantwortliche Stelle den Betroffenen nicht mehr ohne Weiteres identifizieren, stellt die verantwortliche Stelle zumindest die allgemeine Information über die eigene Webseite zur Verfügung, dass pseudonyme Daten verarbeitet werden. Hierbei sind die Herkunft der Daten und die Möglichkeit für Auskunftsansprüche zu benennen.

d. Die antragsabhängigen Betroffenenrechte des Kapitel III der DSGVO (Recht auf Auskunft, Berichtigung, Löschung, Einschränkung der Verarbeitung, Datenportabilität sowie Widerspruch) sind durch den Verantwortlichen vollumfänglich – auch bezogen auf das gespeicherte Pseudonym - zu erfüllen, wenn der Verantwortliche die natürliche Person zu einem Pseudonym identifizieren kann. Der Verantwortliche vergewissert sich jedoch beim Betroffenen, ob eine damit verbundene Re-Identifizierung – bezogen auf das Pseudonym – erwünscht ist.

Verlangt ein Betroffener Auskunft und kann der Verantwortliche den Betroffenen über das Pseudonym nicht identifizieren, weil er bestimmte hierfür notwendige Informationen über den Betroffenen nicht

hat, so hat er dies - sofern möglich - dem Betroffenen mitzuteilen. Im Rahmen der Information muss zumindest der Hinweis auf die Datenherkunft erfolgen sowie auf eine mögliche Identifizierung, wenn der Betroffene die zu seiner Identifikation erforderlichen Informationen bereitstellt (vgl. Art. 11 Abs. 2 DSGVO).

4. Regelungen zum Zusammenführen mit Einzelangaben bestimmter Personen

Sollen die Ergebnisse der pseudonymen Datenverarbeitung mit den Einzelangaben von Betroffenen zusammengeführt werden oder auf den Betroffenen zurückgeführt werden (Re-Identifizierung), so kann dies entweder Teil der ursprünglich geplanten Verarbeitung sein oder eine Zusatzleistung, die auf der Auswertung der kompatiblen Weiterverarbeitung beruht.

a. Wird die Pseudonymisierung als Schutzmaßnahme im Rahmen einer legitimen Datenverarbeitung eingesetzt, die grundsätzlich auch mit Klardaten der Betroffenen zulässig ist, bedarf es über die ursprüngliche Legitimation zur Datenverarbeitung keiner weiteren Erlaubnis mehr zur Rückführung der Pseudonyme auf Einzelpersonen.

b. Handelt es sich bei der Pseudonymisierung um eine Maßnahme zur Ermöglichung der kompatiblen Weiterverarbeitung bereits erhobener Daten, so erstreckt sich die Legitimation nach Art. 6 Abs. 4 DSGVO nur auf die Weiterverarbeitung der Daten, nicht jedoch auf die Rückbeziehung auf einzelne Personen. In diesen Fällen bedarf es für die Rückbeziehung daher der Zurverfügungstellung eines Einwilligungs-Mechanismus für die Betroffenen, der den Anforderungen des Art. 7 DSGVO entspricht. Dies gilt auch für eine Verarbeitung auf Basis einer Interessensabwägung gem. Art. 6 Abs. 1 lit. f DSGVO.

5. Dokumentation

Die Voraussetzungen für eine rechtssichere Pseudonymisierung sowie die Prozessschritte zur Durchführung einer Pseudonymisierung sind zu dokumentieren. Dies kann entweder über ein eigenständiges Pseduonymisierungskonzept oder über eine allgemeine Beschreibung im Rahmen der Darlegung technisch-organisatorischer Maßnahmen für ein Verfahren erfolgen.

D. Technisch-organisatorische Anforderungen an die Pseudonymisierung

D.1 Begriffserklärungen

In den folgenden Abschnitten werden folgende Begriffe verwendet. Die kurzen Erklärungen der Begriffe dienen dazu, ihre Verwendung im Folgenden verständlich zu machen, und können nicht dem Anspruch vollständiger Definitionen genügen.

D.1.1 k-Anonymität

Eine (pseudonymisierte) Datensammlung bietet k-Anonymität, falls die darin noch enthaltenen Identitätsdaten jeder einzelnen Person mit mindestens k - 1 anderen Personen übereinstimmen. K ist hier eine natürliche Zahl.

D.1.2 Aufdeckbarkeit

Ein Pseudonym heißt aufdeckbar, wenn es möglich ist, vom Pseudonym auf die Identitätsdaten der dazugehörigen Person zu schließen. Hierzu ist unter Umständen ein geheimer, nur bestimmten Stellen zugänglicher kryptographischer Schlüssel notwendig.

D.1.3 Aufzählungsangriff

Wenn sämtliche Details (inklusive der dabei verwendeten kryptographischen Schlüssel) eines Pseudonymisierungsverfahrens bekannt sind, können aus einem vorliegenden Pseudonym die zugehörigen Identitätsdaten durch einen Aufzählungsangriff (auch „vollständige Exhaustion“ oder „Probeverschlüsselung“) bestimmt werden. Hierzu werden sämtliche infrage kommenden Identitätsdaten der Pseudonymisierung unterworfen und mit dem vorliegenden Pseudonym verglichen.

Wenn beispielsweise f eine kryptographische Hashfunktion ist und der Wert y = f(name) bekannt, name aber unbekannt ist, können für sämtliche infrage kommenden Namen der Wert f(name) berechnet und mit y verglichen werden, um „name“ zu bestimmen.

D.1.4 Blockchiffrierverfahren

Ein Verschlüsselungsverfahren, welches einen Datenblock fester Länge (z.B. 128 Bit) in Abhängigkeit eines kryptographischen Schlüssels in einen Block derselben Länge transformiert. Das heutzutage geläufigste Blockchiffrierverfahren ist der AES (Advanced Encryption Standard), welcher 128-Bit-Blöcke anhand eines 128-, 192- oder 256-Bit-Schlüssels verschlüsselt.

D.1.5 Datensammlung

Aus mehreren Datensätzen bestehendes Datenmaterial aus möglicherweise unterschiedlichen Quellen oder Jahren, das zu statistischen Zwecken ausgewertet werden soll und aus diesem Grunde pseudonymisiert werden soll.

D.1.6 Datensatz

Eine zu einer Person gehörige Information, die Identitäts- und Inhaltsdaten enthält und die es zu pseudonymisieren gilt.

D.1.7 Datentreuhänder

Siehe Vertrauensstelle.

D.1.8 l-Diversität

Eine (pseudonymisierte) Datensammlung bietet l-Diversität, falls es zu jeder Gruppe von darin noch enthaltenen identischen Identitätsdaten mindestens l unterschiedliche Ausprägungen der Inhaltsdaten gibt. l ist hier eine natürliche Zahl.

D.1.9 Einwegfunktion

Funktion f, die leicht berechenbar aber schwer umzukehren ist; es soll praktisch unmöglich sein, aus einem Funktionswert y Rückschlüsse auf x mit f(x) = y zu ziehen.

Bemerkung:

Für eine Einwegfunktion ist es notwendig, dass der Definitionsbereich von f sehr groß ist, da ansonsten für alle infrage kommende x der Wert f(x) berechnet und mit y verglichen werden könnte. Für ein Beispiel siehe: Aufzählungsangriff.

D.1.10 Entropie

Ein Maß für die Unbestimmtheit einer Zeichenfolge. Beispielsweise liefern zehn von einander unabhängige Münzwürfe (Kopf/Zahl) zehn Bit Entropie. Wird eine Folge mithilfe eines Pseudozufallszahlengenerators aus einem Anfangswert („Seed“) berechnet, so kann diese nie eine höhere Entropie als der Anfangswert erreichen.
Ein kryptographischer Schlüssel sollte eine Entropie von mindestens 100 Bit enthalten.

D.1.11 HMAC

Siehe: Kryptographische Prüfsumme.

D.1.12 Homonymfehler

Ein Homonymfehler entsteht, wenn bei Verkettbarkeit leistenden Pseudonymisierungsverfahren Identitätsdaten von unterschiedlichen Personen fälschlicherweise zu gleichen Pseudonymen führen.

D.1.13 Identitätsdaten

Alle eine Person betreffende Daten, die es ermöglichen, die Person näher zu bestimmen.

D.1.14 Inhaltsdaten

In einer Datensammlung im Wesentlichen alle Daten, die nicht zu den Identitätsdaten gehören. Nichtsdestotrotz kann aus Inhaltsdaten ein Personenbezug hergestellt werden, wenn sie z.B. einmalig sind und diese Information mit einer Person in Verbindung gebracht werden kann.

Bemerkung:

Mitunter kann es zu Überschneidungen zwischen Inhaltsdaten und Identitätsdaten kommen, etwa in der Datensammlung für eine Studie, die Aussagen über die Abhängigkeit von Alter oder Beruf zu bestimmten Merkmalen untersuchen soll. In diesem Fall würden Alter und Beruf (auch) zu den Inhaltsdaten gezählt.

D.1.15 Kontrollnummer

Siehe: Pseudonym.

D.1.16 Kryptographische Hashfunktion

Eine Hashfunktion ist eine Funktion, die einer Zeichenkette beliebiger Länge eine Zeichenkette fester Länge (etwa 256 Bit) zuordnet. Eine kryptographische Hashfunktion hat zudem die Eigenschaft einer Einwegfunktion. Gilt zusätzlich, dass es praktisch unmöglich ist, zwei unterschiedliche Eingabewerte zu finden, die denselben Funktionswert liefern, spricht man von einer kollisionsresistenten Hashfunktion. International genormte kryptographische Hashfunktionen sind etwa MD5, SHA256 oder SHA-3.

D.1.17 Kryptographischen Prüfsumme

Eine Bitfolge fester Länge (etwa 256 Bit), die sich aus einer Zeichenkette beliebiger Länge anhand eines kryptographischen Schlüssels errechnet. Bei Kenntnis des Schlüssels ist es anhand der Prüfsumme möglich, die Unversehrtheit der Zeichenkette festzustellen. Ohne Kenntnis des Schlüssels ist es unmöglich, für eine Zeichenkette eine gültige kryptographische Prüfsumme zu erstellen. Eine international genormte kryptographische Prüfsumme wird mit dem HMAC-Algorithmus ((Keyed-Hash Message Authentication Code) berechnet.

D.1.18 Kryptographischer Schlüssel

Eine Zeichenkette, anhand derer eine Datenmenge mittels einer kryptographischen Funktion (Verschlüsselung oder Signatur) transformiert wird. Je nach Anwendungsfall ist der Schlüssel geheim zu halten.

D.1.19 Pseudonym

Eine Zeichenkette, die Identitätsdaten einer Person ersetzt und damit diese Person repräsentiert. Von einem Pseudonym soll, wenn überhaupt, nur unter fest definierten Bedingungen auf die Identitätsdaten geschlossen werden können (siehe Aufdeckbarkeit).

D.1.20 Pseudonymisierungsliste

Eine Liste, die Identitätsdaten und Pseudonyme gegenüberstellt. Anhand einer Pseudonymisierungsliste können aus den Identitätsdaten einer Person direkt deren Pseudonyme und umgekehrt aus einem Pseudonym einer Person deren Identitätsdaten ermittelt werden.

D.1.21 Pseudonymisierungsstufe

Wird ein Pseudonym nicht direkt aus den Identitätsdaten erstellt, sondern in voneinander unabhängigen Schritten über Zwischenergebnisse, spricht man von Pseudonymisierungsstufen.

Bemerkung:
Eine Pseudonymisierung in mehreren Stufen findet etwa bei der Beteiligung einer oder mehrerer Vertrauensstellen statt.

D.1.22 Pseudonymisierungsverfahren
Ein Verfahren, welches aus Identitätsdaten einer Person ein Pseudonym generiert.

D.1.23 Record Linkage
In der Fachliteratur wird das Zusammenführen von Datensätzen einer pseudonymisierten Datensammlung anhand von verkettbaren Pseudonymen als Record Linkage bezeichnet.

D.1.24 Re-Identifizierung
Siehe Aufdeckbarkeit.

D.1.25 Synonymfehler
Entsteht, wenn bei einem verkettbaren Pseudonymisierungsverfahren Identitätsdaten derselben Person fälschlicherweise zu unterschiedlichen Pseudonymen führen, obwohl dies nicht beabsichtigt war.

D.1.26 Verkettbarkeit der Pseudonyme
Ein Pseudonymisierungsverfahren gewährleistet Verkettbarkeit der Pseudonyme, wenn Identitätsdaten zur selben Person in der Regel zu gleichen oder ähnlichen Pseudonymen führen. Die Pseudonyme bzw. die Datensätze der Person sind dann „verkettbar“: Aus identischen Pseudonymen lässt sich in der Regel auf identische Personen schließen.

Bemerkungen:
Die Verkettbarkeit von pseudonymisierten Daten mit Personen ohne Kenntnis des Pseudonymisierungsverfahrens bzw. der Pseudonymisierungstabelle ist dabei nicht gemeint und ist zu vermeiden.

Bei verkettbaren Pseudonymen kann es dennoch zu Homonym- oder Synonymfehlern (siehe dort) kommen.

D.1.27 Verschlüsselung
Ein Verfahren, welches einen Klartext in Abhängigkeit eines kryptographischen Schlüssels in einen Geheimtext umwandelt. Die Umkehrung, also aus dem Geheimtext den Klartext wiederherzustellen, nennt man Entschlüsselung.

D.1.28 Vertrauensstelle
Eine von der Datenerhebung und der Datenauswertung räumlich und organisatorisch unabhängige Stelle. Die einzige Aufgabe der Vertrauensstelle besteht hier in der Unterstützung der Umwandlung von Identitätsdaten in Pseudonyme.

Bemerkung:
Gegebenenfalls können bei einem Pseudonymisierungsverfahren mehrere Vertrauensstellen beteiligt sein, die die Pseudonyme in mehreren Pseudonymisierungsstufen erstellen.

D.1.29 Zuordnungstabelle
Siehe Pseudonymisierungsliste.

D.2 Maßnahmen

D.2.1 Grundsätzliches
Bei der Pseudonymisierung sind Grundsätze einzuhalten, die es bei jedem Verfahren zu beachten gilt:

a. Kenntnis, nur wenn nötig

b. Löschen von Daten, wenn immer möglich

c. Vermeidung der Ansammlung von zu viel Wissen an einer Stelle (z.B. hinsichtlich Klartextdaten und pseudonymisierten Daten über eine Person)

d. Pseudonyme nur dann, wenn die Notwendigkeit dafür besteht; ansonsten Anonymisierung

Abhängig vom Kontext können dabei unterschiedliche Arten von Pseudonymen zum Einsatz kommen:

- Personen-Pseudonyme, die an Stelle von Identitätsdaten wie z.B.: Name, Ausweisnummer oder Mobiltelefonnummer stehen
- Rollen-Pseudonyme, bei denen eine oder ggf. mehrere Personen einem Pseudonym zugeordnet sind (z.B. IP-Nummer)
- Beziehungs-Pseudonyme, bei denen eine Person für jede (Kommunikations-) Beziehung ein anderes Pseudonym verwendet, z.B. unterschiedliche Spitznamen
- Rollen-Beziehungs-Pseudonyme, die eine Kombination der beiden Pseudonym-Arten sind
- Transaktions-Pseudonyme, bei denen für jede Transaktion ein neues Pseudonym genutzt wird, was z.B. beim Online-Banking zum Einsatz kommt

Generell ist die Verkettbarkeit von Personen-Pseudonymen höher als von Rollen- bzw. Beziehungs-Pseudonymen. Noch geringer ist die Verkettbarkeit von Rollen-Beziehungs-Pseudonymen und Transaktions-Pseudonyme; sie sind prinzipiell nicht verkettbar. Grundsätzlich gilt, dass umso geringer die Verkettbarkeit der Pseudonymisierung ist, umso größer die mögliche Anonymität der Daten für Dritte

ist. Eine geringe Verkettbarkeit erhöht zugleich die Stärke der Pseudonymisierung.

Daneben sind bei der technisch-organisatorischen Umsetzung einer Pseudonymisierung verschiedene Verfahrensschritte zu durchlaufen, die sich typischerweise wie folgt darstellen:

D.2.2 Schaffung eines Pseudonyms (Pseudonymisierung des Datensatzes)

Jede Pseudonymisierung beginnt mit der Erstellung von Pseudonymen, die Datensätze mit zugehörigen natürlichen Personen verbindet. Das Pseudonym kann ggf. zur Re-Identifizierung eines Datensatzes dienen, ist gesondert aufzubewahren und durch technisch-organisatorische Maßnahmen zu schützen.

Bei den zu pseudonymisierenden Daten wird zwischen Identitätsdaten der beteiligten Personen und Inhaltsdaten unterschieden. Eine strikte Trennung zwischen beiden Datenarten ist nicht in allen Fällen möglich, so dass auch Inhaltsdaten Angaben zu einer Person enthalten können (z.B. Geschlecht, Berufsgruppe und Geburtsjahr) und dadurch ein Personenbezug möglich wird.

Die Art der gewählten Pseudonymisierung kann grundsätzlich Einfluss auf die Handlungsspielräume des Anwenders haben. Mit einer starken Pseudonymisierung können in der Regel kritischere Datenverarbeitungen ausreichend geschützt werden, als mit einer schwachen Pseudonymisierung. Ebenso gilt im Bereich der kompatiblen Weiterverarbeitung, dass mit stärkerer Pseudonymisierung auch eher von einer gegebenen Kompatibilität der beabsichtigten Weiterverarbeitung mit dem Ausgangszweck ausgegangen werden kann.

Bei der Schaffung eines Pseudonyms stehen grundsätzlich zwei Verfahren zur Verfügung: Pseudonymisierungslisten und Pseudonyme durch Berechnungsverfahren.

D.2.2.1 Pseudonymisierungslisten

Eine Pseudonymisierungsliste ordnet Identitätsdaten anhand einer Tabelle Pseudonymen zu. Die Pseudonyme haben dabei keinen inhaltlichen oder funktionalen Bezug zu den Identitätsdaten.

Beispiel 1:
Pseudonyme werden durchnummeriert.

Identitätsdaten	Pseudonym
Peter Müller **geb. 31.01.1965** **in Köln**	2022917
Maria Schulze **geb. 03.05.1959** **in Hürth**	2022918
Max Klein **geb. 31.10.1967** **in Bornheim**	2022919

Beispiel 2:
Pseudonyme werden zufällig oder pseudozufällig erzeugt.

Identitätsdaten	Pseudonym
Peter Müller **geb. 31.01.1965** **in Köln**	2184578
Maria Schulze **geb. 03.05.1959** **in Hürth**	3654425
Max Klein **geb. 31.10.1967** **in Bornheim**	8745124

Anmerkungen:

1. Bei der Durchnummerierung der Pseudonyme lassen sich eventuell Rückschlüsse auf Identitätsdaten ziehen. Etwa, wenn die Ausgangsdaten alphabetisch sortiert sind. Oder zu welchem Zeitpunkt die Pseudonyme erzeugt wurden (Beispiel: Spanische Kfz-Kennzeichen liefern Aufschluss über die Erstzulassung des Fahrzeugs).

2. Bei zufälligen Pseudonymen sollte die Länge der Pseudonyme nicht zu kurz gewählt werden, da es ansonsten zu Kollisionen und somit zu Homonymfehlern kommen kann. Als Faustformel gilt, dass bei n möglichen Pseudonymen es nach der Quadratwurzel aus n gebildeten Pseudonymen mit Wahrscheinlichkeit von 50 % zu einer Kollision kommt. Wenn also die Pseudonyme als zehnstellige Dezimalzahlen gewählt werden, kommt es nach 10000 zufällig erzeugten Pseudonymen mit Wahrscheinlichkeit von 50 % zu zwei gleichen Pseudonymen (Stichwort „Geburtstagsparadoxon“[3]).

3. Als Quelle des Zufalls sollte nicht die Zufallsfunktion verwendet werden, die von einer Programmiersprache angeboten wird (etwa die Funktion rand() in der Programmiersprache C). Beispielsweise kann als Zufallsquelle der iterierte Output einer kryptographischen Hashfunktion verwendet werden:

A1 = Hash(A0),
Pseudonym1 = Bit 1 bis 40 von A1

A2 = Hash(A1),
Pseudonym2 = Bit 1 bis 40 von A2

A3 = Hash(A2),
Pseudonym3 = Bit 1 bis 40 von A3

3 https://de.wikipedia.org/wiki/Geburtstagsparadoxon.

Dabei ist A0 ein von der Pseudonymisierungsstelle zu wählender echt zufälliger Wert mit einer Entropie von mindestens 100 Bit. Zur Wahl der Bitanzahl (hier 40) siehe Anmerkung 2.

4. Wenn mehrere Datenlieferanten am Pseudonymisierungsverfahren beteiligt sind und evtl. eine Re-Identifizierung des Datenlieferanten anhand eines Pseudonyms möglich sein soll, kann die Identität des Datenlieferanten ebenfalls pseudonymisiert werden und den Pseudonymen der Personen vorangestellt werden.

D.2.2.2 Pseudonyme durch Berechnungsverfahren

Eine weitere Möglichkeit ist, die Pseudonyme aus Identitätsdaten algorithmisch zu berechnen.

Der Transformationsprozess hat ein Verfahren nach State-of-the-Art zu berücksichtigen (z.B. BSI- Richtlinie TR-02102-11 oder ENISA-Richtlinie zu Kryptoverfahren), um Schwachstellen einer Verschlüsselung, die zu einer Aufdeckung einer Person führen können, zu vermeiden.

Um nicht vom Pseudonym auf die Identitätsdaten (ID) schließen zu können, muss die Berechnung von einem geheimen Parameter, einem sog. kryptographischen Schlüssel K, abhängig sein. Als Berechnungsmethoden bieten sich an:

Verschlüsselung mit einem Verschlüsselungsverfahren:
Pseudonym = EK(ID).

Hier bezeichnet EK die Verschlüsselung mit einem Blockchiffrieralgorithmus, etwa AES, mit dem Schlüssel K.

Bildung einer kryptographischen Prüfsumme:
Pseudonym = HMAC_K(ID).

Hier bezeichnet HMAC = einen Keyed-Hash Message Authentication Code, siehe etwa RFC2104.

Anmerkungen:

1. Die Entropie von K sollte mindestens 100 Bit betragen.

2. Zur Berechnung des Pseudonyms brauchen nicht alle Identitätsdaten herangezogen zu werden. Im Allgemeinen ist es ausreichend, eine Auswahl der Identitätsdaten zu treffen, sodass dadurch die Person in der zu pseudonymisierenden Datensammlung zu identifizieren ist. Siehe auch Abschnitt E.2.

3. Als Pseudonym braucht nicht der gesamte Output der Berechnung verwendet zu werden. Siehe Anmerkung 2 aus Abschnitt D.2.2.1.

4. Obwohl es sich bei einer kryptographischen Hashfunktion um eine Einwegfunktion handelt, ist es nicht ausreichend die Pseudonymberechnung ausschließlich durch die Hashfunktion durchzuführen, also etwa

- **Pseudonym = Hash(PID)**

Es könnte nämlich bei Vorliegen eines Pseudonyms durch einen Aufzählungsangriff (exhaustive Suche) aller infrage kommender Werte für PID dasjenige PID bestimmt werden, dessen Hashwert das Pseudonym ergibt. In Deutschland würde sich, je nach Zusammensetzung von PID, diese Suche auf lediglich maximal 80 Millionen Hashwertberechnungen beschränken.

5. In einer Datensammlung können die Identitätsdaten unter Umständen durch mehrere Pseudonyme ersetzt werden, die sich aus unterschiedlichen Attributen der Identitätsdaten errechnen.

Beispiel:

Pseudonym1 =
EK(Krankenversicherungsnummer)

Pseudonym2 =
EK(Name | Geburtstag | Geburtsort)

Pseudonym3 =
EK(Geburtsname | Geburtstag | Geburtsort)

6. Die Erzeugung und Verwaltung (u.a. Verteilung, Speicherung, Verwendung, Löschung) geheimer Parameter (kryptographische Schlüssel) sind durch nach Stand der Technik geeignete technische und organisatorische Maßnahmen zu realisieren.

7. Die Sicherheit des gewählten Pseudonymisierungsverfahrens kann dadurch erhöht werden, dass – zeit- oder datenvolumenabhängig – geeignete Intervalle definiert werden, in denen ein Wechsel verwendeter geheimer Parameter (kryptographischer Schlüssel) erfolgt. Ebenso können, je nach Art des gewählten Verfahrens und abhängig vom Risiko für Betroffene, mehrere Pseudonymisierungsstufen eingebaut werden, um eine Aufdeckbarkeit auszuschließen (sog. „Überverschlüsselung“).

D.2.2.3 Mehrstufige und gemischte Pseudonymisierungsverfahren

Die Sicherheit eines Pseudonymisierungsverfahren kann erhöht werden, wenn die Bildung der Pseudonyme von mehreren unabhängigen Stellen durchgeführt wird. Hierbei können sowohl Pseudonymisierungslisten als auch Berechnungsverfahren zum Einsatz kommen.

Beispiel:

1. A, B und C erheben Daten von Personen (A, B und C können beispielsweise Arztpraxen sein, die Patientendaten erheben).

2. A, B und C bilden für die Datensätze mithilfe eines Berechnungsverfahrens und eines kryptographischen Schlüssels K1 (der bei allen datenerhebenden Stellen zur Verfügung steht) Pseudonyme P1.

3. A, B und C liefern die pseudonymisierten Datensätze an eine Vertrauensstelle V.

4. V bildet aus den erhaltenen Pseudonymen P1 mithilfe eines Berechnungsverfahrens und eines kryptographischen Schlüssels K2 für die Datensätze neue Pseudonyme P2 und ersetzt die erhaltenen Pseudonyme P1 durch die neuen Pseudonyme P2.

5. V leitet die Datensätze mit den neuen Pseudonymen P2 an eine Sammelstelle S weiter.

6. S führt anhand der Pseudonyme P2 mittels Record Linkage die erhaltenen Datensätze zusammen.

7. Die Daten sollen an Stellen X, Y und Z (unter verschiedenen Gesichtspunkten) ausgewertet werden. Hierzu filtert S die Datensammlung und stellt für X, Y und Z die jeweils notwendigen Datensätze aus der Datensammlung zusammen.

8. Aus der (Teil-)Datensammlung für X (und ebenso für Y und Z) werden die Pseudonyme P2 entfernt und durch neue Pseudonyme P3 ersetzt, die sich aus einer Pseudonymisierungsliste LX ergeben, die den Pseudonymen P2 die Pseudonyme P3 zuordnet. Die Pseudonymisierungslisten LX, LY und LZ für X, Y und Z sind dabei unterschiedlich und unabhängig voneinander.

Bemerkung:
Durch die unterschiedlichen Listen ist sichergestellt, dass nicht mehrere Datenauswerter die ihnen zur Verfügung gestellten Datensammlungen anhand der darin enthaltenen Pseudonyme zusammenführen können.

D.2.2.4 Vor- und Nachteile unterschiedlicher Pseudonymisierungsverfahren

Verfahren	Vorteile	Nachteile
Zuordnungstabellen	**1.** Kein Schlüsselmanagement erforderlich	**1.** Schlechte Skalierbarkeit (Tabelle kann sehr groß werden) **2.** Tabelle muss dauerhaft geschützt werden **3.** Pseudonymisierer benötigt dauerhaft Zugriff auf gesamt Tabelle **4.** Aufdeckbarkeit verlangt Zugriff auf gesamte Tabelle **5.** Verkettbarkeit verlangt Zugriff auf gesamte Tabelle **6.** Zugriff auf Tabelle impliziert Verkettbarkeit und Aufdeckbarkeit (Verkettbarkeit und Aufdeckbarkeit nicht differenziert steuerbar) **7.** Rollenbindung erfordert rollenspezifische Tabellenkopien
Berechnungsverfahren	**1.** Gute Skalierbarkeit, keine Tabellenverwaltung **2.** Kontrolle der Kenntnis geheimer Parameter erlaubt Zugriffskontrolle auf Berechnungsvorschriften **3.** Verschiedene Parameter für Pseudonymisierung, Verkettbarkeit und Aufdeckung möglich, daher differenziert steuerbar **4.** Nur die kryptographischen Schlüssel müssen sicher geschützt werden **5.** Rollenbindung über rollenspezifische Parameterbereitstellung leicht möglich **6.** Zweckgebundene technische Parameterrekonstruktion liefert technisch zweckgebundene Verkettung/Aufdeckung	**1.** Schlüsselmanagement erforderlich (ggf. weitere geheime oder öffentliche Parameter)

D.2.3 Getrennte Aufbewahrung des kryptographischen Schlüssels

D.2.3.1 Zugriffskontrolle (Berechtigungskonzept)

Eine getrennte Aufbewahrung des kryptographischen Schlüssels bedarf eines dokumentierten Berechtigungskonzepts. Hierbei sind mindestens zwei sich voneinander unterscheidende Rollen zu definieren: 1) Die Rolle mit Zugriffsberechtigung auf den Schlüssel zur Re-Identifizierung; 2) Die Rolle mit Zugriff auf die pseudonymisierten Inhaltsdaten.

Es bietet sich an, für ein Pseudonymisierungsverfahren folgende Rollen zu definieren:

1. Daten liefern

2. Daten pseudonymisieren sowie ggf. re-identifizieren

3. Daten sammeln und anhand der Pseudonyme zusammenführen („Record Linkage“)

4. Daten auswerten

Unbedingt erforderlich ist es, dass die Rollen 2. und 4. getrennt voneinander existieren.

Es sollte hierbei vermieden werden, dass eine Person eine Berechtigung für mehrere Rollen erhält. Dies gilt auch für Administratoren. Entsprechende Ausnahmeregelungen sind zu begründen und zu dokumentieren.

Der Zugriff auf einen kryptographischen Schlüssel muss auf ein absolutes Minimum an vertrauenswürdigen Personen eingeschränkt werden (Need-to-Know-Prinzip).

Die Möglichkeit der Re-Identifizierung sollte nicht in der Abteilung einer Organisation bestehen, in der zu einem Pseudonym zugehörige Inhaltsdaten verarbeitet werden. Entsprechende Ausnahmeregelungen sind zu begründen und zu dokumentieren.

D.2.3.2 Vier-Augen-Prinzip

Jeder Zugriff auf einen kryptographischen Schlüssel zur Re-Identifizierung von Identitätsdaten hat nach dem Vier-Augen-Prinzip zu erfolgen. Dies kann technisch oder organisatorisch gelöst werden. Ferner sollte keine der beteiligten Personen eine Zugriffsberechtigung sowohl auf den kryptographischen Schlüssel, das Pseudonym als auch das zugehörige Inhaltsdatum haben. Ist das Vier-Augen-Prinzip nicht möglich, muss zumindest der Zugriff auf den kryptographischen Schlüssel personenbezogen protokolliert werden.

D.2.4 Dokumentation technisch-organisatorischer Maßnahmen zur Nichtzuordenbarkeit

Technisch-organisatorische Maßnahmen zur Gewährleistung einer Nichtzuordenbarkeit eines Pseudonyms zu Identitätsdaten, so beispielsweise im Falle einer fehlenden Legitimation, sind zu dokumentieren. Dies kann in einem Pseudonymisierungskonzept erfolgen. Das Konzept ist in ein IT-Sicherheitsmanagement (z.B. ISO/IEC 27001) einzubinden. Das IT-Sicherheitsmanagementsystem soll dokumentiert und dessen Wirksamkeit regelmäßig überprüft werden.

D.2.5 Regeln zur Aufdeckung

Da ggf. eine Re-Identifizierung von Identitätsdaten bei der Pseudonymisierung möglich ist, ist eine geplante Aufdeckung eines Pseudonyms zu regeln. Hierzu bedarf es einer dokumentierten Definition von Fällen einer gewünschten Aufdeckung. Der Vorgang der Re-Identifizierung des Betroffenen ist zu protokollieren. Aus der Protokollierung muss hervorgehen, welche Personen die Re-Identifizierung durchgeführt haben. Aus der Protokollierung dürfen keine Rückschlüsse auf die einem Pseudonym zugrunde liegenden Identitätsdaten gezogen werden können. Daher ist die Protokollierung in ihrem Umfang einzuschränken. Protokolldaten dürfen nur zeitlich begrenzt gespeichert werden.

D.2.6 Wegfall des Verarbeitungszwecks

Die Zwecke und die Dauer des Pseudonymisierungsverfahrens sind vorab festzulegen und die Maßnahmen für das Verfahrensende, einschließlich der technischen Umsetzung einer Datenlöschung, zu dokumentieren.

Fällt der Zweck für eine Pseudonymisierung weg, z.B. aufgrund Erreichung der damit verbundenen Zwecke, sind pseudonymisierte Daten datenschutzkonform zu löschen oder zu anonymisieren. Eine solche Anonymisierung ist in der Regel nicht durch ein Löschen der Pseudonyme zu erreichen, sondern muss als eigenständiges Verfahren erfolgen, für das spezielle Anforderungen gelten, auf die an dieser Stelle nicht detailliert eingegangen werden kann. Bei einer Anonymisierung ist im Übrigen in regelmäßigen Abständen zu prüfen, ob die Daten weiterhin als anonym einzuordnen sind. Hat ein Betroffener ein Recht auf Löschung seiner Daten, bezieht sich dieses Recht auf personenbezogene Daten sowie pseudonymisierte Daten, nicht auf anonyme Daten. Gesetzlich Aufbewahrungsfristen sind hierbei zu beachten.

E. Best Practices

E.1 Verkettbare Pseudonymisierungsverfahren

Ein Pseudonymisierungsverfahren liefert verkettbare Pseudonyme, wenn für Personen mit den gleichen oder ähnlichen Identitätsdaten gleiche oder ähnliche Pseudonyme erzeugt werden. In diesem Fall können Datensätze anhand der Pseudonyme zusammengeführt werden. Verkettbare Verfahren sind etwa für Langzeitstudien von Bedeutung oder wenn die Datensätze von unterschiedlichen Quellen stammen und für eine Studie zusammengeführt werden sollen. Der Prozess der Zusammenführung anhand von verkettbaren Pseudonymen wird in der Fachliteratur als Record Linkage bezeichnet.

Beispiele:

1. Für Studien zur Legalbewährung von Straftätern werden die Inhaltsdaten (Straftat, Strafmaß, Alter etc.) in einer Datenbank gesammelt. Aus Datenschutzgründen dürfen die Einträge keinen Personenbezug aufweisen. Behörden sind regelmäßig verpflichtet, Daten zu Vorstrafen von Personen nach gesetzlich vorgegebenen Zeiträumen zu löschen. Um dennoch Langzeitstudien über die Rückfälligkeit von Straftätern durchführen zu können, kann das Datenmaterial mit verkettbaren Pseudonymen versehen werden.

2. Bei den deutschen epidemiologischen Krebsregistern werden Datensätze über Krebspatienten pseudonymisiert gesammelt, um den Erfolg unterschiedlicher Behandlungsmethoden untersuchen zu können. Datenlieferanten sind beispielsweise Ärzte, Krankenhäuser und Sterberegister. Die Daten erstrecken sich zum Teil über lange Zeiträume und können sogar aus unterschiedlichen Bundesländern stammen, da die Patienten den Wohnort gewechselt haben können. Aussagekräftige Studien lassen sich nur anhand verkettbarer Pseudonyme erstellen.

Anmerkung:

Liegen in der Datensammlung für einen Datensatz mehrere Pseudonyme vor (siehe Anmerkung 5 in Abschnitt D.2.2.2), so können die Datensätze verkettet werden, wenn nur eines der Pseudonyme übereinstimmt.

E.2 Auswahl der Identitätsdaten

Alle eine Person betreffende Attribute, die es ermöglichen, die Person näher zu bestimmen, gehören zu den Identitätsdaten der Person. Dies können beispielsweise sein:

- Vor-, Familien- und Geburtsname
- Geschlecht
- Geburtsdatum und -ort
- Wohnort und Nationalität
- Anzahl der Geschwister
- Beruf oder Berufsgruppe
- Krankenversicherungs- oder Personalausweisnummer
- u. v. a. m.

E.2.1 Identitätsdaten für die Berechnung von Pseudonymen

Die Identitätsdaten einer Person können, wie in Abschnitt D.2.2.2 beschrieben, zur Berechnung des Pseudonyms zu der Person verwendet werden.

Dabei ist zu berücksichtigen, dass bei Verwendung einer kryptographischen Funktion zur Berechnung der Pseudonyme zwar gleiche Identitätsdaten gleiche Pseudonyme liefern, aber bereits geringe Abweichungen in den Identitätsdaten zu komplett anderen Pseudonymen führen. Gründe für eine Änderung der Pseudonyme können sein:

- Schreib- und Tippfehler oder Zahlendreher
- Namensänderung durch Hochzeit oder Scheidung
- Unterschiedliche Schreibweisen des Vornamens (z.B. Hans/Johannes, Inge/Ingrid)
- Wohnortwechsel
- Bezeichnungswechsel einer Ortschaft wegen Gebietsreform
- Unkenntnis eines Attributs (etwa Geburtsort)
- u. v. a. m.

Wenn der Fall auftritt, dass einer Person zu unterschiedlichen Zeitpunkten oder von unterschiedlichen Stellen unterschiedliche Pseudonyme zugeordnet werden, spricht man von einem Synonymfehler. In diesem Fall ist eine Verkettbarkeit der Pseudonyme zu dieser Person nicht mehr gewährleistet.

Die Synonymfehlerrate kann verringert werden durch folgende Maßnahmen:

- Weglassen eines Attributs bei der Berechnung des Pseudonyms, beispielsweise wird nur das Geburtsjahr anstelle des vollständigen Geburtsdatums verwendet
- Beschränkung beim Namen auf den oder die (etwa drei) Anfangsbuchstaben

- Verwendung eines Namens- oder Phonetikcodes anstelle des Namens (siehe etwa de.wikipedia.org/wiki/Kölner_Phonetik)
- Verwendung der Gemeindekennziffer statt Wohn- oder Geburtsort
- u. v. a. m.

Erhalten andererseits unterschiedliche Personen zu unterschiedlichen Zeitpunkten oder von unterschiedlichen Stellen dasselbe Pseudonym, spricht man von einem Homonymfehler. Sofern die Pseudonyme aus den Identitätsdaten berechnet werden, entstehen Homonymfehler immer dann, wenn die Identitätsdaten, aus denen die Pseudonyme berechnet werden, bei beiden Personen übereinstimmen.

Die Homonymfehlerrate kann verringert werden durch folgende Maßnahmen:

- Aufnahme zusätzlicher Attribute zur Berechnung des Pseudonyms, beispielsweise kann das vollständige Geburtsdatum anstelle nur des Geburtsjahrs verwendet werden
- Verwendung von langlebigen eindeutigen Merkmalen zur Berechnung der Pseudonyme, etwa die Rentenversicherungs- oder Krankenversicherungsnummer
- u. v. a. m.

Anmerkungen:

1. Bei einer hohen Synonymfehlerrate werden Werte im Allgemeinen unterschätzt (etwa die Rückfallquote bei einer Untersuchung zur Legalbewährung oder die Sterblichkeitsrate bei einer bestimmten Behandlungsmethode)
2. Bei einer hohen Homonymfehlerrate werden Werte im Allgemeinen überschätzt
3. Eine Verringerung der Synonymfehlerrate hat in der Regel eine Erhöhung der Homonymfehlerrate zur Folge – und umgekehrt
4. Ein Kompromiss zwischen Synonym- und Homonymfehlerrate hängt stark von der zugrundeliegenden oder zu erwartenden Datensammlung ab. Entsprechend sind die Attribute der Identitätsdaten, die zur Berechnung der Pseudonyme verwendet werden sollen, auszuwählen.

E.2.1 Identitätsdaten in den Inhaltsdaten

In pseudonymisierten Datensammlungen können die Inhaltsdaten weiterhin Identitätsdaten enthalten, sofern diese für die intendierte Forschung mithilfe der Datensammlung von Bedeutung sein können. Beispielsweise kann das Geschlecht, das Alter, der Wohnort (als fünfstellige Postleitzahl) oder der ausgeübte Beruf von Interesse sein. In gewissen Fällen kann es aber möglich sein, allein anhand der in den Inhaltsdaten enthaltenden Identitätsdaten Personen zu identifizieren. Beispielsweise ist es denkbar, dass es im Postleitzahlbereich 65432 nur einen einzigen Fliesenleger gibt. Dieser wäre dann in der Datensammlung zweifellos identifizierbar. Doch, selbst wenn es mehrere Fliesenleger mit der Postleitzahl 65432 gibt, wäre zu gewährleisten, dass diese nicht alle gemeinsam ein bestimmtes Merkmal, etwa eine bestimmte Krankheit, aufweisen, da man ansonsten von einer Person, von der man weiß, dass sie von Beruf Fliesenleger ist und die Postleitzahl 65432 hat, sofort wüsste, dass sie unter dieser Krankheit leidet.

Für eine pseudonymisierte Datensammlung muss daher k-Anonymität und l-Diversität gewährleistet sein.

Eine Datensammlung bietet k-Anonymität, falls die darin enthaltenen Identitätsdaten jeder einzelnen Person mit mindestens k - 1 anderen Personen übereinstimmen.

Eine Datensammlung bietet l-Diversität, falls es zu jeder Gruppe von darin enthaltenen identischen Identitätsdaten mindestens l unterschiedliche Ausprägungen der Inhaltsdaten gibt.

k und l sind hier natürliche Zahlen.

Anmerkungen:

1. Größere Werte für k und l repräsentieren in diesem Kontext eine größere Anonymität
2. k-Anonymität und l-Diversität kann durch Aggregation der Attribute in den Identitätsdaten erreicht werden.

Beispiele:

- Statt „Fliesenleger“ wird als Beruf „Handwerker“ angegeben.
- Alle Postleitzahlen in der Datensammlung, die mit 654 beginnen, werden zusammengefasst. Statt 65432 wird dann 654xx in der Datensammlung abgespeichert.

3. k-Anonymität und l-Diversität sind von der pseudonymisierenden Stelle (siehe Abschnitt D.2.3.a) herzustellen. Hierzu muss die pseudonymisierende Stelle Zugriff auf die in den Inhaltsdaten enthaltenen Attribute der Identitätsdaten haben.

E.3 Einbindung einer Vertrauensstelle

Die Sicherheit von Pseudonymisierungsverfahren wird im Allgemeinen erhöht, wenn die in Abschnitt D.2.3.a genannten Rollen organisatorisch und örtlich getrennt werden. Eine Vertrauensstelle nimmt dabei die Datensammlung des oder der Datenlieferanten entgegen, pseudonymisiert sie und leitet sie an die Datensammelstelle weiter. Die Datensammelstelle führt die erhaltenen Daten der Datensammlung sodann anhand der Pseudonyme zusammen. Die Datensammelstelle gibt sie schließlich an den oder die Datenauswerter weiter. Auf diese Weise kommen weder Datensammelstelle noch Datenauswerter zu irgendeinem Zeitpunkt mit den Identitätsdaten in Kontakt.

Nach der Pseudonymisierung in der Vertrauensstelle kann die Vertrauensstelle verpflichtet werden, die Identitätsdaten unwiederbringlich zu löschen, sofern keine Notwendigkeit der Re-Identifizierung der Pseudonyme besteht (siehe Abschnitte D.2.5 und E.4). Nach Abschluss des Gesamtverfahrens kann die Vertrauensstelle ggf. verpflichtet werden, auch die verwendeten kryptographischen Schlüssel zu löschen.

Für die Vertrauensstelle besteht dabei keine Notwendigkeit der Kenntnis der Inhaltsdaten, sondern muss bei einem verkettbaren Pseudonymisierungsverfahren lediglich die Identitätsdaten kennen. Es empfiehlt sich daher, die Inhaltsdaten auf einem getrennten Übertragungsweg von den Datenlieferanten direkt an die Datensammelstelle zu übermitteln. Der getrennte Übertragungsweg kann dabei physikalischer Natur sein; die Inhaltsdaten können aber auch über die Vertrauensstelle laufen und mit einem Chiffrierverfahren überschlüsselt sein, bei dem ausschließlich die Datensammelstelle in der Lage ist, die Daten zu entschlüsseln.

E.4 Aufdeckbarkeit von Pseudonymen/Re-Identifizierung

Unter bestimmten Voraussetzungen kann es notwendig sein, von einem Pseudonym auf die zugehörige Person bzw. deren Identitätsdaten zurückzuschließen.

Im Falle, dass das Pseudonym durch ein Berechnungsverfahren aus den Identitätsdaten entstanden ist, ist es für die Aufdeckbarkeit notwendig, dass die verwendeten kryptographischen Schlüssel nicht gelöscht wurden. Sofern die Bildung der Pseudonyme ein Verschlüsselungsverfahren verwendet wurde, kann das Pseudonym unmittelbar entschlüsselt werden, um an die Identitätsdaten zu gelangen. Wurde das Pseudonym durch eine kryptographischen Prüfsumme gebildet, ist eine Aufdeckung der Identitätsdaten nicht unmittelbar möglich. Sofern der verwendete Schlüssel K nicht gelöscht wurde, lassen sich die Identitätsdaten jedoch durch eine vollständige Exhaustion über alle infrage kommende Identitätsdaten (vgl. Anmerkung 4 in Abschnitt D.2.2.2) bestimmen.

Im Falle, dass das Pseudonym durch eine Pseudonymisierungsliste aus den Identitätsdaten entstanden ist, ist es für die Aufdeckbarkeit notwendig, dass die verwendete Pseudonymisierungsliste nicht gelöscht wurde.

Für mehrstufige und gemischte Verfahren sind alle zur Bildung verwendeten kryptographischen Schlüssel und Pseudonymisierungslisten für eine Aufdeckbarkeit notwendig.

Beim Beispielszenario aus Abschnitt D.2.2.3 wäre eine Re-Identifizierung eines Pseudonyms P3, welches beim Datenauswerter X vorliegt, wie folgt möglich:

1. X liefert das Pseudonym P3 an S

2. S bestimmt aus P3 anhand der Liste LX das Pseudonym P2

3. S liefert das Pseudonym P2 an V

4. V berechnet aus P2 anhand des Schlüssels K2 das Pseudonym P1

5. V liefert einer berechtigten Stelle, die Kenntnis vom Schlüssel K1 hat, das Pseudonym P1

Die berechtigte Stelle bestimmt aus P1 anhand des Schlüssels K1 die zugehörigen Identitätsdaten.