

DIN EN ISO 25237

ICS 35.240.80

**Medizinische Informatik –
Pseudonymisierung (ISO 25237:2017);
Deutsche Fassung EN ISO 25237:2017**

Health informatics –
Pseudonymization (ISO 25237:2017);
German version EN ISO 25237:2017

Informatique de santé –
Pseudonymisation (ISO 25237:2017);
Version allemande EN ISO 25237:2017

Gesamtumfang 80 Seiten

DIN-Normenausschuss Medizin (NAMED)

Nationales Vorwort

Dieses Dokument (EN ISO 25237:2017) wurde vom Technischen Komitee ISO/TC 215 „Health informatics“ (Sekretariat: ANSI, USA) unter Beteiligung deutscher Experten erarbeitet und vom Technischen Komitee CEN/TC 251 „Medizinische Informatik“ (Sekretariat: NEN, Niederlande) übernommen.

Das zuständige deutsche Normungsgremium ist der Arbeitsausschuss NA 063-07-04 AA „Sicherheit“ im DIN-Normenausschuss Medizin (NAMed).

Für die in diesem Dokument zitierte Internationale Norm wird im Folgenden auf die entsprechende Deutsche Norm hingewiesen:

ISO 27799	siehe DIN EN ISO 27799
ISO/IEC 15408-2	siehe DIN ISO/IEC 15408-2*)

Nationaler Anhang NA (informativ)

Literaturhinweise

DIN EN ISO 27799, *Medizinische Informatik — Informationssicherheitsmanagement im Gesundheitswesen bei Verwendung der ISO/IEC 27002*

DIN ISO/IEC 15408-2*), *Informationstechnik — IT-Sicherheitsverfahren — Evaluationskriterien für IT-Sicherheit — Teil 2: Funktionale Sicherheitsanforderungen (ISO/IEC 15408-2:2005); Text Englisch*

*) Die Norm wurde ersatzlos zurückgezogen.

Deutsche Fassung

Medizinische Informatik —
Pseudonymisierung
(ISO 25237:2017)

Health informatics —
Pseudonymization
(ISO 25237:2017)

Informatique de santé —
Pseudonymisation
(ISO 25237:2017)

Diese Europäische Norm wurde vom CEN am 14. Dezember 2016 angenommen.

Die CEN-Mitglieder sind gehalten, die CEN/CENELEC-Geschäftsordnung zu erfüllen, in der die Bedingungen festgelegt sind, unter denen dieser Europäischen Norm ohne jede Änderung der Status einer nationalen Norm zu geben ist. Auf dem letzten Stand befindliche Listen dieser nationalen Normen mit ihren bibliographischen Angaben sind beim Management-Zentrum des CEN-CENELEC oder bei jedem CEN-Mitglied auf Anfrage erhältlich.

Diese Europäische Norm besteht in drei offiziellen Fassungen (Deutsch, Englisch, Französisch). Eine Fassung in einer anderen Sprache, die von einem CEN-Mitglied in eigener Verantwortung durch Übersetzung in seine Landessprache gemacht und dem Management-Zentrum mitgeteilt worden ist, hat den gleichen Status wie die offiziellen Fassungen.

CEN-Mitglieder sind die nationalen Normungsinstitute von Belgien, Bulgarien, Dänemark, Deutschland, der ehemaligen jugoslawischen Republik Mazedonien, Estland, Finnland, Frankreich, Griechenland, Irland, Island, Italien, Kroatien, Lettland, Litauen, Luxemburg, Malta, den Niederlanden, Norwegen, Österreich, Polen, Portugal, Rumänien, Schweden, der Schweiz, Serbien, der Slowakei, Slowenien, Spanien, der Tschechischen Republik, der Türkei, Ungarn, dem Vereinigten Königreich und Zypern.



EUROPÄISCHES KOMITEE FÜR NORMUNG
EUROPEAN COMMITTEE FOR STANDARDIZATION
COMITÉ EUROPÉEN DE NORMALISATION

CEN-CENELEC Management-Zentrum: Avenue Marnix 17, B-1000 Brüssel

Inhalt

	Seite
Europäisches Vorwort.....	4
Vorwort.....	5
Einleitung.....	6
1 Anwendungsbereich.....	7
2 Normative Verweisungen.....	7
3 Begriffe.....	7
4 Abkürzungen.....	14
5 Anforderungen an den Schutz von personenbezogenen Daten im Gesundheitswesen.....	14
5.1 Ziele des Schutzes von personenbezogenen Daten.....	14
5.2 Allgemeines.....	15
5.3 Entpersonalisierung als Prozess der Risikominderung.....	16
5.3.1 Allgemeines.....	16
5.3.2 Pseudonymisierung.....	17
5.3.3 Anonymisierung.....	17
5.3.4 Direkte und indirekte Bezeichner.....	17
5.4 Schutz der personenbezogenen Daten von Entitäten.....	17
5.4.1 Vergleich von personenbezogenen und entpersonalisierten Daten.....	17
5.4.2 Konzept der Pseudonymisierung.....	19
5.5 Pseudonymisierung in der realen Welt.....	22
5.5.1 Begründung.....	22
5.5.2 Stufen des Schutzes von personenbezogenen Daten.....	22
5.6 Kategorien der betroffenen Personen.....	25
5.6.1 Allgemeines.....	25
5.6.2 Zu betreuende Personen.....	25
5.6.3 Ärztliches Personal und Organisationen.....	26
5.6.4 Gerätedaten.....	26
5.7 Klassifizierungsdaten.....	26
5.7.1 Nutzdaten.....	26
5.7.2 Beobachtungsdaten.....	26
5.7.3 Pseudonymisierte Daten.....	27
5.7.4 Anonymisierte Daten.....	27
5.8 Forschungsdaten.....	27
5.8.1 Allgemeines.....	27
5.8.2 Generierung von Forschungsdaten.....	28
5.8.3 Sekundäre Nutzung von personenbezogenen Gesundheitsinformationen.....	28
5.9 Zur Identifizierung geeignete Daten.....	28
5.9.1 Allgemeines.....	28
5.9.2 Patientenbezeichner.....	28
5.10 Daten von Opfern von Gewalttaten und öffentlich bekannten Personen.....	29
5.10.1 Allgemeines.....	29
5.10.2 Geninformationen.....	29
5.10.3 Vertrauenswürdiger Dienst.....	29
5.10.4 Bedarf bezüglich der Wiedererkennung von pseudonymisierten Daten.....	29

5.10.5	Eigenschaften von Pseudonymisierungsdiensten	30
6	Datenschutz durch Pseudonymisierung	31
6.1	Begriffsmodell der Problembereiche	31
6.2	Eignung von personenbezogenen Informationen zur direkten und zur indirekten Identifizierung.....	31
6.2.1	Allgemeines	31
6.2.2	Zur Identifizierung von Personen geeignete Variablen.....	32
6.2.3	Kumulationsvariablen	32
6.2.4	Ausreißervariablen	33
6.2.5	Strukturierte Datenvariablen	34
6.2.6	Nicht-strukturierte Datenvariablen.....	34
6.2.7	Inferenzrisikobewertung.....	35
6.2.8	Privatheit und Sicherheit.....	35
7	Wiedererkennungsprozess	35
7.1	Allgemeines.....	35
7.2	Teil der normalen Verfahren.....	36
7.3	Ausnahme	36
7.4	Technische Realisierbarkeit.....	37
	Anhang A (informativ) Szenarien für die Pseudonymisierung im Gesundheitswesen	38
	Anhang B (informativ) Anforderungen an die Analyse des Risikos für den Schutz von personenbezogenen Daten.....	53
	Anhang C (informativ) Pseudonymisierungsprozess (Verfahren und Implementierung)	64
	Anhang D (informativ) Spezifikation von Verfahren und Implementierung	70
	Anhang E (informativ) Grundstruktur einer Policy für den Betrieb von Pseudonymisierungsdiensten (Verfahren und Implementierung)	72
	Anhang F (informativ) Genetische Informationen	76
	Literaturhinweise	77

Europäisches Vorwort

Dieses Dokument (EN ISO 25237:2017) wurde vom Technischen Komitee ISO/TC 215 „Health informatics“ in Zusammenarbeit mit dem Technischen Komitee CEN/TC 251 „Medizinische Informatik“ erarbeitet, dessen Sekretariat von NEN gehalten wird.

Diese Europäische Norm muss den Status einer nationalen Norm erhalten, entweder durch Veröffentlichung eines identischen Textes oder durch Anerkennung bis Juli 2017, und etwaige entgegenstehende nationale Normen müssen bis Juli 2017 zurückgezogen werden.

Es wird auf die Möglichkeit hingewiesen, dass einige Elemente dieses Dokuments Patentrechte berühren können. CEN [und/oder CENELEC] sind nicht dafür verantwortlich, einige oder alle diesbezüglichen Patentrechte zu identifizieren.

Entsprechend der CEN-CENELEC-Geschäftsordnung sind die nationalen Normungsinstitute der folgenden Länder gehalten, diese Europäische Norm zu übernehmen: Belgien, Bulgarien, Dänemark, Deutschland, die ehemalige jugoslawische Republik Mazedonien, Estland, Finnland, Frankreich, Griechenland, Irland, Island, Italien, Kroatien, Lettland, Litauen, Luxemburg, Malta, Niederlande, Norwegen, Österreich, Polen, Portugal, Rumänien, Schweden, Schweiz, Serbien, Slowakei, Slowenien, Spanien, Tschechische Republik, Türkei, Ungarn, Vereinigtes Königreich und Zypern.

Anerkennungsnotiz

Der Text von ISO 25237:2017 wurde von CEN als EN ISO 25237:2017 ohne irgendeine Abänderung genehmigt.

Vorwort

ISO (die Internationale Organisation für Normung) ist eine weltweite Vereinigung von Nationalen Normungsorganisationen (ISO-Mitgliedsorganisationen). Die Erstellung von Internationalen Normen wird normalerweise von ISO Technischen Komitees durchgeführt. Jede Mitgliedsorganisation, die Interesse an einem Thema hat, für welches ein Technisches Komitee gegründet wurde, hat das Recht, in diesem Komitee vertreten zu sein. Internationale Organisationen, staatlich und nicht-staatlich, in Liaison mit ISO, nehmen ebenfalls an der Arbeit teil. ISO arbeitet eng mit der Internationalen Elektrotechnischen Kommission (IEC) bei allen elektrotechnischen Themen zusammen.

Die Verfahren, die bei der Entwicklung dieses Dokuments angewendet wurden und die für die weitere Pflege vorgesehen sind, werden in den ISO/IEC-Direktiven, Teil 1 beschrieben. Im Besonderen sollten die für die verschiedenen ISO-Dokumentenarten notwendigen Annahmekriterien beachtet werden. Dieses Dokument wurde in Übereinstimmung mit den Gestaltungsregeln der ISO/IEC-Direktiven, Teil 2 erarbeitet (siehe www.iso.org/directives).

Es wird auf die Möglichkeit hingewiesen, dass einige Elemente dieses Dokuments Patentrechte berühren können. ISO ist nicht dafür verantwortlich, einige oder alle diesbezüglichen Patentrechte zu identifizieren. Details zu allen während der Entwicklung des Dokuments identifizierten Patentrechten finden sich in der Einleitung und/oder in der ISO-Liste der empfangenen Patenterklärungen (siehe www.iso.org/patents).

Jeder in diesem Dokument verwendete Handelsname wird als Information zum Nutzen der Anwender angegeben und stellt keine Anerkennung dar.

Eine Erläuterung der Bedeutung ISO-spezifischer Benennungen und Ausdrücke, die sich auf Konformitätsbewertung beziehen, sowie Informationen über die Beachtung der Grundsätze der Welthandelsorganisation (WTO) zu technischen Handelshemmnissen (TBT, en: Technical Barriers to Trade) durch ISO enthält der folgende Link: www.iso.org/iso/foreword.html.

Das für dieses Dokument verantwortliche Komitee ist ISO/TC 215, *Health informatics*.

Einleitung

Die Pseudonymisierung gilt als wichtiges Verfahren für den Schutz von personenbezogenen Gesundheitsinformationen. Die entsprechenden Dienste können sowohl auf nationaler Ebene als auch für die grenzüberschreitende Kommunikation genutzt werden.

Zu den Anwendungsfeldern gehören u. a.:

- die indirekte Nutzung von klinischen Daten (z. B. zu Forschungszwecken);
- klinische Studien und die Überwachung nach Markteinführung;
- pseudonymisierte Pflege;
- Patientenerkennungssysteme;
- Überwachung und Bewertung der öffentlichen Gesundheit;
- vertrauliche Berichte zur Patientensicherheit (z. B. zu Nebenwirkungen von Medikamenten);
- vergleichende Berichte zu Qualitätsindikatoren;
- Kreuzgutachten (Peer-Reviews);
- Verbrauchergruppen;
- Außendienst.

Dieses Dokument stellt ein Begriffsmodell der Problembereiche, Anforderungen an vertrauenswürdige Praktiken sowie Spezifikationen zur Unterstützung der Planung und Implementierung von Pseudonymisierungsdiensten zur Verfügung.

Die Spezifikation eines allgemeinen Ablaufplans dient in Verbindung mit einer Politik bzw. Policy für vertrauenswürdige Vorgänge sowohl als allgemeiner Leitfaden für Implementierer als auch zu Qualitätssicherungszwecken, indem sie den Anwendern der Pseudonymisierungsdienste dabei hilft, zu ermitteln, inwieweit sie diesen Diensten vertrauen können. Dieser Leitfaden wird auch Ausbildungsorganisationen von Nutzen sein, indem diese befähigt werden, Pseudonymisierungsdienste selbst in einer Weise auszuführen, die ausreicht, um den gewünschten Gütegrad und die gewünschte Risikominderung zu erreichen.

1 Anwendungsbereich

Dieses Dokument enthält Grundsätze für und Anforderungen an den Datenschutz für den Fall, dass zum Schutz von personenbezogenen Gesundheitsinformationen Pseudonymisierungsdienste genutzt werden. Dieses Dokument gilt für Organisationen, die Pseudonymisierungsprozesse für ihre eigenen Zwecke ausführen wollen, oder für Organisationen, die die Vertrauenswürdigkeit an Pseudonymisierungsdiensten beteiligter Vorgänge in Anspruch nehmen.

Dieses Dokument

- legt ein Grundkonzept für die Pseudonymisierung fest (siehe Abschnitt 5);
- legt eine grundlegende Methodik für Pseudonymisierungsdienste einschließlich organisatorischer und technischer Aspekte fest (siehe Abschnitt 6);
- legt einen Rahmen für die anzuwendende Policy sowie Mindestanforderungen an die geregelte Wiedererkennung fest (siehe Abschnitt 7);
- gibt einen Überblick über verschiedene Anwendungsfälle der Pseudonymisierung, die sowohl umkehrbar als auch unumkehrbar sein kann (siehe Anhang A);
- stellt einen Leitfaden für die Risikobewertung bei der Wiedererkennung zur Verfügung (siehe Anhang B);
- liefert ein Beispiel für ein System, das die Entpersonalisierung verwendet (siehe Anhang C);
- liefert informative Anforderungen an eine Interoperabilität für Pseudonymisierungsdienste (siehe Anhang D); und
- legt einen Rahmen für die anzuwendende Policy sowie Mindestanforderungen an vertrauenswürdige Praktiken für die Aktivitäten eines Pseudonymisierungsdienstes fest (siehe Anhang E).

2 Normative Verweisungen

Die folgenden Dokumente werden im Text in solcher Weise in Bezug genommen, dass einige Teile davon oder ihr gesamter Inhalt Anforderungen des vorliegenden Dokuments darstellen. Bei datierten Verweisungen gilt nur die in Bezug genommene Ausgabe. Bei undatierten Verweisungen gilt die letzte Ausgabe des in Bezug genommenen Dokuments (einschließlich aller Änderungen).

ISO 27799, *Health informatics — Information security management in health using ISO/IEC 27002*

3 Begriffe

Für die Anwendung dieses Dokuments gelten die folgenden Begriffe.

ISO und IEC unterhalten terminologische Datenbanken für die Verwendung in der Normung unter den folgenden Adressen:

- IEC Electropedia: unter <http://www.electropedia.org/>
- ISO Online browsing platform: unter <http://www.iso.org/obp>

3.1

Zugriffskontrolle

(en: access control)

Mittel zur Sicherstellung, dass auf die Ressourcen eines Datenverarbeitungssystems nur von dazu berechtigten Entitäten und in autorisierter Art und Weise zugegriffen werden kann

[QUELLE: ISO/IEC 2382:2015, 2126294]

3.2

Anonymisierung

(en: anonymization)

Prozess, bei dem *personenbezogene Daten* (3.37) entweder vom für die Verarbeitung der Daten Verantwortlichen allein oder in Zusammenarbeit mit anderen unumkehrbar so verändert werden, dass sich die betroffene Person danach weder direkt noch indirekt identifizieren lässt

Anmerkung 1 zum Begriff: Der Begriff ist absolut und kann in der Praxis schwierig umgesetzt werden.

[QUELLE: ISO/IEC 29100:2011, 2.2, modifiziert]

3.3

anonymisierte Daten

(en: anonymized data)

Daten (3.14), die als Ergebnis eines *Anonymisierungsprozesses* (3.2) erzeugt wurden

[QUELLE: ISO/IEC 29100:2011, 2.3, modifiziert]

3.4

anonymer Bezeichner

(en: anonymous identifier)

Bezeichner (3.27) einer Person, der die *Identifikation* (3.26) der betroffenen *natürlichen Person* (3.34) nicht erlaubt

3.5

Authentisierung

(en: authentication)

Beibringung eines Belegs für die behauptete Identität

3.6

Angreifer

(en: attacker)

Person, die bewusst Schwachstellen in technischen und nicht-technischen Sicherheitskontrollen nutzt, um Informationssysteme und Netzwerke zu stehlen oder zu beeinträchtigen oder die Verfügbarkeit für berechnete Benutzer des Informationssystems und Netzwerk-Ressourcen zu beeinträchtigen

[QUELLE: ISO/IEC 27033-1:2015, 3.3]

3.7

chiffrierter Text

(en: ciphertext)

Daten (3.14), die durch Anwendung einer Verschlüsselung erzeugt wurden, deren semantischer Inhalt nicht ohne Anwendung von kryptographischen Techniken verfügbar ist

[QUELLE: ISO/IEC 2382:2015, 2126285]

3.8**Vertraulichkeit**

(en: confidentiality)

Eigenschaft, dass *Informationen* (3.29) keinen unberechtigten Personen, Entitäten oder Prozessen zur Verfügung gestellt oder diesen offenbart werden

[QUELLE: ISO 7498-2:1989, 3.3.16]

3.9**Content-Encryption-Schlüssel**

(en: content-encryption key)

kryptographischer Schlüssel, der zur Verschlüsselung des Inhalts einer Kommunikation verwendet wird

3.10**für die Verarbeitung Verantwortlicher**

(en: controller)

natürliche oder juristische Person, Behörde, Einrichtung oder jede andere Stelle, die allein oder gemeinsam mit anderen über die Zwecke und Mittel der *Verarbeitung von personenbezogenen Daten* (3.40) entscheidet

3.11**Kryptographie**

(en: cryptography)

Disziplin, die Grundsätze, Mittel und Verfahren für die Umwandlung von *Daten* (3.14) verkörpert, um deren Informationsinhalte zu verbergen sowie ihre unentdeckte Änderung und/oder unautorisierte Verwendung zu verhindern

[QUELLE: ISO 7498-2:1989, 3.3.20]

3.12**kryptographischer Algorithmus****Verschlüsselungsalgorithmus**

(en: cryptographic algorithm)

⟨Chiffrierung⟩ Verfahren für die Umwandlung von *Daten* (3.14) mit dem Ziel, deren Informationsgehalt zu verbergen und ihre unerkannte Änderung und/oder nicht autorisierte Nutzung zu verhindern

3.13**Chiffrierschlüsselverwaltung****Schlüsselverwaltung**

(en: cryptographic key management, key management)

Erzeugung, Speicherung, Verteilung, Löschung, Archivierung und Anwendung von *Schlüsseln* (3.31) in Übereinstimmung mit einer *Sicherheits-Policy* (3.46)

[QUELLE: ISO 7498-2:1989, 3.3.33]

3.14**Daten**

(en: data)

wieder herleitbare Darstellung von *Informationen* (3.29) in einer formalisierten Art und Weise, die zur Kommunikation, Interpretation oder Verarbeitung geeignet ist

Anmerkung 1 zum Begriff: Daten können durch Menschen oder automatische Mittel verarbeitet werden.

[QUELLE: ISO/IEC 2382:2015, 2121272]

3.15

Datenintegrität

(en: data integrity)

Eigenschaft, dass *Daten* (3.14) nicht in unautorisierter Art und Weise verändert oder zerstört worden sind

[QUELLE: ISO 7498-2:1989, 3.3.21]

3.16

Datenverknüpfung

(en: data linking)

Abgleich und Zusammenführung von *Daten* (3.14) aus mehreren Datenbanken

3.17

Datenschutz

(en: data protection)

technisches und soziales System für die Aushandlung, Verwaltung und Sicherstellung der *Vertraulichkeit* (3.39) und Sicherheit von Informationen

3.18

betroffene Person

(en: data subject)

Person, auf die sich *Daten* (3.14) beziehen

3.19

Entschlüsselung

(en: decryption)

Prozess der Konvertierung von verschlüsselten *Daten* (3.14) zurück in die ursprüngliche verständliche Form

3.20

Entpersonalisierung

(en: de-identification)

allgemeine Benennung für jeden Prozess, der die Verbindung zwischen einer Menge von zur Identifizierung geeigneten *Daten* (3.14) und der *betroffenen Person* (3.18) verringert

3.21

zur direkten Identifizierung geeignete Daten

(en: directly identifying data)

Daten (3.14), anhand derer eine einzelne Person direkt identifizierbar ist

Anmerkung 1 zum Begriff: Direkte Bezeichner sind diejenigen *Daten*, die dazu verwendet werden können, eine Person ohne zusätzliche Informationen oder durch Quervernetzung anhand anderer öffentlich zugänglicher Informationen zu identifizieren.

3.22

Offenlegung

(en: disclosure)

Preisgabe von oder Ermöglichung des Zugriffs auf *Daten* (3.14)

Anmerkung 1 zum Begriff: Ob der Empfänger sich die *Daten* tatsächlich ansieht, sie zur Kenntnis nimmt oder sie aufbewahrt, ist ohne Bedeutung für die Frage, ob sie offengelegt wurden.

3.23

Verschlüsselung

(en: encryption)

Prozess der Konvertierung von *Informationen* (3.29) oder *Daten* (3.14) in eine Chiffre oder einen Code

3.24

Patientenbezeichner

(en: healthcare identifier, subject of care identifier)

Bezeichner (3.27) einer Person für die hauptsächliche Nutzung durch ein Gesundheitssystem

3.25

identifizierbare Person

(en: identifiable person)

jemand, der direkt oder indirekt identifiziert werden kann, insbesondere durch Zuordnung zu einer Kennnummer oder zu einem oder mehreren spezifischen Elementen, die Ausdruck ihrer physischen, physiologischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität sind

[QUELLE: Richtlinie 95/46/EG]

3.26

Identifizierung

(en: identification)

Prozess der Nutzung von behaupteten oder beobachteten Attributen einer Entität mit dem Ziel, diese aus den anderen Entitäten in einer Reihe von Identitäten auszusondern

Anmerkung 1 zum Begriff: Die Identifizierung einer Entität in einem bestimmten Kontext ermöglicht es einer anderen Entität, die Entitäten, mit denen sie interagiert, voneinander zu unterscheiden.

3.27

Bezeichner

(en: identifier)

Information (3.29), die vor einer möglichen Bestätigung durch einen entsprechenden Authentifikator verwendet wird, um eine Identität zu beanspruchen

[QUELLE: ENV 13608-1:2000, 3.44]

3.28

zur indirekten Identifizierung geeignete Daten

(en: indirectly identifying data)

Daten (3.14), anhand deren eine einzelne Person nur dann identifizierbar ist, wenn sie in Verbindung mit anderen zur indirekten Identifizierung geeigneten Daten verwendet werden

Anmerkung 1 zum Begriff: Indirekte Bezeichner können die Grundgesamtheit, zu der die Person gehört, verkleinern, und zwar bei kombinierter Verwendung unter Umständen bis auf eins.

BEISPIEL Postleitzahl, Geschlecht, Alter, Geburtsdatum.

3.29

Information

(en: information)

Wissen in Bezug auf Objekte, das innerhalb eines bestimmten Kontexts eine spezielle Bedeutung hat

[QUELLE: ISO/IEC 2382:2015, 2121271, modifiziert]

3.30

Unumkehrbarkeit

(en: irreversibility)

Situation, in der es für einen beliebigen Übergang von identifizierbar zu pseudonym rechnerisch unmöglich ist, vom *Pseudonym* (3.43) auf den ursprünglichen *Bezeichner* (3.27) zu schließen

3.31

Schlüssel

(en: key)

Symbolfolge, die die Operationen der *Verschlüsselung* (3.23) und der *Entschlüsselung* (3.19) steuert

[QUELLE: ISO 7498-2:1989, 3.3.32]

3.32

Verknüpfung von Informationsobjekten

(en: linkage of information objects)

Prozess, der die Herstellung einer logischen Verbindung zwischen verschiedenen Informationsobjekten ermöglicht

3.33

longitudinale oder lebenslange Gesundheitsakte

permanente koordinierte Aufzeichnung wesentlicher Informationen in chronologischer Reihenfolge

Anmerkung 1 zum Begriff: Sie kann alle historisch erfassten Daten beinhalten oder abgerufen werden als vom Benutzer bestimmte Zusammenfassung signifikanter demographischer, genetischer, klinischer und umweltbezogener Fakten und Ereignisse, die in einem automatischen System verwaltet werden.

[QUELLE: ISO/TR 21089:2004, 3.61, modifiziert]

3.34

natürliche Person

(en: natural person)

echter Mensch im Gegensatz zu einer juristischen Person, bei der es sich um eine private oder öffentliche Organisation handeln kann

3.35

Personenidentifizierung

(en: person identification)

Prozess der Herstellung einer Verbindung zwischen einem Informationsobjekt und einer physischen Person

3.36

Personenbezeichner

(en: personal identifier)

Informationen, deren Zweck darin besteht, eine Person innerhalb eines bestimmten Kontexts eindeutig zu identifizieren

3.37

personenbezogene Daten

(en: personal data)

Informationen über eine bestimmte oder bestimmbare *natürliche Person* (3.34) („betroffene Person“)

[QUELLE: Richtlinie 95/46/EG]

3.38

Hauptnutzung personenbezogener Daten

(en: primary use of personal data)

Nutzungen und *Offenlegungen* (3.22), die für die erfassten *Daten* (3.14) beabsichtigt sind

3.39**Privatsphäre**

(en: privacy)

Nichtstattfinden eines Eindringens in das Privatleben oder die Angelegenheiten einer Person, sofern dieses Eindringen aus der unzulässigen oder illegalen Erfassung und Nutzung von *Daten* (3.14) über diese Person resultiert

[QUELLE: ISO/IEC 2382:2015, 2126263]

3.40**Verarbeitung personenbezogener Daten**

(en: processing of personal data)

mit oder ohne Hilfe automatisierter Verfahren ausgeführter Vorgang oder ausgeführte Vorgangsreihe im Zusammenhang mit *personenbezogenen Daten* (3.37) wie das Erheben, das Speichern, die Organisation, die Aufbewahrung, die Anpassung oder Veränderung, das Auslesen, das Abfragen, die Benutzung, die Weitergabe durch Übermittlung, Verbreitung oder jede andere Form der Bereitstellung, der Abgleich oder die Verknüpfung sowie das Sperren, Löschen oder Vernichten

[QUELLE: Richtlinie 95/46/EG]

3.41**Auftragsverarbeiter**

(en: processor)

natürliche oder juristische Person, Behörde, Einrichtung oder jede andere Stelle, die *personenbezogene Daten* (3.37) im Auftrag des *für die Verarbeitung Verantwortlichen* (3.10) verarbeitet

Anmerkung 1 zum Begriff: Siehe Richtlinie 95/46/EG des Europäischen Parlaments und des Rates vom 24. Oktober 1995 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr.

3.42**Pseudonymisierung**

(en: pseudonymization)

spezielle Art der *Entpersonalisierung* (3.20), die sowohl die Verbindung mit einer *betroffenen Person* (3.18) entfernt und eine Verbindung zwischen einem bestimmten Satz von Eigenschaften in Bezug auf die betroffene Person und einem oder mehreren *Pseudonymen* (3.43) hinzufügt

3.43**Pseudonym**

(en: pseudonym)

Personenbezeichner (3.36), der sich vom üblicherweise verwendeten Personenbezeichner unterscheidet und mit pseudonymisierten Daten verwendet wird, um für Kohärenz innerhalb des Datensatzes zu sorgen, die alle Informationen über die betroffene Person miteinander verknüpft, ohne die Identität dieser Person in der realen Welt offenzulegen

Anmerkung 1 zum Begriff: Dies kann in umkehrbarer oder unumkehrbarer Art und Weise entweder aus dem üblicherweise verwendeten Personenbezeichner abgeleitet oder völlig ohne Zusammenhang sein.

Anmerkung 2 zum Begriff: Der Begriff des Pseudonyms wird üblicherweise auf einen Bezeichner beschränkt, der eine direkte Ableitung aus dem normalen Personenbezeichner nicht zulässt. Derartige pseudonyme Informationen sind daher funktionell anonym. Unter Umständen kann ein vertrauenswürdiger Dritter in der Lage sein, den normalen Personenbezeichner aus dem Pseudonym abzuleiten.

3.44**Empfänger**

(en: recipient)

natürliche oder juristische Person, Behörde, Einrichtung oder jeder andere Stelle, die *Daten* (3.14) erhält

3.45

sekundäre Nutzung von personenbezogenen Daten

(en: secondary use of personal data)

Nutzungen und *Offenlegungen* (3.22), die sich vom ursprünglich vorgesehenen Verwendungszweck der erfassten *Daten* (3.14) unterscheiden

3.46

Sicherheits-Policy

(en: security policy)

Aktionsplan oder Vorgehensweise für die Sicherstellung der Rechtersicherheit

[QUELLE: ISO/IEC 2382:2015, 2126246]

3.47

vertrauenswürdiger Dritter

(en: trusted third party)

Sicherheitsinstanz oder deren Vertreter, die bzw. der von einer anderen Entität mit Aktivitäten in Bezug auf die Sicherheit betraut wurde

[QUELLE: ISO/IEC 18014-1:2008, 3.20]

4 Abkürzungen

DICOM Digitale Bildbearbeitung und Kommunikation in der Medizin (en: Digital Imaging and Communication in Medicine)

HIPAA Health Insurance Portability and Accountability Act

GIS Gesundheitsinformationssystem (en: Health Information System, HIS)

HIV Humanes Immundefizienz-Virus (en: Human Immunodeficiency Virus)

IP Internetprotokoll (en: Internet Protocol)

VoV Opfer der Anwendung von Gewalttaten (en: Victim of Violence use)

5 Anforderungen an den Schutz von personenbezogenen Daten im Gesundheitswesen

5.1 Ziele des Schutzes von personenbezogenen Daten

Das Ziel des Schutzes von personenbezogenen Daten als Teil des mit der Vertraulichkeit verbundenen Ziels der Sicherheit ist, jede unbefugte oder unerwünschte Offenlegung von Informationen über eine Person zu verhindern, die in der Folge Einfluss auf rechtliche, organisatorische und/oder finanzielle Risikofaktoren haben kann. Der Schutz von personenbezogenen Daten ist ein Teilbereich des allgemeinen Datenschutzes, der der Definition nach weitere sensibel auf Verletzungen des Datenschutzes reagierende Entitäten, wie z. B. Organisationen, einschließt. Da der Schutz von personenbezogenen Daten der am besten geregelte und umfassendste ist, konzentriert sich dieses Begriffsmodell darauf. Auf den Schutz von personenbezogenen Daten ausgerichtete Lösungen lassen sich auch auf den Datenschutz anderer Entitäten bzw. betroffener Personen übertragen. Dies kann in Ländern von Nutzen sein, in denen der Schutz der Daten von betroffenen Personen oder Organisationen gesetzlich geregelt ist.

Der Schutz von personenbezogenen Daten hat zwei Ziele, nämlich einerseits den Schutz der personenbezogenen Daten beim Umgang mit Onlineanwendungen (wie z. B. Browsen im Internet) und andererseits den Schutz von in Datenbanken gespeicherten personenbezogenen Daten. Dieses Dokument befasst sich nur mit dem letzteren Ziel.

Daten können aus Datenbanken extrahiert werden. Das Ziel ist, das Risiko der Offenlegung der Identität der betroffenen Personen zu verringern. Forscher arbeiten mit „Fällen“, d.h. mit Langzeit-Patientendaten, die über bestimmte Zeiträume erfasst wurden und/oder aus verschiedenen Quellen stammen. Für die Zusammenfassung unterschiedlicher Datenelemente in derartigen Fällen ist es jedoch notwendig, ein Verfahren anzuwenden, das eine Kumulierung ohne Gefährdung des Schutzes der personenbezogenen Daten der betroffenen Personen ermöglicht. Dies kann durch Pseudonymisierung der Daten erreicht werden.

Die Entpersonalisierung wird angewendet, um in einer Vielzahl von unterschiedlichen Situationen Risiken für den Schutz von personenbezogenen Daten zu verringern.

Die am weitesten gehende Entpersonalisierung erfolgt bei Schulungsmaterialien, die einer breiten Öffentlichkeit zur Verfügung gestellt werden, denen sich aber dennoch Einzelheiten entnehmen lassen sollten, um für die Ausbildung von Medizinerinnen von Nutzen zu sein (Es gibt ein IHE-Profil für die Automatisierungsunterstützung für die Ausführung dieser Art von Entpersonalisierung. Ein großer Teil dieses Prozesses wird je nach Bedarf an den einzelnen Patienten und dem Schulungszweck entsprechend angepasst.).

Das öffentliche Gesundheitswesen nutzt entpersonalisierte Datenbanken, um Krankheiten und deren Ausbreitung nachzuverfolgen und zu verstehen.

In klinischen Studien wird entpersonalisiert, sowohl um personenbezogene Daten zu schützen als auch um unbewusste systematische Abweichungen zu verhindern, indem andere Informationen, wie z.B. ob der Patient ein Placebo oder ein experimentelles Medikament erhalten hat, entfernt werden.

In geringem Maße wird in vielen klinischen Berichten entpersonalisiert, wobei dem Leser gegenüber die Identität des behandelnden Arztes, des Krankenhauses, des Patienten usw. geheim gehalten werden, um sowohl Risiken für den Schutz von personenbezogenen Daten zu verringern als auch um unbewusste systematische Abweichungen zu eliminieren. Diese Art der Entpersonalisierung verhindert nur, dass dem Leser irgendwelche Informationen zufällig offengelegt werden. Jedes planmäßige Vorgehen führt zur alsbaldigen Aufdeckung der Patientenidentität usw.

Bei der Erstellung von Statistiken oder Analysen zur Arbeitsbelastung in Krankenhäusern oder in Bezug auf Behandlungen, die im Rahmen von Verträgen mit Beauftragten oder Kunden von Pflegediensten ausgeführt werden, ist es notwendig, einzelne Patienten aussondern zu können, ohne deren Identität im Einzelnen zu kennen. Dies ist ein Beispiel für die Entpersonalisierung in Geschäftszusammenhängen.

Der Prozess der Risikoschichtung (z.B. erneuter Krankenhausaufenthalt) kann durch Nutzung von Datensätzen aus den primären und sekundären Patientenpflegediensten erfolgen. Die Datensätze werden für die Analyse entpersonalisiert, falls aber Patienten als hochgradig gefährdet erkannt werden, können diese von einem entsprechenden Mediziner wieder identifiziert werden, um Folgemaßnahmen einleiten zu können. Für Einzelheiten zur Entpersonalisierung im Gesundheitswesen siehe Anhang A.

5.2 Allgemeines

Entpersonalisierung ist der allgemeine Begriff für jeden Prozess der Reduktion der Zuordnung zwischen einer Menge von zur Identifizierung geeigneten Daten und der betroffenen Person für einen oder mehrere vorgesehene Verwendungszwecke des resultierenden Datenbestandes. Pseudonymisierung ist eine Unterkategorie der Entpersonalisierung. Das Pseudonym ist das Mittel, mit dessen Hilfe pseudonymisierte Daten der betreffenden Person oder den betreffenden Informationssystemen verknüpft werden, ohne die Identität der Person offenzulegen. Entpersonalisierung kann inhärent den Nutzen des resultierenden Datenbestandes einschränken. Die Pseudonymisierung kann mit oder ohne die Möglichkeit der Wiedererkennung (en: re-identification) der betroffenen Person erfolgen (umkehrbare oder unumkehrbare Pseudonymisierung). Es gibt mehrere Szenarien für die Anwendung der Pseudonymisierung im Gesundheitswesen, wobei sie besonders bei der zunehmenden elektronischen Verarbeitung von Patientendaten im Verbund mit steigenden Erwartungen der Patienten hinsichtlich des Datenschutzes Anwendung findet. In Anhang A werden mehrere Beispiele gegeben.

Es ist wichtig, zu beachten, dass über die gesamte Dauer der Existenz von pseudonymisierten Daten ein Risiko der unbefugten Wiedererkennung besteht. Insofern besteht Ähnlichkeit mit dem Verfahren der Verschlüsselung, weil sich mit genügend Aufwand letztlich jede Verschlüsselung knacken lässt, das Ziel aber ist, dies so schwierig zu gestalten, dass die damit verbundenen Kosten untragbar werden. In Bezug auf die Entpersonalisierung liegen weniger Erfahrungen vor als in Bezug auf die Verschlüsselung, weswegen die damit verbundenen Risiken nicht genauso gut bekannt sind.

5.3 Entpersonalisierung als Prozess der Risikominderung

5.3.1 Allgemeines

Im Prozess der Entpersonalisierung sollten die für den resultierenden Datenbestand geltenden Sicherheits- und Datenschutzregeln berücksichtigt werden. Nur selten lässt sich das Risiko so weit vermindern, dass keine weiter gehenden Befugniscontrollen für den Datenbestand erforderlich sind.

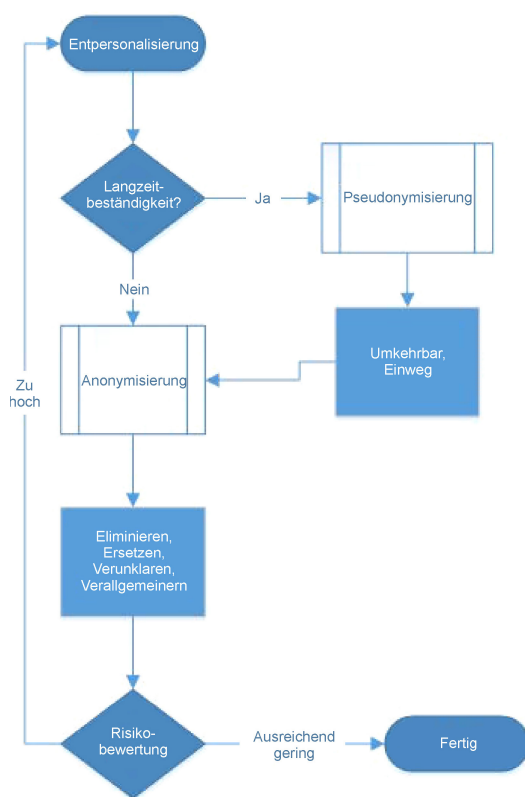


Bild 1 — Visualisierung des Entpersonalisierungsprozesses

Bild 1 ist ein informatives Diagramm zur bildlichen Darstellung dieses Entpersonalisierungsprozesses. Es zeigt, dass an der Spitze der Begriff der Entpersonalisierung als Prozess steht. Bei diesem Prozess kommen Teilprozesse, nämlich die Pseudonymisierung und/oder die Anonymisierung, zur Anwendung. Für diese Teilprozesse wiederum werden verschiedene speziell für den betreffenden Datenelementtyp und das betreffende Risikominderungsverfahren entwickelte Tools (Hilfsprogramme) angewendet.

Der Ausgangszustand ist, dass die Menge der Daten, die das System durchlaufen dürfen, gleich null ist. Jedes Element sollte anhand des vorgesehenen Verwendungszwecks des resultierenden Datenbestandes begründet werden. Der vorgesehene Verwendungszweck des Datenbestandes hat erheblichen Einfluss auf den Entpersonalisierungsprozess.

5.3.2 Pseudonymisierung

Falls Langzeitbeständigkeit erforderlich ist, kann die Pseudonymisierung wirksam zur Entpersonalisierung eingesetzt werden. Dies kann dazu dienen, eine Menge von Datensätzen zusammenzuhalten, die miteinander verbunden sein sollten, falls sie ohne diese Langzeitbeständigkeit voneinander getrennt werden könnten. Dies ist von Nutzen, wenn sämtliche Datensätze zu einem Patienten unter einem Pseudonym zusammengehalten werden sollen. Die Pseudonymisierung kann auch dazu dienen, sicherzustellen, dass jedes Mal, wenn Daten in einen entpersonalisierten Datensatz extrahiert werden, neue Einträge ebenfalls mit demselben Pseudonym verbunden werden. Bei der Pseudonymisierung kann der angewendete Algorithmus entweder absichtlich umkehrbar oder absichtlich unumkehrbar sein. Im Falle der Umkehrbarkeit kann eine geheime Nachschlagtabelle verwendet werden, um, sofern genehmigt, die ursprüngliche Identität aufzudecken. In einem nicht umkehrbaren Schema kann im Verlauf des Prozesses eine temporäre Tabelle verwendet werden, die nach Abschluss des Prozesses vernichtet wird.

5.3.3 Anonymisierung

Bei der Anonymisierung handelt es sich um den Prozess und die Tools, die anzuwenden sind, wenn keine Langzeitbeständigkeit erforderlich ist. Der Anonymisierungsprozess wird auch dann angewendet, wenn eine Pseudonymisierung vorgenommen wurde, um die verbleibenden Datenattribute zu behandeln. Bei der Anonymisierung werden Tools, wie redaktionelle Bearbeitung, Löschen, Ausblenden, Ersetzen, Randomisieren, Verschieben, Verzerren, Kappen, Gruppieren usw., angewendet. Anonymisierung kann zu einer verringerten Verknüpfungsmöglichkeit führen.

Für jedes Element, dem ein Durchlaufen des Systems gestattet wird, sollte eine Begründung gegeben werden. Jedes Element sollte das beim vorgesehenen Verwendungszweck des resultierenden Datenbestandes kleinste mögliche Risiko darstellen. Auf diese Art kann, falls der vorgesehene Verwendungszweck des resultierenden Datenbestandes keine feinkörnigen Codes erfordert, eine Gruppierung von Codes vorgenommen werden.

5.3.4 Direkte und indirekte Bezeichner

Im Entpersonalisierungsprozess werden drei Arten von Daten angesprochen: direkte Bezeichner, anhand derer sich der Patient direkt identifizieren lässt; indirekte Bezeichner, die bei der Anwendung in Verbindung mit anderem indirektem oder externem Wissen eine Korrelation ermöglichen; und nicht identifizierende Daten, d.h. alle anderen Daten.

Üblicherweise wird der Entpersonalisierungsprozess auf einen Datenbestand angewendet, der aus Einträgen besteht, die viele Attribute haben können; z. B. ein Arbeitsblatt mit nach Spalten geordneten Datenzeilen.

Der Entpersonalisierungsprozess wird einschließlich der Pseudonymisierung und Anonymisierung auf alle Daten angewendet. Die Pseudonymisierung wird im Allgemeinen auf direkte Bezeichner angewendet, kann aber auch auf indirekte Bezeichner angewendet werden, sofern dies für die Risikominderung bei Fortbestehen der Langzeiterfordernisse des vorgesehenen Verwendungszwecks des resultierenden Datenbestandes angemessen ist. Anonymisierungstools werden bei allen Datenformen angewendet, sofern dies zur Risikominderung angemessen ist.

5.4 Schutz der personenbezogenen Daten von Entitäten

5.4.1 Vergleich von personenbezogenen und entpersonalisierten Daten

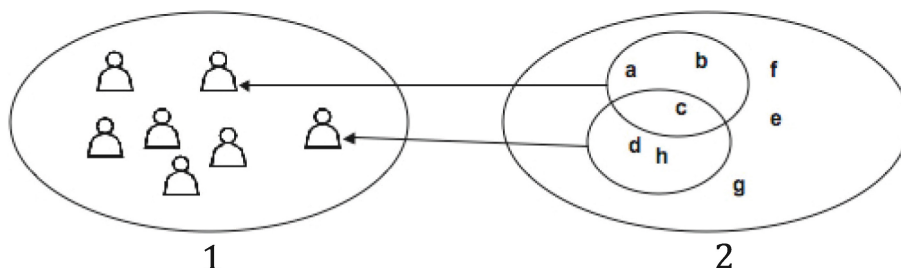
5.4.1.1 Definition der personenbezogenen Daten

Nach Literaturhinweis [18] sind unter „personenbezogenen Daten“ alle Informationen über eine bestimmte oder bestimmbare natürliche Person („betroffene Person“) zu verstehen; als bestimmbar wird eine Person angesehen, die direkt oder indirekt identifiziert werden kann, insbesondere durch Zuordnung zu einer

Kennnummer oder zu einem oder mehreren spezifischen Elementen, die Ausdruck ihrer physischen, physiologischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität sind.

Dieser Begriff wird in anderen nationalen Gesetzen behandelt, die dieselben Grundsätze anwenden, die in dieser Definition zu finden sind (z. B. HIPAA).

5.4.1.2 Idealisiertes Konzept von Identifizierung und Entpersonalisierung



Legende

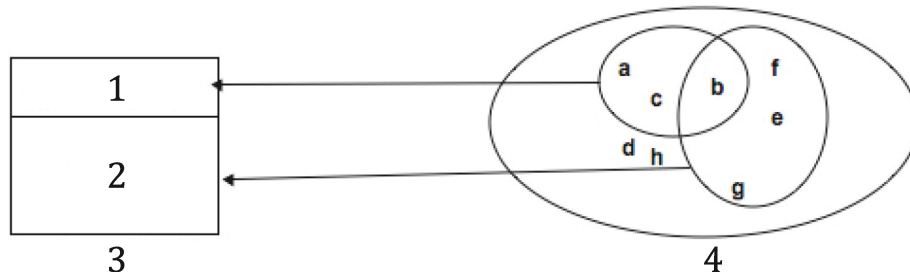
- 1 Satz von betroffenen Personen
- 2 Satz von Eigenschaften

Bild 2 — Identifizierung von betroffenen Personen

In diesem Unterabschnitt wird ein idealisiertes Konzept von Identifizierung und Entpersonalisierung beschrieben. Es wird davon ausgegangen, dass es außerhalb des Modells, wie in Bild 2 als Beispiel gezeigt, keine Daten gibt, die sich z. B. mit im Modell genutzten Daten verknüpfen lassen, um betroffene Personen (indirekt) zu identifizieren.

In 5.4.1 werden mögliche außerhalb des Datenmodells befindliche Informationsquellen berücksichtigt. Dies ist erforderlich, um die Wiedererkennungsrisiken erörtern zu können. Informations- und Kommunikationstechnologieprojekte stellen Daten, die nicht innerhalb des Modells verwendet werden, bei der Abdeckung von Aspekten des Funktionsdesigns niemals bildlich dar. Wenn sich Kritiker jedoch auf die Identifizierbarkeit konzentrieren, führen sie Informationen ein, die von einem Angreifer abgezapft werden können, um betroffene Personen zu identifizieren oder weitere Informationen über sie (z. B. zur Zugehörigkeit zu einer Gruppe) zu erhalten.

Wie in Bild 1 dargestellt, hat eine betroffene Person eine Reihe von Eigenschaften (z. B. Name, Geburtsdatum, medizinische Daten), die in einer medizinischen Datenbank gespeichert werden und bei denen es sich um personenbezogene Daten der betroffenen Person handelt. Innerhalb einer Gruppe von betroffenen Personen ist eine betroffene Person dann identifiziert, wenn sie ausgesondert werden kann. Das bedeutet, dass sich eine Reihe von mit der betroffenen Person verbundenen Eigenschaften ermitteln lässt, anhand derer sich die betreffende Person eindeutig identifizieren lässt. In einigen Fällen genügt bereits eine einzige Eigenschaft, um die betroffene Person zu identifizieren (z. B. wenn es sich bei der Zahl um eine eindeutige nationale Personenkennziffer handelt). In anderen Fällen sind mehrere Eigenschaften erforderlich, um eine betroffene Person auszusondern, wie z. B. dann, wenn die Anschrift eines unter derselben Anschrift gemeldeten Familienmitgliedes verwendet wird. Einige Zuordnungen zwischen Eigenschaften und betroffenen Personen (z. B. Geburtsdatum, Geburtsort) sind dauerhafter als andere (wie z. B. eine E-Mail-Adresse).



Legende

- 1 zur Identifizierung geeignete Daten
- 2 Nutzdaten
- 3 personenbezogene Daten
- 4 Satz von Eigenschaften

Bild 3 — Trennung von personenbezogenen Daten und Nutzdaten

Vom Begriffsstandpunkt aus lassen sich personenbezogene Daten nach den Kriterien der Identifizierbarkeit in zwei Kategorien einteilen (siehe Bild 3):

- Nutzdaten: der Teil der Daten, der Eigenschaften enthält, die keine eindeutige Identifizierung der betroffenen Person ermöglichen; begrifflich enthalten die Nutzdaten anonyme Daten (z.B. klinische Messwerte, maschinell erhaltene Messwerte);
- zur Identifizierung geeignete Daten: der zur Identifizierung geeignete Teil der Daten, der eine Reihe von Eigenschaften enthält, die eine eindeutige Identifizierung der betroffenen Person ermöglichen (z. B. demographische Daten).

Es ist zu beachten, dass die begriffliche Unterscheidung zwischen „zur Identifizierung geeigneter Daten“ und „Nutzdaten“ zu Widersprüchen führen kann. Dies ist dann der Fall, wenn zur direkten Identifizierung geeignete Daten als „Nutzdaten“ betrachtet werden. Jedes Pseudonymisierungsverfahren sollte darauf ausgerichtet sein, die Menge der zur direkten Identifizierung geeigneten Daten z. B. durch Zusammenfassung dieser Daten in Gruppen zu verringern. In bestimmten Fällen, in denen dies nicht möglich ist (wie z. B. beim Geburtsdatum von Säuglingen), sollte auf das damit verbundene Risiko im Policy-Dokument hingewiesen werden. In einem der folgenden Abschnitte dieses Dokuments wird die Aufteilung der Daten in den Nutzdaten-Teil und den Teil der zur Identifizierung geeigneten Daten statt vom Begriffsstandpunkt von einem praktischen Standpunkt aus behandelt. Vom Begriffsstandpunkt aus genügt es, dass eine solche Unterscheidung möglich ist. Es ist wichtig, zu beachten, dass die Unterscheidung zwischen zur Identifizierung geeigneten Eigenschaften und Nutzdaten nicht absolut ist. Einige der auch zur Identifizierung geeigneten Daten können für die Forschung notwendig sein, so z. B. das Jahr und der Monat der Geburt. Diese Unterscheidungen werden im Folgenden behandelt.

5.4.2 Konzept der Pseudonymisierung

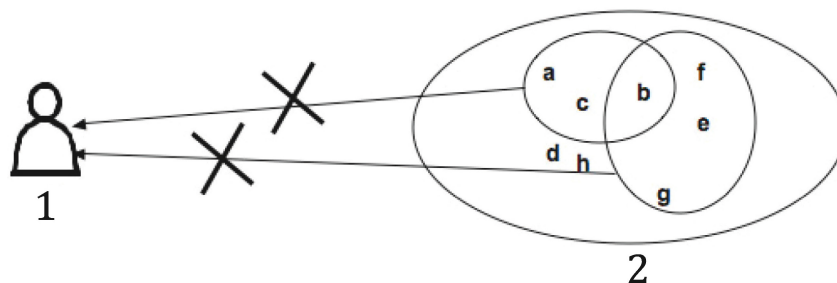
Praxis und Fortschritt in der Medizin erfordern die Freigabe von Bestandteilen privater Patientenakten für die Zwecke von Lehre, Forschung und Qualitätskontrolle und für weitere Zwecke. Sowohl aus wissenschaftlichen Gründen als auch im Sinne des Schutzes von personenbezogenen Daten müssen diese Datensatzelemente modifiziert werden, um die Identitäten der betroffenen Personen geheim zu halten.

Es gibt kein Entpersonalisierungsverfahren, das allein in der Lage wäre, sämtliche unterschiedlichen Erfordernisse der diversen medizinischen Nutzungen abzudecken und dabei noch die betreffenden Identitäten geheim zu halten. Jeder Datensatzfreigabeprozess muss einer Risikobewertung unterzogen werden, um Folgendes zu bewerten:

- a) den Zweck der Datenfreigabe (z. B. Analyse);
- b) die für diesen Zweck mindestens freizugebenden Informationen;
- c) die Risiken der Offenlegung (einschließlich Wiedererkennung);
- d) die Klassifizierung der Information (z. B. Kennzeichnung oder Etikettierung);
- e) welche Freigabestrategien zur Verfügung stehen.

Anhand dieser Informationen — der Einzelheiten des Freigabeprozesses und der Risikobewertung — ist eine Strategie für die Identitätsgeheimhaltung festzulegen. Diese Festlegung muss für jeden neuen Freigabeprozess getroffen werden, obgleich für viele verschiedene Freigabeprozesse eine gemeinsame Freigabestrategie und gemeinsame Einzelheiten zur Freigabe gewählt werden können. Die meisten der in der Lehre verwendeten Dateien haben gemeinsame Eigenschaften hinsichtlich des Zwecks und des Mindestinformationsgehalts. Vielen klinischen Arzneimittelstudien liegt bei veränderlichen Einzelheiten dieselbe Strategie zugrunde. Die Entpersonalisierung deckt mehr Erfordernisse als nur den Schutz der Vertraulichkeit ab. Häufig erfordern auch Blind- oder Doppelblindstudien eine Entpersonalisierung zur Verblindung. Dies hat Einfluss auf die Entscheidungen zu den Freigabeverfahren.

Dieser Unterabschnitt enthält die Terminologie, die zur Beschreibung der Geheimhaltung von zur Identifizierung geeigneten Informationen angewendet wird.



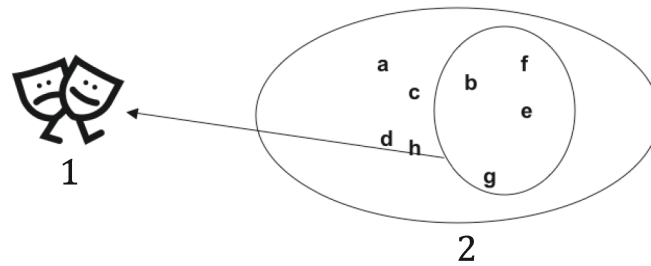
Legende

- 1 betroffene Person
- 2 Satz von Eigenschaften

Bild 4 — Anonymisierung

Die Anonymisierung (siehe Bild 4) ist der Prozess, der die Zuordnung zwischen dem zur Identifizierung geeigneten Datenbestand und der betroffenen Person aufhebt. Dies kann auf zwei verschiedene Weisen erfolgen:

- durch Eliminierung oder Umwandlung von Eigenschaften im zugehörigen Eigenschaftsdatenbestand, so dass die Zuordnung nicht mehr eindeutig ist und sich auf mehr als eine betroffene Person beziehen kann und keine direkte Beziehung zu einer Person bestehen bleibt;
- durch Vergrößerung der Grundgesamtheit der betroffenen Personen, sodass die Zuordnung zwischen dem Datenbestand und der betroffenen Person nicht mehr eindeutig ist und keine direkte Beziehung zu einer Person bestehen bleibt.



Legende

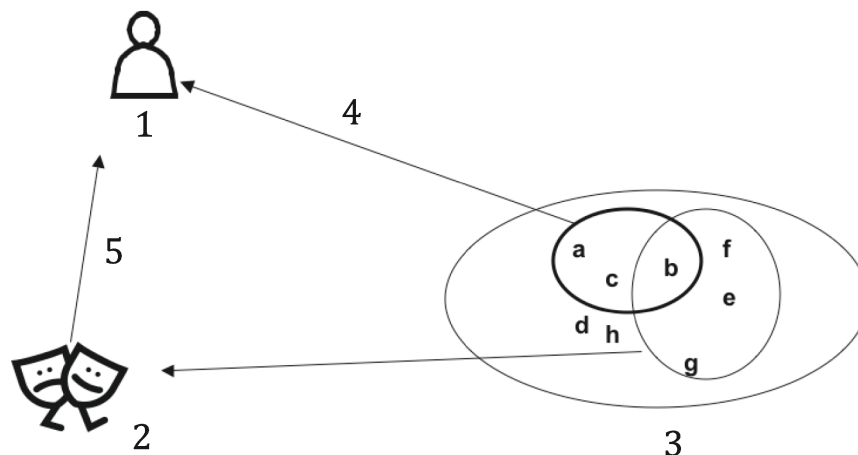
- 1 Pseudonym(e)
- 2 Satz von Eigenschaften

Bild 5 — Pseudonymisierung

Die Pseudonymisierung (siehe Bild 5) hebt die Zuordnung zu einer bestimmten betroffenen Person auf und fügt eine Zuordnung zwischen einem bestimmten Satz von Eigenschaften zur betroffenen Person und einem oder mehreren Pseudonymen hinzu.

Von einem funktionellen Standpunkt aus betrachtet, können pseudonyme Datenbestände zugeordnet werden, weil die Pseudonyme Zuordnungen zwischen Eigenschaftsgruppen, nicht aber die Zuordnung zur betroffenen Person ermöglichen. Auf diese Weise können z. B. Langzeitstudien durchgeführt werden, um auf realen Patientendaten basierende Fälle zusammenzutragen und die Identität der betreffenden Patienten zu schützen.

Im Falle der unumkehrbaren Pseudonymisierung enthält das Begriffsmodell kein Verfahren für die Ableitung einer Zuordnung zwischen der betroffenen Person und dem Eigenschaftsbestand aus dem Pseudonym.



Legende

- 1 betroffene Person
- 2 Pseudonyme
- 3 Satz von Eigenschaften
- 4 a) abgeleitet von
- 5 b) abgeleitet von

Bild 6 — Umkehrbare Pseudonymisierung

Bei der umkehrbaren Pseudonymisierung (siehe Bild 6) umfasst das Begriffsmodell eine Möglichkeit, wie sich der Datenbestand der betroffenen Person wieder zuordnen lässt.

Für dieses Ziel stehen zwei Verfahren zur Verfügung:

- a) Ableitung aus den Nutzdaten; dies kann z. B. bei gemeinsamer Verschlüsselung von zur Identifizierung geeigneten Informationen mit den Nutzdaten der Fall sein;
- b) Ableitung aus dem Pseudonym oder mittels einer Nachschlagetabelle.

Die umkehrbare Pseudonymisierung lässt sich auf verschiedene Weisen umsetzen, wobei davon ausgegangen wird, dass die Umkehrung der Pseudonymisierung nur unter geregelten Umständen und von einer entsprechend berechtigten Entität vorgenommen werden sollte. Das Policy-Rahmenwerk zur Wiedererkennung wird in Abschnitt 7 beschrieben. Im Gegensatz zur unumkehrbaren Pseudonymisierung erfordert die umkehrbare Pseudonymisierung üblicherweise einen erhöhten Schutz der Entität, die die Pseudonymisierung ausführt.

Anonymisierte Daten unterscheiden sich von pseudonymisierten Daten, weil pseudonymisierte Daten ein Verfahren zur Gruppierung von Daten enthalten, das auf Kriterien beruht, die aus den personenbezogenen Daten abgeleitet werden, von denen sie abgeleitet wurden.

5.5 Pseudonymisierung in der realen Welt

5.5.1 Begründung

In 5.4 ist der begriffliche Ansatz für die Pseudonymisierung dargelegt, wobei Begriffe wie z. B. „zugeordnet“, „identifizierbar“, „pseudonym“ usw. als absolut angesehen werden. In der Praxis lässt sich das Risiko der Wiedererkennung von Datenbeständen häufig nur schwer beurteilen. Daher werden die Begriffe der Pseudonymisierung und der unerwünschten/unbeabsichtigten Identifizierbarkeit in diesem Unterabschnitt verfeinert. Als Ausgangspunkt wird an dieser Stelle die Europäische Datenschutzschutzrichtlinie in Bezug genommen.

Es gibt in vielen Rechtssystemen diverse Vorschriften, die für verschiedene Zwecke die Erstellung von entpersonalisierten Daten verlangen. Es gibt auch Vorschriften, die den Schutz von privaten Informationen verlangen, ohne die dafür anzuwendenden Mechanismen festzulegen. Diese Vorschriften verwenden in der Regel aufwands- oder schwierigkeitsbezogene Ausdrücke, was wegen des aufgrund der in Entwicklung befindlichen Entpersonalisierungstechnologien rasch veränderlichen Schwierigkeitsgrades auch angemessen ist.

Formulierungen wie „alle Mittel (...), die vernünftigerweise (...) eingesetzt werden könnten“ (en: „all the means likely reasonable“) und „von einem Dritten“ (en: „by any other person“) sind noch zu vage. Da die Definitionen von „identifizierbar“ und „pseudonym“ von einem nicht genau festgelegten Verhalten („alle Mittel (...), die vernünftigerweise (...) eingesetzt werden könnten“) und von nicht genau festgelegten Akteuren („von einem Dritten“) abhängen, sollte das Begriffsmodell in diesem Dokument „vernünftige“ Annahmen über „alle Mittel (...), die vernünftigerweise (...) von einem Dritten (für die Zuordnung von Eigenschaften zu betroffenen Personen) eingesetzt werden könnten“ einschließen.

Das Begriffsmodell wird künftig noch verfeinert werden, um den Unterschieden in Bezug auf die Identifizierbarkeit Rechnung zu tragen, und in das Begriffsmodell wird „Beobachtungsdatenbanken“ und „Angreifer“ einbezogen werden.

5.5.2 Stufen des Schutzes von personenbezogenen Daten

5.5.2.1 Allgemeines

Den aktuellen Definitionen fehlt es bei der Beschreibung von Begriffen wie „pseudonym“ oder „identifizierbar“ an Genauigkeit. Es ist unrealistisch, anzunehmen, dass sich alle Ungenauigkeiten in der Terminologie beseitigen ließen, weil Pseudonymisierung immer auch Gegenstand von Statistik ist. Das