Arbeitshilfe zur Pseudonymisierung/Anonymisierung

Arbeitsgruppe Datenschutz

Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS)
Arbeitsgruppe „Datenschutz und IT-Sicherheit im Gesundheitswesen“

Autoren (alphabetisch)

Sonja Holst	Charité - Universitätsmedizin Berlin
Bernd Schütze	Deutsche Telekom Healthcare and Security GmbH
Gerald Spyra	Ratajczak und Partner mbB Rechtsanwälte

Stand: 29. Juni 2018

Haftungsausschluss

Das vorliegende Werk ist nach bestem Wissen erstellt, der Inhalt wurde von den Autoren mit größter Sorgfalt zusammengestellt. Dennoch ist diese Ausarbeitung nur als Standpunkt der Autoren aufzufassen, eine Haftung für die Angaben übernehmen die Autoren nicht. Die in diesem Werk gegebenen Hinweise dürfen daher nicht direkt übernommen werden, sondern müssen vom Leser für die jeweilige Situation anhand der geltenden Vorschriften geprüft und angepasst werden.

Die Autoren sind bestrebt, in allen Publikationen die Urheberrechte der verwendeten Texte zu beachten, von ihnen selbst erstellte Texte zu nutzen oder auf lizenzfreie Texte zurückzugreifen.

Alle innerhalb dieses Dokumentes genannten und ggf. durch Dritte geschützten Marken- und Warenzeichen unterliegen uneingeschränkt den Bestimmungen des jeweils gültigen Kennzeichenrechts und den Besitzrechten der jeweiligen eingetragenen Eigentümer. Allein aufgrund der bloßen Nennung ist nicht der Schluss zu ziehen, dass Markenzeichen nicht durch Rechte Dritter geschützt sind!

Copyright

Für in diesem Dokument veröffentlichte, von den Autoren selbst erstellte Objekte gilt hinsichtlich des Copyrights die folgende Regelung:

Dieses Werk ist unter einer Creative Commons-Lizenz (4.0 Deutschland Lizenzvertrag) lizenziert. D. h. Sie dürfen:

- Teilen: Das Material in jedwedem Format oder Medium vervielfältigen und weiterverbreiten
- Bearbeiten: Das Material remixen, verändern und darauf aufbauen

und zwar für beliebige Zwecke, sogar kommerziell. Der Lizenzgeber kann diese Freiheiten nicht widerrufen, solange Sie sich an die Lizenzbedingungen halten.

Die Nutzung ist unter den folgenden Bedingungen möglich:

- Namensnennung: Sie müssen angemessene Urheber- und Rechteangaben machen, einen Link zur Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Diese Angaben dürfen in jeder angemessenen Art und Weise gemacht werden, allerdings nicht so, dass der Eindruck entsteht, der Lizenzgeber unterstütze gerade Sie oder Ihre Nutzung besonders.
- Weitergabe unter gleichen Bedingungen: Wenn Sie das Material remixen, verändern oder anderweitig direkt darauf aufbauen, dürfen Sie Ihre Beiträge nur unter derselben Lizenz wie das Original verbreiten.
- Keine weiteren Einschränkungen: Sie dürfen keine zusätzlichen Klauseln oder technische Verfahren einsetzen, die anderen rechtlich irgendetwas untersagen, was die Lizenz erlaubt.

Im Weiteren gilt:

- Jede der vorgenannten Bedingungen kann aufgehoben werden, sofern Sie die Einwilligung des Rechteinhabers dazu erhalten.
- Diese Lizenz lässt die Urheberpersönlichkeitsrechte unberührt.

Um sich die Lizenz anzusehen, gehen Sie bitte ins Internet auf die Webseite:

https://creativecommons.org/licenses/by-sa/4.0/deed.de

bzw. für den vollständigen Lizenztext

https://creativecommons.org/licenses/by-sa/4.0/legalcode

Inhaltsverzeichnis

Tabellenverzeichnis

Zusammenfassung

Personenbezogene Daten sind „alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen“. Dieser Begriff ist somit sehr umfassend, denn er erfasst somit alle Daten, die einer individuellen Person direkt oder indirekt zurechenbar sind.

Wie die bisherigen nationalen Datenschutzregelungen enthält auch die EU-Datenschutz-Grundverordnung (DS-GVO) ein grundsätzliches Verbot der Verarbeitung personenbezogener Daten. Daher gilt nach wie vor, dass jede Verarbeitung personenbezogener Daten verboten ist. Somit benötigt jede Verarbeitung einen gesetzlich geregelten Erlaubnistatbestand. Insbesondere ist zu beachten, dass je sensibler die zu verarbeitenden Daten sind, desto notwendiger die Gewährleistung eines angemessen hohen Schutzniveaus für diese Daten ist. Gewährleisten muss den Schutz personenbezogener Daten für die gesamte Dauer der Verarbeitung, also über den gesamten Lebenszyklus der Daten hinweg, der „Verantwortliche“. Art. 4 Ziff. 7 DS-GVO definiert einen Verantwortlichen als „die natürliche oder juristische Person, Behörde, Einrichtung oder andere Stelle, die allein oder gemeinsam mit anderen über die Zwecke und Mittel der Verarbeitung von personenbezogenen Daten entscheidet“. Somit ist derjenige, welcher über die Mittel und Zwecke der Verarbeitung entscheidet, vollumfänglich für alles verantwortlich, was mit diesen Daten geschieht.

Die DS-GVO sieht die Pseudonymisierung als eine mögliche Maßnahme an, deren Einsatz die Gewährleistung eines angemessenen Schutzniveaus unterstützen kann. Die DS-GVO verweist an verschiedenen Stellen, wie z. B. bei den Anforderungen bzgl. Privacy by Design/Default (Art. 25) oder bei der Verarbeitung von Forschungsarbeiten, auf diese Maßnahme.

Die im BDSG noch erhaltene Begrifflichkeit der Anonymisierung wird von der DS-GVO nicht mehr explizit definiert. Jedoch bedeutet dieses nicht, dass eine Anonymisierung mit Geltung der DS-GVO per se nicht mehr möglich ist. Vielmehr adressiert sowohl die DS-GVO als auch etwaige nationale Regelungen an mehreren Stellen diese Möglichkeit. Bedingt durch die Regelungen der DS-GVO bzgl. den Begrifflichkeiten „personenbezogene Daten“ und „Pseudonymisierung“ änderten sich die Anforderungen an eine Anonymisierung. Eine „faktische“ Anonymisierung von Daten, wie sie bspw. bisher im bis zum 24. Mai 2018 geltenden BDSG vorgesehen war, wird mit Geltung der DS-GVO, d. h. ab dem 25. Mai 2018, so nicht mehr möglich sein.

Sowohl die Pseudonymisierung als auch die Anonymisierung können bzw. sollten daher verwendet werden. Einerseits um die Risiken der Verarbeitung personenbezogener Daten für von dieser Verarbeitung betroffene Personen zu verringern, andererseits um den aus der DS-GVO resultierenden rechtlichen Anforderungen hinsichtlich eines angemessen hohen Schutzniveaus zu genügen.

Dabei ist zu beachten, dass sowohl die Anonymisierung als auch die Pseudonymisierung eine Verarbeitung im Sinne der DS-GVO darstellen. D. h. es wird eine Rechtsgrundlage („Erlaubnistatbestand“) für die Durchführung einer Anonymisierung oder auch einer Pseudonymisierung benötigt. Dies ist nur eine der datenschutzrechtlichen Rahmenbedingungen, welche in der DS-GVO zu finden sind.

In dieser Ausarbeitung werden verschiedene Rahmenbedingungen besprochen, die im Hinblick auf eine Pseudonymisierung oder Anonymisierung zu beachten sind.

1 Einleitung

Sowohl die Pseudonymisierung als auch die Anonymisierung wird verwendet um Risiken der Verarbeitung für von der Verarbeitung betroffene Personen zu verringern. D. h. die Methoden stellen Maßnahmen dar, welche dem Schutz von personenbezogenen Daten dienen. In Bezug auf die Pseudonymisierung schrieb der europäische Gesetzgeber in ErwGr. 28 DS-GVO: „Die Anwendung der Pseudonymisierung auf personenbezogene Daten kann […] die Verantwortlichen und die Auftragsverarbeiter bei der Einhaltung ihrer Datenschutzpflichten unterstützen".

Wie aus dieser Formulierung ersichtlich wird, kann es einen oder auch mehrere Verantwortliche geben. Art. 4 Ziff. 7 DS-GVO definiert „Verantwortlicher" als „die natürliche oder juristische Person, Behörde, Einrichtung oder andere Stelle, die allein oder gemeinsam mit anderen über die Zwecke und Mittel der Verarbeitung von personenbezogenen Daten entscheidet". Entsprechend wird in Art. 26 DS-GVO erklärt, unter welchen Umständen eine Verarbeitung durch gemeinsam Verantwortliche erfolgen kann. Dies gilt selbstverständlich auch für die Pseudonymisierung oder auch die Anonymisierung, die jeweils eine Verarbeitung darstellen. Und auch diese Verarbeitungen können von einem oder auch von mehreren gemeinsam Verantwortlichen durchgeführt und genutzt werden.

Eine Pseudonymisierung ersetzt nicht zwangsläufig andere Datenschutzmaßnahmen, sondern ist eher als begleitende Maßnahme zu verstehen (ErwGr. 28 S. 2 DS-GVO). Die DS-GVO nennt die Pseudonymisierung als begleitende Maßnahme an verschiedenen Stellen wie z. B.:

- Art. 6 Abs. 4 DS-GVO, um bei Zweckänderung geeignete Garantien für die Sicherheit abzubilden
- Art. 25 Abs. 1 DS-GVO, als eines der Mittel, um „Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen" („Privacy by Design/Privacy by Default") umzusetzen
- Art. 32 Abs. 1 lit. a DS-GVO, als eine der zu berücksichtigenden Anforderungen bei der Gewährleistung eines angemessenen Schutzniveaus
- Art. 89 Abs. 1 DS-GVO, als eine mögliche Maßnahme, um Rechte und Freiheiten betroffener Person bei der Verarbeitung personenbezogener Daten von im öffentlichen Interesse liegenden Archivzwecken, zu wissenschaftlichen oder historischen Forschungszwecken oder zu statistischen Zwecken zu gewährleisten.

Anonyme oder pseudonyme Daten können z. B. genutzt werden für:

- Schulungszwecke
- Krankheitsregister
- Klinische Studien
- Statistische Auswertungen/Analysen.

Art. 40 Abs. 2 lit. d DS-GVO verweist darauf, dass Verhaltensregeln die Pseudonymisierung personenbezogener Daten konkreter ausgestalten können. Demnach enthalten die Vorgaben der DS-GVO die Rahmenbedingungen, unter welchen eine Pseudonymisierung anwendbar ist. Bzgl. der Ausgestaltung der Pseudonymisierung gibt es aber Bedarf an Auslegung. Dies will diese Ausarbeitung leisten: eine Klarstellung, wie mit Pseudonymisierung und Anonymisierung mit Geltung der DS-GVO umzugehen ist.

2 Abgrenzung / Klarstellung

Die vorliegende Ausarbeitung bezieht sich auf Daten in der Gesundheitsversorgung, d. h. besondere Kategorien von Daten im Sinne der DS-GVO. Grundsätzlich ist eine Pseudonymisierung oder Anonymisierung natürlich auch bei anderen Daten sinnvoll. Die vorliegend dargestellten Ausführungen / Methoden sind daher i. d. R. auch auf diese Daten übertragbar.

Diese Ausarbeitung stellt keine Verhaltensregel i.S.v. Art. 40 Abs. 2 lit. d DS-GVO dar, sondern beschreibt, welche Rahmenbedingungen beim Vorgehen bzgl. Anonymisierung oder Pseudonymisierung aus Sicht der DS-GVO mindestens beachtet werden sollten. Weiterhin werden einige Methoden zu dieser Thematik vorgestellt, ohne dass diesbezüglich von den Verfassern ein Anspruch auf Vollständigkeit bzgl. der Darstellung erhoben wird.

3 Allgemeines

Im Rahmen der folgenden Darstellung werden des Öfteren Beispiele zur Veranschaulichung genutzt. Alle Beispiele basieren auf dem folgenden onkologischen Beispieldatensatz:

Vorname	Nachname	Geschlecht	Geb.-Datum	PLZ	ICD	Diagnose
Heike	Richter	W	11.05.1983	10115	C43.9	Bösartiges Melanom der Haut, nicht näher bezeichnet
Jan	Schröder	M	03.12.1965	10115	D22.9	Melanozytennävus, nicht näher bezeichnet
Hugo-Egon	Meyer	M	27.08.1977	10178	C85.9	Non-Hodgkin-Lymphom, Typ nicht näher bezeichnet
Eckbert	Schneider	M	23.12.1981	10247	D44.8	Neubildung unsicheren oder unbekannten Verhaltens: Beteiligung mehrerer endokriner Drüsen
Jürgen	Stillstand	M	29.11.1985	10319	C18.4	Bösartige Neubildung: Colon transversum
Hiltrud	Niemand	W	15.07.1987	10407	D46.1	Refraktäre Anämie mit Ringsideroblasten
Uwe	Müller	M	31.03.1988	10435	C16.3	Bösartige Neubildung: Antrum pyloricum
Michael	Matuschek	M	13.04.1968	10439	D46.2	Refraktäre Anämie mit Blastenüberschuss
Anke	Schmidt	W	01.04.1978	10585	C50.3	Bösartige Neubildung unterer innerer Quadrant Brustdrüse
Kunigunde	Gewaltig	W	21.01.1969	10707	C91.10	Chronische lymphatische Leukämie: Ohne Angabe einer kompletten Remission
Franz	Herrlich	M	17.11.1967	10717	D12.8	Gutartige Neubildung: Rektum
Berthold	Koch	M	28.08.1991	10717	D12.6	Gutartige Neubildung: Kolon, nicht näher bezeichnet
Fieda	Fischer	W	15.11.1987	10787	C50.8	Bösartige Neubildung: Brustdrüse, mehrere Teilbereiche überlappend
Gerfriede	Jensen	W	23.07.1983	10827	C50.1	Bösartige Neubildung: Zentraler Drüsenkörper der Brustdrüse
Käthe	Albers	W	27.05.1975	10963	C83.0	Non-Hodgkin-Lymphom: Kleinzellig (diffus)

Tabelle 1: Beispieldaten mit onkologischen Erkrankungen

4 Begriffsbestimmungen

4.1 Personenbezogene Daten

Entsprechend Art. 4 Ziff. 1 DS-GVO sind personenbezogene Daten „alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen“. Damit fallen nicht nur Informationen darunter, welche direkt eine Person identifizieren, sondern auch alle Informationen, welche über Zwischenschritte eine Person identifizieren. D. h. soweit und solange die Informationen aus sich heraus Rückschluss auf eine einzelne Person zulassen, handelt es sich um Daten einer bestimmten Person[1].

Der Begriff „identifizierbar“ muss daher im Sinne von „als Einzelperson wahrnehmbar“ bzw. einer „Einzelperson zuordenbar“ verstanden werden.

Die Identifizierbarkeit ist damit Dreh- und Angelpunkt hinsichtlich der Beurteilung, ob Daten als anonym oder pseudonym angesehen werden können.

4.2 Pseudonymisierung

Der Begriff der Pseudonymisierung wird in Art. 4 Ziff. 5 DS-GVO definiert. Dort heißt es:

> „"Pseudonymisierung" die Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden“.

Dieser Definition folgend charakterisiert eine Pseudonymisierung daher Nachfolgendes:

- Die Pseudonymisierung ist eine Verarbeitung personenbezogener Daten.
- Pseudonyme Daten sind Daten, die ohne weitere Informationen einer spezifischen Person nicht zuordenbar sind.
- Die zur Zuordenbarkeit benötigten Informationen stehen dem Verantwortlichen nicht zur Verfügung, sondern
  - werden gesondert aufbewahrt und
  - sind durch technische und organisatorische Maßnahmen vor dem Zugriff durch den Verantwortlichen geschützt.

→ Für den Verantwortlichen besteht bei der Verarbeitung pseudonymisierter Daten keine Möglichkeit der Identifizierung der betroffenen Person.

Hinweis: ErwGr. 26 führt aus, dass zur Feststellung, ob eine natürliche Person identifizierbar ist, alle Mittel berücksichtigt werden sollten, die von dem Verantwortlichen oder einer anderen Person nach allgemeinem Ermessen wahrscheinlich genutzt werden, um die natürliche Person direkt oder indirekt zu identifizieren.
Bei der Feststellung, ob Mittel nach allgemeinem Ermessen wahrscheinlich zur Identifizierung der natürlichen Person genutzt werden, sollten alle objektiven Faktoren, wie die Kosten der Identifizierung und der dafür erforderliche Zeitaufwand, herangezogen werden, wobei die zum Zeitpunkt der Verarbeitung verfügbare Technologie und technologische Entwicklungen zu berücksichtigen sind.

[1] Karg M. (2015) Anonymität, Pseudonyme und Personenbezug revisited. DuD: 520-526

4.3 Pseudonyme Daten

Entsprechend der in der DS-GVO enthaltenen Definition von Pseudonymisierung sind demnach pseudonyme Daten solche Daten, welche der oder die Verantwortlichen keiner spezifischen Person zuordnen können, jedoch für andere durch die Einbeziehung weitergehender Informationen („Zuordnungsregeln") die grundsätzliche Möglichkeit der Zuordnung besteht. Dafür ist es nicht erforderlich, dass die betroffene Person durch die „Re-Identifizierung" mit bürgerlichem Namen zu identifizieren ist[1]. Ausreichend ist vielmehr, wenn durch das Datum bzw. die Daten die betroffene Person individualisiert wird und Aussagen über deren sachliche und persönliche Verhältnisse möglich sind; ein Name muss nicht vorhanden sein[2].

4.4 Anonyme Daten

Entsprechend ErwGr. 26 DS-GVO sollten die Vorgaben der DS-GVO nicht für anonyme Daten gelten. D. h. für anonyme Daten gelten die Anforderungen der DS-GVO nicht. Jedoch muss der Verantwortliche, u. a. der Rechenschaftspflicht aus Art. 5 Abs. 2 DS-GVO folgend, zu jedem Zeitpunkt der Verarbeitung (und damit insbesondere auch während der gesamten Speicherdauer) nachweisen können, dass es sich um anonyme Daten handelt.

Daraus ergibt sich im Umkehrschluss, dass anonyme Daten weder direkt personenbezogene Daten noch pseudonymisierte Daten sein können. D. h., anonyme Daten sind Daten, bei denen keine Zuordnungsmöglichkeit zu einer spezifischen betroffenen Person existiert[3].

Cave: Anders als im BDSG a.F. bedeutet „anonym" unter der DS-GVO, dass keine Möglichkeit zur Re-Identifikation besteht; eine Abwägung wie in § 3 Ziff. 6 BDSG a.F. „[…] oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft […]" ist unter der DS-GVO nicht vorgesehen. Daher gilt, dass wenn eine Möglichkeit der Zuordnung der Daten zu einer spezifischen betroffenen Person existiert, die Daten keine anonymen Daten sind.

4.5 Anonymisierung

Anonymisierung ist die Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten nicht mehr einer spezifischen betroffenen Person zugeordnet werden können.

Zur Klarstellung: Sowohl pseudonyme als auch anonyme Daten sind daher für den Verantwortlichen keiner spezifischen betroffenen Person zuordenbar. Der Unterschied zwischen anonymen und pseudonymen Daten liegt darin, dass bei pseudonymen Daten außerhalb der Zugriffsmöglichkeiten des Verantwortlichen grundsätzlich eine Zuordnungsmöglichkeit besteht oder bestehen könnte, bei anonymen Daten hingegen für niemanden eine Zuordnungsmöglichkeit vorhanden ist.

[2] Artikel-29-Datenschutzgruppe. WP 136 „Stellungnahme 4/2007 zum Begriff 'personenbezogene Daten'", S. 16: […] ein Name zur Identifizierung einer Person jedoch keineswegs immer notwendig ist". [Online, zitiert am 2018-04-24]; Verfügbar unter http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf

[3] Voigt P, von dem Bussche A. The EU General Data Protection Regulation (GDPR) - A Practical Guide. Springer Verlag, 2017. ISBN 978-3-319-57958-0. PP 13-16, chapter „2.1.2.2 Anonymisation and Pseudonymisation": „Anonymised data is either information that does not relate to an identified or identifiable individual or personal data that was rendered anonymous in such a manner that the person is not or no longer identifiable."

Da es sich sowohl bei der Pseudonymisierung als auch bei der Anonymisierung um eine Verarbeitung gemäß Art. 4 Ziff. 2 DS-GVO handelt, ist daher auch für eine Anonymisierung bzw. Pseudonymisierung von Gesundheitsdaten ein Erlaubnistatbestand gem. Art. 9 Abs. 2,4 DS-GVO bzw. Art. 6 Abs. 1, 2 DS-GVO für Daten, die nicht zu den besonderen Kategorien zählen, erforderlich.

5 Rechtliche Rahmenbedingungen

5.1 Erlaubnistatbestand Pseudonymisierung/Anonymisierung

Wie bereits dargestellt stellen sowohl Anonymisierung als auch Pseudonymisierung Verarbeitungen im Sinne von Art. 4 Ziff. 2 DS-GVO dar. Nach Art. 6 sowie Art. 9 DS-GVO ist die Verarbeitung personenbezogener Daten verboten, außer es existiert ein Erlaubnistatbestand.

5.1.1 Einwilligung

Existiert kein gesetzlicher Erlaubnistatbestand, so ist zur Pseudonymisierung oder Anonymisierung die Einholung einer rechtsgültigen Einwilligung der betroffenen Person bzw. Personen erforderlich. Hierbei sind alle die von der DS-GVO genannten Vorgaben / Rahmenbedingungen einzuhalten.[4]

5.1.2 Sonderfall Forschung

Die DS-GVO privilegiert Verarbeitungen personenbezogener Daten zum Zwecke der Forschung an unterschiedlichen Stellen. Insbesondere enthält die DS-GVO privilegierende Bestimmungen für wissenschaftliche und historische Forschungszwecke[5]. Hinsichtlich der Nutzung besonderer Kategorien personenbezogener Daten findet sich in Art. 9 Abs. 2 lit. j DS-GVO ein datenschutzrechtlicher Erlaubnistatbestand zur Nutzung von Daten zu Zwecken der wissenschaftlichen Forschung. Hiernach ist eine Verarbeitung gestattet, wenn die Verarbeitung gemäß Art. 89 Abs. 1 DS-GVO erforderlich ist und ein nationales oder europäisches Recht für die Nutzung existiert, welches den besonderen Anforderungen von Art. 89 Abs. 1 DS-GVO genügt.

5.1.2.1 Forschung ohne Einwilligung nach § 27 BDSG n.F.

In Deutschland erfolgte eine nationale Konkretisierung dieser Öffnungsklausel in § 27 BDSG n.F. Die Regelung in § 27 BDSG n.F. gestattet die Verarbeitung besonderer Kategorien von Daten im Sinne des Art. 9 Abs. 1 DS-GVO zu Forschungszwecken auch ohne Einwilligung, wenn

a) die Verarbeitung zu diesen Zwecken erforderlich ist und
b) die Interessen der Verantwortlichen an der Verarbeitung die Interessen der betroffenen Person an der Nicht-Verarbeitung ihrer Daten erheblich überwiegen.

§ 27 Abs. 1 BDSG n.F. stellt somit einen Erlaubnistatbestand zur Verarbeitung von Gesundheitsdaten zu Forschungszwecken dar, wobei allerdings zu beachten ist, dass spezifischeres Bundes- oder Landesrecht vorrangig gelten kann (§ 1 Abs. 1 Nr. 2 BDSG n.F.)[6]. Weiterhin gilt § 27 Abs. 1 BDSG n.F. nur für die Verarbeitung von Daten i.S.v. Art. 9 Abs. 1 DS-GVO; für Daten, die nicht unter Art. 9 Abs. 1 DS-GVO fallen, müssen andere Erlaubnistatbestände gefunden werden wie z. B. die in Art. 6 DS-GVO zu Findenden.

---

[4] Näheres siehe: Ausarbeitung der GMDS AG DIG „Anforderungen der DS-GVO an die Einwilligung“. [Online, zitiert am 2018-04-24]; Verfügbar unter http://ds-gvo.gesundheitsdatenschutz.org/html/einwilligung.php

[5] Zu den Begriffsbestimmungen bzgl. „Forschung“, „wissenschaftliche Forschung“ und „historische Forschung“ siehe Ausarbeitung „Datenschutzrechtliche Anforderungen an die medizinische Forschung unter Berücksichtigung der EU Datenschutz-Grundverordnung (DS-GVO)“, herausgegeben von der Arbeitsgruppe „Datenschutz und IT-Sicherheit im Gesundheitswesen“ der GMDS und der Arbeitsgruppe „Datenschutz und Datensicherheit im Gesundheits- und Sozialwesen“ der GDD. [Online, zitiert am 2018-04-25]; Verfügbar unter http://ds-gvo.gesundheitsdatenschutz.org/html/forschung.php

[6] Beispiele für vorrangige bereichsspezifische Regelungen sind, soweit sie den Vorgaben der DS-GVO entsprechen:

- Bundesrecht: Sozialgesetzbücher, Arzneimittelgesetz, Gendiagnostikgesetz usw.
- Landesrecht: Krankenhausgesetze, Krebsregistergesetze

Die von § 27 Abs. 1 DS-GVO geforderte Interessenabwägung stellt hohe Anforderungen: die Interessen des oder der Verantwortlichen müssen nicht nur überwiegen, sondern sie müssen *erheblich* überwiegen. Dies entspricht damit den Anforderungen aus § 28 Abs. 6 Nr. 4 BDSG a.F., weshalb diese Anforderungen Forschenden in Deutschland wohl bekannt sein dürften. Dementsprechend kann ein erhebliches Überwiegen des Interesses an der Forschung angenommen werden, wenn ein Forschungsvorhaben „erhebliche Verbesserungen für die Gesundheit oder soziale Sicherheit der Bevölkerung mit sich bringt“[7].

Gemäß § 27 Abs. 3 BDSG n.F. sind besondere Kategorien personenbezogener Daten im Sinne des Art. 9 Abs. 1 DS-GVO zu anonymisieren, sobald dies nach dem Forschungs- oder Statistikzweck möglich ist, es sei denn, berechtigte Interessen der betroffenen Person stehen einer Anonymisierung entgegen. Ist dies der Fall, sind die Merkmale, mit denen Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren Person zugeordnet werden können, gesondert zu speichern (§ 27 Abs. 3 S. 2 BDSG n.F.). Die Neuregelung des § 27 Abs. 3 BDSG n.F. verlangt somit, Forschung grundsätzlich mit anonymen Daten durchzuführen und in Ausnahmefällen mit pseudonymen Daten.

Von einer Unvereinbarkeit einer Anonymisierung mit der Zweckerreichung eines Forschungsvorhabens kann zumindest immer dann ausgegangen werden, wenn eine fortlaufende Zuordnung von neuen Daten zu bereits vorhandenen Daten erforderlich ist[8]. Gleiches gilt, wenn im Rahmen des Forschungsvorhabens die betroffene Person ggf. kontaktiert werden muss, z. B. weil Forschungsergebnisse ihre Behandlung beeinflussen könnten. Weiterhin schließt die Nutzung von Biomaterialien, deren enthaltene genetische Informationen prinzipiell einen Personenbezug erlauben, eine Anonymisierung aus[8].

Eine Re-Identifizierung, die bei einer Verarbeitung pseudonymer Daten grundsätzlich ja möglich ist, darf entsprechend § 27 Abs. 3 S. 3. BDSG n.F. nur durchgeführt werden, wenn dies der Forschungs- oder Statistikzweck erfordert[9].

5.2 Betroffenenrechte

5.2.1 Anonyme Daten und Betroffenenrechte

Entsprechend ErwGr. 26 gelten für anonyme Daten die Anforderungen der DS-GVO nicht. Damit gelten auch nicht die aus den Artt. 12 bis 22 DS-GVO resultierenden Anforderungen hinsichtlich der Erfüllung der Betroffenenrechte. Dies bezieht sich jedoch nicht auf die Zeit vor der Anonymisierung. Denn dann sind die jeweiligen Betroffenenrechte sehr wohl vollumfänglich zu beachten. Somit ist bei einer Anonymisierung die betroffene Person ggf. u. a. hinsichtlich des Vorgehens einer Anonymisierung und – sofern vorhanden – der Zweckänderung bei der Verarbeitung der Daten zu informieren.

[7] Buchner B, Tinnefeld M-T. § 27 BDSG RN. 12 in Kühling/Buchner, Kommentar zur Datenschutz-Grundverordnung/BDSG. C.H.Beck Verlag 2. Auflage 2018. ISBN 978-3-406-71932-5

[8] Buchner B, Tinnefeld M-T. § 27 BDSG RN. 24 in Kühling/Buchner, Kommentar zur Datenschutz-Grundverordnung/BDSG. C.H.Beck Verlag 2. Auflage 2018. ISBN 978-3-406-71932-5

[9] Bzgl. „erforderlich“ siehe Kap. 4.7 „Erforderlichkeit, Notwendigkeit“ in der Ausarbeitung „Datenschutzrechtliche Anforderungen an die medizinische Forschung unter Berücksichtigung der EU Datenschutz-Grundverordnung (DS-GVO)“, herausgegeben von der Arbeitsgruppe „Datenschutz und IT-Sicherheit im Gesundheitswesen“ der GMDS und der Arbeitsgruppe „Datenschutz und Datensicherheit im Gesundheits- und Sozialwesen“ der GDD. [Online, zitiert am 2018-04-25]; Verfügbar unter http://ds-gvo.gesundheitsdatenschutz.org/html/forschung.php

5.2.2 Pseudonyme Daten und Betroffenenrechte

Pseudonyme Daten gelten als personenbezogene Daten, daher gelten auch die Betroffenenrechte voll umfänglich. Pseudonyme Daten weisen die Besonderheit auf, dass der Verantwortliche die betroffene Person nicht identifizieren kann. Das wiederum macht es notwendig, dass jegliche Kommunikation mit der betroffenen Person nur über die Personen laufen kann, welche die Daten (ursprünglich) erhoben haben, d. h. welche die betroffene Person kennen und die ggf. mittels einer Zuordnungstabelle eine Re-Identifizierung vornehmen können.

5.2.3 Information bei Zweckänderung

Häufig erfolgt eine Anonymisierung/Pseudonymisierung von Daten, um die Daten zu einem anderen Zweck als den ursprünglichen zu verwenden. Gemäß Artt. 13 Abs. 4, 14 Abs. 4 DS-GVO muss der Verantwortliche vorher (also insbesondere noch vor der Pseudonymisierung/Anonymisierung) der bzw. den betroffenen Personen Informationen über diesen anderen Zweck und alle anderen maßgeblichen Informationen gemäß Art. 14 Abs. 2 DS-GVO zur Verfügung stellen.

5.3 Privacy by Design/Default

Entsprechend Art. 25 DS-GVO muss die Pseudonymisierung bzw. Anonymisierung über den vollständigen Lebenszyklus der Daten, d. h. von der Erhebung bis zur endgültigen Vernichtung, aufrechterhalten werden. Änderungen in der technischen Entwicklung müssen während dieser Zeit betrachtet, bzgl. der Auswirkungen auf die pseudonymisierten/anonymisierten Daten bewertet und ggf. erforderliche Maßnahmen abgeleitet und umgesetzt werden[10].

[10] Hinweise bzgl. des Vorgehens hierzu finden sich in der Praxishilfe „Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen (Art. 25 DS-GVO)“. [Online, zitiert am 2018-06-20]; Verfügbar unter http://ds-gvo.gesundheitsdatenschutz.org/html/privacy_design_default.php

5.4 Datenschutz-Folgenabschätzung

Die Aufsichtsbehörden der jeweiligen Bundesländer veröffentlichten Listen mit Kriterien, wann eine Datenschutzfolgenabschätzung erforderlich ist[11]. Demnach erfordert in einigen Bundesländern die Anonymisierung von Daten nach Art. 9 DS-GVO eine Datenschutz-Folgenabschätzung entsprechend Art. 35 DS-GVO[12]:

Kriterium	Bundesland
Anonymisierung von besonderen personenbezogenen Daten nach Artikel 9 DS-GVO, falls diese (ggf. vermeintlich) anonymen Daten an Dritte weitergegeben oder zu nicht nur internen statistischen Zwecken verarbeitet werden sollen	– Brandenburg – Bremen – Mecklenburg-Vorpommern – Saarland – Sachsen – Sachsen-Anhalt – Schleswig-Holstein – Thüringen
Umfangreiche Anonymisierung von besonderen Kategorien personenbezogener Daten nach Artikel 9 DS-GVO, falls diese (ggf. vermeintlich) anonymen Daten an Dritte weitergegeben oder zu nicht nur internen statistischen Zwecken verarbeitet werden sollen	– Niedersachsen
Umfangreiche Verarbeitung personenbezogener Daten im Rahmen der amtlichen Statistik, deren Erhebung, Speicherung und Verarbeitung, insbesondere der Anonymisierungsprozesse sowie deren Anonymisierung und statistische Aufbereitung vor/für die Übermittlung der Informationen an Dritte.	– Brandenburg – Niedersachsen – Nordrhein-Westfalen – Thüringen

[11] GMDS AG „Datenschutz und IT-Sicherheit im Gesundheitswesen“: Listen gemäß Art. 35 Abs. 4 DS-GVO der deutschen Datenschutz-Aufsichtsbehörden. [Online, zitiert am 2018-06-20]; Verfügbar unter http://ds-gvo.gesundheitsdatenschutz.org/html/dsfa_liste_aufsichtsbehoerden.php

[12] Hinweise bzgl. des Vorgehens hierzu finden sich in der Praxishilfe „Datenschutz-Folgenabschätzung gemäß Art. 35 DS-GVO“. [Online, zitiert am 2018-06-20]; Verfügbar unter http://ds-gvo.gesundheitsdatenschutz.org/html/dsfa.php

6 Sonderfall: Genetische Daten / Biomaterial

Schon 2005 wurde auf der 27. Internationalen Konferenz der Beauftragten für Datenschutz und den Schutz der Privatsphäre in Montreux in Bezug auf genetische Daten Folgendes festgehalten[13]:

„Die Datenschutzbeauftragten

1. […]
7. Sind sich bewusst, dass aufgrund des rasch wachsenden Kenntnisstandes im Bereich der Genetik Daten über die menschliche DNA zu den sensibelsten überhaupt werden können, und dass die Gewährleistung eines angemessenen rechtlichen Schutzes dieser Daten angesichts der beschleunigten Wissensentwicklung wachsende Bedeutung erlangt,
8. […]“

Aktuellere Untersuchungen haben ergeben, dass genetische Daten nicht als anonym angesehen werden können[14]. Genetische Daten können daher maximal pseudonymisiert werden bzw. als pseudonym angesehen werden.

6.1 Biomaterial und Einwilligung

Biomaterial enthält durch seinen genetischen Inhalt prinzipiell auch Daten, durch die Rückschlüsse auf etwaige dem Betroffenen verwandte Personen wie beispielsweise Kinder, Eltern und/oder Enkel möglich sind. Unter den Vorgaben der DS-GVO können diese Daten mittels Einwilligung kaum verarbeitet werden, da eine Einwilligung immer nur für die eigenen Daten, nicht aber für die Daten Dritter gegeben werden kann.

Daraus folgt, dass wenn eine Person heute beispielsweise Biomaterial abgibt, so kann ein heute noch nicht geborenes Enkelkind dieser Person aufgrund der Kenntnis dieses gespendeten Biomaterials in der Zukunft durch eine Erkrankung diskriminiert werden, die heute noch gar nicht bekannt ist, die aber dem Erbgut innewohnt. Auch diesen Personen muss ein „Recht auf Nichtwissen“ zugestanden werden, eine Einwilligung der das Biomaterial abgebenden Person kann dies nicht legitimieren.

[13] 27. Konferenz vom 14. - 16. September 2005 in Montreux. [Online, zitiert am 2018-06-20]; Verfügbar unter https://www.lda.brandenburg.de/sixcms/detail.php/bb1.c.272605.de?_aria=ds bzw. pdf-Datei beim BfDI unter http://www.bfdi.bund.de/SharedDocs/Publikationen/Entschliessungssammlung/IntDSK/IntDSK2005-MontreuxErklaerungDSInEinerGlobalisiertenWelt.pdf?__blob=publicationFile

[14] Siehe z. B.

- McGuire A, Gibbs RA (2006) No longer de-identified. Science 312: 370–371
- El Emam K. (2011) Methods for the de-identication of electronic health records for genomic research. Genome Medicine 3:25. [Online, zitiert am 2018-06-20]; Verfügbar unter http://genomemedicine.com/ content/3/4/25
- El Emam K, Jonker E, Arbuckle L, Malin B (2011) A systematic review of re-identification attacks on health data. PLoS One 2011; 6: e28071. [Online, zitiert am 2018-06-20]; Verfügbar unter https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3229505/
- Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y (2013) Identifying personal genomes by surname inference. Science 339: 321–324

Diskussion des Themas z. B. Hansson et al. (2016) The risk of re-identification versus the need to identify individuals in rare disease research. Eur J Hum Genet 24(11): 1553–1558. [Online, zitiert am 2018-06-20]; Verfügbar unter https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5110051/

Daher wird für die Verarbeitung derartiger personenbezogener bzw. personenbeziehbarer Daten ein gesetzlicher Erlaubnistatbestand benötigt. Wenn seitens der Gesellschaft ein Konsens besteht, dass die Vorratsdatenspeicherung von Biomaterial mit recht allgemeiner Zweckbindung („medizinische Forschung“) gewünscht ist, muss somit ein Gesetz diese Forschung explizit erlauben, so dass eine Einwilligung nicht benötigt wird. Denn eine Einwilligung alleine des Spenders wird kaum die damit verbundene Verarbeitung der über Dritte (= Verwandte) enthaltenen Informationen legitimieren können.

7 Exkurs: HIPAA und De-Identification – der amerikanische Weg

Der amerikanische „Health Insurance Portability and Accountability Act[15]“ (HIPAA) kennt den Begriff der De-identification, also den Prozess, um Personen vor der Identifikation zu schützen. Das Vorgehen hierzu wird in Abschnitt 164.514 von HIPAA beschrieben[16]. Das U. S. Department of Health & Human Services veröffentlichte eine diesbezügliche Leitlinie wie De-Identifikation in Bezug auf medizinische Daten umgesetzt werden kann[17].

§ 16.4514(b) HIPAA enthält eine Spezifizierung bzgl. der Umsetzung der De-Identifikation und kennt zwei Methoden:

1) § 16.4514(b)(1)
   Eine Person mit angemessenem Wissen und Erfahrung bzgl. allgemein anerkannten statistischen und wissenschaftlichen Prinzipien und Methoden hinsichtlich einer De-Identifikation
   i. entscheidet, dass das angewandte Verfahren zur De-Identifikation nur sehr geringe Risiken auch in Kombination mit Informationen, die anderen Stellen (Dritten i. S. d. DS-GVO) zur Verfügung stehen, hinsichtlich der Re-Identifizierung der betroffenen Person(en) beinhaltet und
   ii. dokumentiert die angewandten Methoden und die Analyse bzgl. der Risiken.
2) § 16.4514(b)(2) (nach Angaben des U.S. Department of Health & Human Services[17])
   Es müssen nachfolgende Informationen von betroffenen Personen oder von Verwandten, Arbeitgebern, oder Haushaltsmitgliedern entfernt werden:
   a) Namen
   b) Alle ortsbezogenen Angaben kleiner als ein Staat, einschließlich Straßenadresse, Stadt, Bezirks, Bezirks, Postleitzahl, außer den ersten 3 Ziffern der Postleitzahl
   c) Alle Datumsangaben, die direkt auf eine Person schließen lassen, außer dem Jahr
   d) Telefonnummern
   e) Fahrzeugnummern/Seriennummer, KFZ-Kennzeichen
   f) Faxnummern
   g) Geräte-Identifizierer sowie Seriennummern
   h) E-Mailadressen
   i) Web Universal Resource Locators (URLs)
   j) Sozialversicherungsnummern
   k) Internet Protocol (IP) Adressen
   l) Patienten-IDs aus elektronischen Patientenakten
   m) Biometrische Identifikatoren, Fingerprint und Stimme eingeschlossen
   n) Identifikatoren eines Gesundheitsplans

[15] U.S. Government Publishing Service: H. Rept. 104-736 - HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT OF 1996. [Online, zitiert am 2018-06-12]; Verfügbar unter https://www.gpo.gov/fdsys/search/pagedetails.action?granuleId=CRPT-104hrpt736&packageId=CRPT-104hrpt736 b zw. Zusammenfassung unter https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html

[16] §164.514 Other requirements relating to uses and disclosures of protected health information. [Online, zitiert am 2018-06-12]; Verfügbar unter http://www.hipaasurvivalguide.com/hipaa-regulations/164-514.php

[17] U.S. Department of Health & Human Services: Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. [Online, zitiert am 2018-06-12]; Verfügbar unter https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html

o) Fotographien, die das gesamte Gesicht zeigen, und alle vergleichbaren Bildnisse
p) Kontonummern, Abrechnungsnummern
q) Jede andere ein Individuum identifizierende Nummer oder Kennzeichen, ausgenommen Angaben bzgl. § 16.4514(c) HIPAA (Angaben zur Re-Identifikation)
r) Nummern von Zeugnissen, Zertifikaten, Attesten, Führerscheinnummer
s) Den oder dem Verantwortlichen ist nach bestem Wissen und Gewissen keine Möglichkeit bekannt, wie die zu verarbeitenden Daten alleine oder in Kombination mit anderen Informationen zur Re-Identifikation genutzt werden können (ausgenommen § 16.4514(c) HIPAA)

§ 16.4514(c) HIPAA sieht die Möglichkeit einer Re-Identifikation vor. Dazu kann einer betroffenen Person im Rahmen der De-Identifikation ein Code oder ein anderes Identifizierungskennzeichen zugewiesen werden, um dadurch eine Möglichkeit zur Re-Identifikation zu erhalten.

Dieses wiederum hat zur Konsequenz, dass eine De-Identifikation nach HIPAA somit keine Anonymisierung im Sinne des europäischen Rechts darstellt. Eine De-Identifikation nach HIPAA kann jedoch den aus der DS-GVO resultierenden Anforderungen bezüglich einer Pseudonymisierung genügen, wenn der Verantwortliche keinen Zugriff auf die ggf. vorhandene Re-Identifikationsmöglichkeit hat.

8 Hands on: Wie geht man vor?

8.1 Identifizierung der direkten und indirekten identifizierenden Daten

Will man eine Pseudo- bzw. Anonymisierung durchführen, müssen in einem ersten Schritt sowohl direkte als auch indirekte Identifikationsmerkmale im Datensatz identifiziert und bzgl. der Notwendigkeit der Änderung/Löschung hinsichtlich des Prozesses einer Pseudonymisierung oder Anonymisierung bewertet werden.

Die zur Verarbeitung vorgesehenen Daten sind also in drei Kategorien einzuteilen:

a) Direkte Identifikationsmerkmale: Alle Daten, welche eine direkte Identifizierung zulassen. Beispiele für direkte Identifikationsmerkmale sind insbesondere Namen (der bürgerlicher Name sowie alle sonstige Namen wie beispielsweise der Rufname oder der Chat Name), unter denen die Person bekannt ist.

b) Indirekte Identifikationsmerkmale: Alle Daten, welche in Verbindung mit anderem indirektem oder externem Wissen eine Identifikation potentiell ermöglichen. Beispiele für indirekte Identifikationsmerkmale sind:
   - Personenbezeichner (z. B. Patienten-ID, Sozialversicherungsnummer, Steuernummer, Autokennzeichen, Kontonummer, Versicherungsnummer, Geburtsdatum)
   - Erscheinungsmerkmale (z. B. Körpergröße, Haarfarbe, Kleidung, Tätowierungen)
   - Biometrische Kennzeichen (z. B. Gesicht, Stimmprofile, Fingerabdrücke)
   - Genetische Daten
   - Digitale Zertifikate, welche eine Identifikationsmöglichkeit beinhalten (z. B. Zertifikate zur elektronischen Unterschrift)
   - Identifikationsmerkmale basierend auf elektronischer Kommunikation (z B. Telefonnummer, Faxnummer, E-Mailadresse, IP-Adresse)
   - Demographische Daten (z. B. Religion, Geburtsland, Muttersprache, Vorstrafen)
   - Zuordnungsmerkale (z. B. Beruf, Funktion, Anschriften, Vorstrafen, Name der Mutter/des Vaters)
   - Ausreißervariablen (z. B. seltene Diagnosen, Behandlungsbesonderheiten, körperliche Fehlbildungen, für die untersuchte Population untypische Merkmale).

   Je nachdem, welche weiteren Informationen dem oder den Verantwortlichen zur Verfügung stehen, sind indirekte Identifikationsmerkmale ggf. als direkte Identifikationsmerkmale anzusehen.

c) Nicht identifizierende Daten: alle anderen Daten, die weder direkte oder indirekte Identifikationsmerkmale darstellen.

Gerade bei der Analyse der sogenannten indirekten Identifikationsmerkmale ist häufig eine individuelle, kontextbezogene Betrachtung erforderlich. So kann bspw. in einem Fall die Haarfarbe ein indirektes Identifikationsmerkmal darstellen, durch die eine Person eindeutig identifizierbar wird. In anderen Fällen beinhaltet das Datum „Haarfarbe“ vielleicht keine Re-Identifikationsmöglichkeit. Desgleichen können medizinische Bilddaten neben den zur Identifizierung geeigneten Metadaten (z. B. DICOM StudyUID) auch in sich selbst eine Identifikationsmöglichkeit enthalten, z. B. wenn eine 3D-Rekonstruktion des Kopfes eine Gesichtserkennung ermöglichen würde.

8.2 Arten von Pseudonymen und ihre Unterscheidungsmöglichkeiten

Pseudonyme können einerseits durch den Inhaber der Zuordnungsregel unterschieden werden:

a) Pseudonyme werden ausschließlich vom Betroffenen selbst vergeben. Ein Beispiel hierfür ist der „Nickname“ des Nutzers im Chat.
b) Pseudonyme werden natürlich auch vom ursprünglichen Verarbeiter vergeben, wie dies beispielsweise bei der IP-Adress-Vergabe durch einen Internet-Provider erfolgt.
c) Pseudonyme können von einem vertrauenswürdigen Dritten vergeben werden, wie dies beispielsweise häufig bei der Einschaltung einer sog. Trusted-Third-Party in medizinischen Forschungsnetzen geschieht.

Eine andere Unterscheidungsmöglichkeit bzgl. Pseudonyme besteht in der Art ihrer Erzeugung:

a) Das Pseudonym wird durch eine schlüsselabhängige Einweg- oder Hashfunktion aus invarianten Daten (Bsp. Identitätsdaten) erzeugt.
b) Das Pseudonym wird (willkürlich) nach einem festen Einweg-Algorithmus vom Benutzer aus einem Geheimnis (z. B. Passphrase) erzeugt.
c) Das Pseudonym wird (zufällig) frei gewählt oder nach einem Zufallsverfahren erzeugt.

Eine dritte Unterscheidungsmöglichkeit bei Pseudonymen besteht in ihrer gesellschaftlichen Verwendung. Pseudonyme können als

- Personenpseudonyme, z. B.
    - Öffentliches Personenpseudonym (z. B. Telefonnummer)
    - Nichtöffentliches Personenpseudonym (z. B. Kontonummer)
    - Anonymes Personenpseudonym (z. B. Genom)

oder auch als

- Rollenpseudonyme, wie beispielsweise
    - Geschäftsbeziehungspseudonym (z. B. Chat- Name)
    - Transaktionspseudonym (z. B. PIN, TAN)

genutzt werden.

8.3 Methoden zur Pseudonymisierung/Anonymisierung

Sowohl bei der Pseudonymisierung als auch bei der Anonymisierung gibt es unterschiedliche Methoden, die im Einzelfall danach evaluiert werden sollten, wie mit ihnen am besten der individuell verfolgte Zweck erreicht werden kann.

8.3.1 Nichtangabe

Das zu schützende Datum wird bei dieser Methode nicht verwendet, sondern weggelassen, z. B. durch Löschung oder Nicht-Exportieren von Spalten einer Tabelle einer Datenbank oder auch von Teilbereichen der Werte.

Beispiel: Die Daten aus Tabelle 1 wurden durch Weglassen von Vorname und Nachname sowie Tag und Datum beim Geburtsdatum wie auch die letzten Ziffern bei ICD entpersonalisiert:

Geschlecht	Geb.-Datum	PLZ	ICD
w	1983	10115	C43
m	1965	10115	D22
m	1977	10178	C85
m	1981	10247	D44

Geschlecht	Geb.-Datum	PLZ	ICD
m	1985	10319	C18
w	1987	10407	D46
m	1988	10435	C16
m	1968	10439	D46
w	1978	10585	C50
w	1969	10707	C91
m	1967	10717	D12
m	1991	10717	D12
w	1987	10787	C50
w	1983	10827	C50
w	1975	10963	C83

Tabelle 2: Entpersonalisierung von Daten durch Nutzung der Methode der Nichtangabe

Allerdings muss man dabei darauf achten, dass sich durch Nichtangabe von Teilbereichen auch die Information selbst ändert. So kann bspw. die Verringerung des Wertes des ICD im obigen Beispiel zu einer ggf. nicht unerheblichen Änderung der Diagnosen führen:

Original		Nichtangabe	
ICD	**Diagnose**	**Diagnose**	**ICD**
C16.3	Bösartige Neubildung: Antrum pyloricum	Bösartige Neubildung des Magens	C16
C18.4	Bösartige Neubildung: Colon transversum	Bösartige Neubildung des Kolons	C18
C50.1	Bösartige Neubildung: Zentraler Drüsenkörper der Brustdrüse	Bösartige Neubildung der Brustdrüse	C50
C50.3	Bösartige Neubildung unterer innerer Quadrant Brustdrüse	Bösartige Neubildung der Brustdrüse	C50
C83.0	Non-Hodgkin-Lymphom: Kleinzellig (diffus)	Nicht follikuläres Lymphom	C83
D12.6	Gutartige Neubildung: Kolon, nicht näher bezeichnet	Gutartige Neubildung des Kolons, des Rektums, des Analkanals und des Anus	D12
D12.8	Gutartige Neubildung: Rektum	Gutartige Neubildung des Kolons, des Rektums, des Analkanals und des Anus	D12

Tabelle 3: Änderung des Informationswertes einer Diagnose bei Änderung des ICD durch Nichtangabe

Insofern gilt es die Anwender darüber aufzuklären, dass gewisse Daten verändert wurden:

8.3.2 Maskierung/Ersetzung

Zu schützende Daten werden mit einem konstanten oder sich ändernden Wert, Zeichen oder Zeichenkette ersetzt.

Beispiel: Die Daten aus Tabelle 1 wurden maskiert, indem der Tag und der Monat des Datums jeweils auf „01“ geändert wurden, die Namen auf feste Zeichenkette unter Beibehaltung der Geschlechterzuordnung:

Vorname	Nachname	Geschlecht	Geb.-Datum	PLZ	ICD
Anne	Musterfrau	w	01.01.1983	10115	C43.9
Max	Mustermann	m	01.01.1965	10115	D22.9
Max	Mustermann	m	01.01.1977	10178	C85.9
Max	Mustermann	m	01.01.1981	10247	D44.8
Max	Mustermann	m	01.01.1985	10319	C18.4
Anne	Musterfrau	w	01.01.1987	10407	D46.1
Max	Mustermann	m	01.01.1988	10435	C16.3
Max	Mustermann	m	01.01.1968	10439	D46.2
Anne	Musterfrau	w	01.01.1978	10585	C50.3
Anne	Musterfrau	w	01.01.1969	10707	C91.10
Max	Mustermann	m	01.01.1967	10717	D12.8
Max	Mustermann	m	01.01.1991	10717	D12.6
Anne	Musterfrau	w	01.01.1987	10787	C50.8
Anne	Musterfrau	w	01.01.1983	10827	C50.1
Anne	Musterfrau	w	01.01.1975	10963	C83.0

Tabelle 4: Maskiertes Geburtsdatum

8.3.3 Mischung/Shuffeling

Bei der Nutzung dieser Methode werden die in den Datensätzen enthaltenen Werte getauscht („verwürfelt“). Dabei ist zu beachten, dass etwaige, eine Person eindeutig identifizierende Informationen wie bspw. eine Telefonnummer oder eine Kreditkartennummer zur Auflösung des Personenbezugs noch zusätzlich mit einer weiteren Methode verfremdet werden müssen, um einen Personenbezug ausschließen zu können.

Die Grundlage für diese Durchmischung sollte eine Zufallsverteilung sein, die jedem Datenfeld die Daten bzw. Teilmenge der Daten eines anderen Datenfeldes zuordnet, wodurch letztlich ein neuer Datensatz gebildet wird. Bei einer zufälligen Vertauschung ist grundsätzlich nicht auszuschließen, dass ein Datensatz auf sich selbst abgebildet wird, wodurch im Ergebnis keine Veränderung stattfinden würde. Dies ist natürlich durch entsprechende Vorkehrungen auszuschließen.

Beispiel: Die Daten aus Tabelle 1 werden untereinander vermischt, um so die Identifizierung auszuschließen:

Original

Vorname	Nachname	Geb.-Datum	ICD
Käthe	Albers	27.05.1975	C83.0
Fieda	Fischer	15.11.1987	C50.8
Kunigunde	Gewaltig	21.01.1969	C91.10
Franz	Herrlich	17.11.1967	D12.8
Gerfriede	Jensen	23.07.1983	C50.1
Berthold	Koch	28.08.1991	D12.6
Michael	Matuschek	13.04.1968	D46.2
Hugo-Egon	Meyer	27.08.1977	C85.9
Uwe	Müller	31.03.1988	C16.3

Mischung

Vorname	Nachname	Geb.-Datum	ICD
Uwe	Albers	28.08.1991	C16.3
Michael	Fischer	31.03.1988	C18.4
Kunigunde	Gewaltig	15.11.1987	C43.9
Käthe	Herrlich	15.07.1987	C50.1
Jürgen	Jensen	29.11.1985	C50.3
Jan	Koch	23.07.1983	C50.8
Hugo-Egon	Matuschek	11.05.1983	C83.0
Hiltrud	Meyer	23.12.1981	C85.9
Heike	Müller	01.04.1978	C91.10

Original				Mischung			
Vorname	**Nachname**	**Geb.-Datum**	**ICD**	**Vorname**	**Nachname**	**Geb.-Datum**	**ICD**
Hiltrud	Niemand	15.07.1987	D46.1	Gerfriede	Niemand	27.08.1977	D12.6
Heike	Richter	11.05.1983	C43.9	Franz	Richter	27.05.1975	D12.8
Anke	Schmidt	01.04.1978	C50.3	Fieda	Schmidt	21.01.1969	D22.9
Eckbert	Schneider	23.12.1981	D44.8	Eckbert	Schneider	13.04.1968	D44.8
Jan	Schröder	03.12.1965	D22.9	Berthold	Schröder	17.11.1967	D46.1
Jürgen	Stillstand	29.11.1985	C18.4	Anke	Stillstand	03.12.1965	D46.2

Tabelle 5: Vermischung der Datensätze, so dass eine Identifizierung nicht möglich ist

8.3.4 Varianzmethode

Bei dieser Methode werden Daten, die auf Zahlen basieren, dadurch verfremdet, dass die Zahlenwerte in festgelegten, zufällig erhöhten oder verringerten Streuungsintervallen geändert werden.

Beispiel: Das Geburtsdatum aus Tabelle 1 wurde mittels der Varianzmethode bearbeitet, wodurch das Geburtsdatum willkürlich verändert wurde, die statistische Aussage der Tabelle jedoch erhalten blieb:

Vorname	Nachname	Geschlecht	Geb.-Datum	Geb.-Datum
Heike	Richter	w	11.05.1983	14.05.1983
Jan	Schröder	m	03.12.1965	05.12.1965
Hugo-Egon	Meyer	m	27.08.1977	29.08.1977
Eckbert	Schneider	m	23.12.1981	21.12.1981
Jürgen	Stillstand	m	29.11.1985	30.11.1985
Hiltrud	Niemand	w	15.07.1987	16.07.1987
Uwe	Müller	m	31.03.1988	03.04.1988
Michael	Matuschek	m	13.04.1968	18.04.1968
Anke	Schmidt	w	01.04.1978	29.03.1978
Kunigunde	Gewaltig	w	21.01.1969	17.01.1969
Franz	Herrlich	m	17.11.1967	22.11.1967
Berthold	Koch	m	28.08.1991	01.09.1991
Fieda	Fischer	w	15.11.1987	18.11.1987
Gerfriede	Jensen	w	23.07.1983	26.07.1983
Käthe	Albers	w	27.05.1975	30.05.1975

Tabelle 6: Anpassung des Geburtsdatums durch die Varianzmethode

8.3.5 Kryptografische Methoden

Hierbei kommen Verschlüsselungs- und/oder Hash-Algorithmen zum Einsatz. Dabei ist zu beachten, dass kryptografische Eigenschaften wie Blocklänge, Ausgabealphabet und Kollisionen der jeweils verwendeten Methoden Auswirkungen auf das Ergebnis der Anonymisierung haben. Weiterhin ist zu beachten, dass hier kryptografische Methoden im speziellen Kontext der Anonymisierung bzw. Pseudonymisierung betrachtet werden, d. h. einige Betrachtungen im anderen Kontext ggf. zu anderen Ergebnissen führen können.

8.3.5.1 Rahmenbedingungen abklären

Mit der Fachseite, welche die pseudonymisierten oder anonymisierten Daten verarbeiten will, müssen die Randbedingungen geklärt werden. So muss z. B. geklärt werden, ob

- der Zeichensatz (arabisch, deutsch, …),
- die Zeichenart (numerisch, Buchstabenerhalt, Sonderzeichen bleibt Sonderzeichen),
- Zeichenlänge

bei der Pseudonymisierung/Anonymisierung erhalten bleiben sollen?
Weiterhin können funktionelle Anforderungen wie z. B.

- Kollisionsfreiheit; d. h. unterschiedliche Eingaben führen immer zu unterschiedlichen Ergebnissen, so dass die Unterschiede erhalten bleiben und ggf. Datensätze trotz Pseudonymisierung/Anonymisierung zusammengeführt werden können,
- Eindeutigkeit; gleiche Eingaben führen immer zu gleichen Abbildungen,
- Erhalt der statistischen Verteilung

hinzukommen.

Grundsätzlich gilt: Je mehr Randbedingungen an die Pseudonymisierung bzw. Anonymisierung seitens der Fachseite gestellt werden, umso größer wird das Risiko der Re-Identifizierbarkeit durch statistische Analysen.

8.3.5.2 Verschlüsselungsverfahren

Moderne kryptografische Methoden sind nahezu ausschließlich Binärchiffren, die sich in Block- und Stromchiffren sowie in symmetrische und asymmetrische Verfahren unterteilen lassen. Hierbei ist Folgendes zu beachten:

1) Stromchiffren müssen zum Erhalt von Eigenschaften mehrfach denselben Schlüsselstrom verwenden, was die kryptografische Stärke der Verfahren abschwächt. Daher sind Stromchiffren für die Pseudonymisierung/Anonymisierung i.d.R. eher ungeeignet.
2) Den Vorteilen im Umgang mit dem Schlüsselmaterial stehen bei asymmetrischen Verfahren sehr hohe Performance-Einbußen und relativ große Chiffrat-Blöcke entgegen.

Dabei erhalten Binärchiffren weder den Zeichensatz noch die Zeichenart oder die Zeichenlänge. Jedoch sind Binärchiffren sowohl kollisionsfrei als auch eindeutig.

Andere Verschlüsselungsverfahren können Anforderungen bzgl. Zeichenart, Zeichensatz und Zeichenlänge ggf. erhalten. Dieses ist z. B. bei entsprechender Implementierung beim symmetrischen Verfahren „One-Time-Pad“ der Fall. Hier wiederum kann ggf. die Anforderung der Eindeutigkeit nicht mehr gegeben sein.

D. h., ob eine Verschlüsselung zur Anonymisierung/Pseudonymisierung genutzt werden kann, ist abhängig von den Anforderungen der Fachabteilung.

8.3.5.3 Hash-Funktionen

Hash-Funktionen (auch: kryptografische Checksumme oder Einwegfunktion genannt) bilden eine beliebig lange Eingabedatenmenge auf einen binären String fester Länge ab. Somit können auch Hash-Funktionen weder den Zeichensatz noch die Zeichenart oder die Zeichenlänge erhalten. Die Möglichkeit der Kollisionsfreiheit ist abhängig von der Ausgabelänge und dem Algorithmus. Bei MD5 beispielsweise ist nachgewiesen, dass Kollisionsfreiheit nicht gegeben ist. Die Anforderung der Eindeutigkeit wird von Hash-Funktionen gewährleistet.

8.3.5.4 Salt

„Salt“ (= „Salz“) bezeichnet in der Kryptografie eine zufällig gewählte Zeichenfolge, die an einen gegebenen Klartext vor der Verwendung als Eingabe einer Hash-Funktion angehängt wird, um die Entropie der Eingabe zu erhöhen, was letztlich zu einer höheren Streuung des Ergebnisses führt. Hierdurch kann z. B. verhindert werden, dass Originaldaten bspw. mit Hilfe von Rainbow-Tabellen identifiziert werden können.

Stand heute wird eine Entropie von 100 Bit als resistent gegen Brute-Force Angriffe mit hohem Angriffspotential angesehen. D. h. der Wertebereich muss eine Mindeststreuung von 2^{100} bzw. 10^{30} haben. Zu jedem Datensatz sollte ein eigener Salt existieren, um den größtmöglichen Schutz zu erhalten. Werden alle Datensätze mit ein und derselben Zeichenfolge kombiniert, so wird dies als „Pepper“ bezeichnet.

Bei der Überprüfung eines Datums wird jedoch nicht jedes Mal ein neuer Salt erzeugt, da sich sonst der entstandene Hashwert von dem gespeicherten unterscheidet und somit der Wahrheitsgehalt der Information nicht überprüft werden kann. D. h. die Anforderung der Eindeutigkeit wäre nicht mehr gegeben. Deshalb wird – sofern die Eindeutigkeit eine einzuhaltende Anforderung darstellt - bei der Generierung der zur jeweiligen Information verwendete Salt zusammen mit dem entstandenen Hashwert gespeichert. Dabei müssen Salt und Hashwert natürlich voneinander getrennt aufbewahrt werden, der Salt unbedingt geheim gehalten werden, da ansonsten der Schutz abgeschwächt wird.

8.3.6 Was wird wann mit welcher Methode erreicht?

Methode	Anonymisierung	Pseudonymisierung
Nichtangabe	Nichtangabe sorgt immer für Anonymität bzgl. des betreffenden Datums	Keine Pseudonymisierung möglich
Maskierung/Ersetzung	Bei Vorgehen – Ersetzen mit gleichbleibendem Wert – Ersetzen mit sich erhöhendem Wert – Zufällige Mischung (Initialisierungsschlüssel vernichtet)	Verwendung – eines Schemas – von pseudozufälliger Ersetzung, z. B. schlüsselabhängige Ersetzung
Mischung/Shuffeling	Bei Vorgehen – Zufällige Mischung (Initialisierungsschlüssel vernichtet)	Verwendung von – pseudozufällige Ersetzung, z. B. schlüsselabhängige Ersetzung
Varianzmethode	Bei ausreichend großer Varianz	Vorgehen mit schlüsselabhängiger Abweichung; Schlüssel wird aufbewahrt
Kryptografische Methoden	Bei Vorgehen – Schlüssel wird vernichtet – Verfahren ist nicht invertierbar	Bei Vorgehen – Schlüssel wird sicher aufbewahrt

8.3.7 k-Anonymität

Direkte und indirekte Identifikationsmerkmale werden zu Gruppen mit gleichen Inhalten zusammengefasst, d. h. die Identifikationsmerkmale so verändert, dass die Merkmale zu Gruppen zusammengefasst werden können. Damit sind die hinter den Daten stehenden Individuen nicht mehr unterscheidbar, d. h. eine eindeutige Identifikation ist nicht mehr möglich.

Um k-Anonymität zu erreichen, können alle oben beschriebenen Methoden eingesetzt werden. Dabei gilt: Je größer die Gruppe, je größer ist das Maß an Anonymität bzw. je kleiner ist die Wahrscheinlichkeit als Angehöriger einer Gruppe mit bestimmten Merkmalen identifiziert zu werden.

Der Parameter k definiert bei der k-Anonymität die Mindestgröße der Gruppen. Er ist damit gleichzeitig das Maß der Anonymität. In einer Gruppe von k Individuen liegt die Wahrscheinlichkeit bei 1/k ein einzelnes Individuum korrekt zu identifizieren.

In der Literatur wird ein Schwellwert von mindestens 5 angegeben, d. h.: Bei jeder Auswertung umfasst das Ergebnis zu jedem Zeitpunkt des Auswertungszeitraumes mindestens 5 Betroffene, sodass kein Rückschluss auf Einzelpersonen gegeben ist. Kann eine Rückführbarkeit auf eine Personengruppe unter 5 Personen nicht ausgeschlossen werden, sind sowohl der Schwellwert aus auch die Merkmale/Items für die jeweilige Auswertung so zu definieren, dass trotzdem der Identifikationsschutz gewährleistet ist.

8.3.8 Beispiele bzgl. Vorgehen

Datentyp	Methode
Zahl	– Neuvergabe der letzten x Stellen (x = abhängig von den Zahlenwerten) – Ersetzen durch Zufallszahlen – Nutzung einer Varianz (z. B. ± x%) – Löschung
String	– Neuvergabe über Tabelle – Ersetzung durch feste Zeichenkette – Ersetzung durch feste Zeichenkette mit laufender Nummer zwecks Beibehaltung der Unterscheidbarkeit
Datum	– Setzen von Tag und Monat auf festen Wert – Setzen des Datums auf einen festen Wert
Postleitzahl	– Neuvergabe von mindestens den letzten 2 Stellen über Umsetzungstabelle – Ersetzen von mindestens den letzten beiden Stellen durch festen Wert – Ersetzen von mindestens den letzten beiden Stellen durch festen Zufallswert
E-Mail-Adresse	– Löschen – Ersetzen durch festen Dummy-Wert
Religion	– Löschen – Ersetzen durch festen Dummy-Wert
Medizinische Code-Systeme wie ICD, OPS, usw.	– Verkürzen der Kodierung – Löschung

Tabelle 7: Beispiel bzgl. Ersetzen von Datentypen

8.4 Darstellung des Risikos der Re-Identifizierung

Eine Re-Identifizierung kann Teil des geplanten Ablaufes sein, wenn eine Re-Identifikation unter zuvor festgelegten Bedingungen beabsichtigt erfolgt, z. B. Kontaktierung des Patienten, da die Verarbeitungsergebnisse Einfluss auf seine Behandlung haben.

Erfolgt eine Re-Identifikation als ungeplantes, insbesondere nicht beabsichtigtes Ereignis, kann eine derartige Re-Identifikation Risiken für die betroffene Person bergen.

8.4.1 Risikodarstellung

Entscheidend für die Beurteilung des Risikos ist bereits das Vorhandensein der abstrakten Möglichkeit zur Identifikation von Betroffenen. Dabei sind die wesentlichen Risikofaktoren für eine Re-Identifizierung statistische Strukturen und Zusatzwissen. Selbst wenn Datensätze hoch effektiv und nach den aktuellsten kryptologischen Methoden geschützt sind, können statistische Auffälligkeiten dazu führen, dass ein Personenbezug - ggf. auch nur teilweise - wiederhergestellt werden kann. Dieses Risiko wird durch Verfügbarkeit von Zusatzwissen erheblich verstärkt.

Beispiel: Die nachfolgende Tabelle scheint zunächst anonyme Daten zu enthalten.

Geschlecht	Geb.-Datum	PLZ	ICD
w	01.01.1983	1011	C43
m	01.01.1965	1011	D22
m	01.01.1977	1017	C85
m	01.01.1981	1024	D44
m	01.01.1985	1031	C18
w	01.01.1987	1040	D46
m	01.01.1988	1043	C16
m	01.01.1968	1043	D46
w	01.01.1978	1058	C50
w	01.01.1969	1070	C91
m	01.01.1967	1071	D12
m	01.01.1991	1071	D12
w	01.01.1987	1078	C50
w	01.01.1983	1082	C50
w	01.01.1975	1096	C83

Tabelle 8: Auf Anonymität zu prüfendes Ergebnis

Vorname	Nachname	Geschlecht	Geb.-Datum	PLZ
Heike	Richter	w	11.05.1983	10115
Jan	Schröder	m	03.12.1965	10115
Hugo-Egon	Meyer	m	27.08.1977	10178
Eckbert	Schneider	m	23.12.1981	10247
Jürgen	Stillstand	m	29.11.1985	10319
Hiltrud	Niemand	w	15.07.1987	10407
Uwe	Müller	m	31.03.1988	10435
Michael	Matuschek	m	13.04.1968	10439
Anke	Schmidt	w	01.04.1978	10585

Vorname	Nachname	Geschlecht	Geb.-Datum	PLZ
Kunigunde	Gewaltig	w	21.01.1969	10707
Franz	Herrlich	m	17.11.1967	10717
Berthold	Koch	m	28.08.1991	10717
Fieda	Fischer	w	15.11.1987	10787
Gerfriede	Jensen	w	23.07.1983	10827
Käthe	Albers	w	27.05.1975	10963

Tabelle 9: Zuordnungsmöglichkeiten durch die Originaldaten

Existiert hingegen auch nur ein teilweise möglicher Zugriff auf die Originaldaten aus Tabelle 9, so ist eine Re-Identifikation der Daten aus Tabelle 8 möglich: Nur Frau Richter und Herr Schröder wohnen in einem Bereich, dessen PLZ mit „1011“ beginnt und durch die Geschlechtsangabe ist eine eindeutige Zuordnung möglich. Die Daten sind also nicht als anonyme Daten anzusehen, sondern als pseudonyme Daten.

8.4.2 Grundbedingung für eine Prüfung

Bei einer Prüfung des Ergebnisses einer Pseudonymisierung/Anonymisierung muss sowohl das Vorgehen der Verarbeitung, also die Methodik und der Umsetzung der Methodik beurteilt werden, wie das vorhandene Ergebnis. Dafür ist es erforderlich, dass alles nachvollziehbar dokumentiert wurde. Die/Der Prüfende benötigt:

- Eine ausführliche Dokumentation der Methodik,
- eine detaillierte Darstellung der Umsetzung der Methodik,
- die Ergebnistabellen.

Bei den Ergebnissen ist darauf zu achten, dass auch statistische Kennzahlen Informationen zu einzelnen Individuen beinhalten können:

- Ein Mittelwert basierend auf wenigen Beobachtungen kann ggf. Rückschlüsse bzgl. gering besetzter Gruppen beinhalten. Um diese Randgruppen identifizieren und diese bzgl. einer Möglichkeit des Rückschlusses auf einzelne Individuen prüfen zu können, muss neben der Fallzahl immer auch Minimum, Maximum und Standardabweichung angegeben werden.
- Die Angabe von Perzentilen (Prozenträngе) bei einer geringen Fallzahl beinhaltet nahezu immer die Möglichkeit von Rückschlüssen auf einzelne.

Hochfellner [18] empfiehlt als Mindestgrößen:

- Mindestens 20 Beobachtungen für die Ausgabe von Mittelwerten
- Mindestens 40 Beobachtungen für die Ausgabe von 50%-Perzentilen
- Mindestens 80 Beobachtungen für die Ausgabe von 25%- oder 75%-Perzentilen
- Mindestens 200 Beobachtungen für die Ausgabe von 10%- oder 90%-Perzentilen
- Mindestens 400 Beobachtungen für die Ausgabe von 5%- oder 95%-Perzentilen
- Mindestens 2000 Beobachtungen für die Ausgabe von 1%- oder 99%-Perzentilen

[18] Hochfellner et al. (2012) FDZ-Methodenreporte: Datenschutz am Forschungsdatenzentrum. (Hrsg.: Bundesagentur für Arbeit) [Online, zitiert am 2018-04-24]; Verfügbar unter http://doku.iab.de/fdz/reporte/2012/MR_06-12.pdf

8.4.3 Risikobeurteilung

Gemäß den Anforderungen von Art. 25 DS-GVO sind sowohl „zum Zeitpunkt der Festlegung der Mittel für die Verarbeitung [also der Planung der Verarbeitung] als auch zum Zeitpunkt der eigentlichen Verarbeitung geeignete technische und organisatorische Maßnahmen“ zu treffen, welche die Rechte der betroffenen Personen schützen. Auch Art. 32 DS-GVO verlangt abhängig vom Risiko für die betroffenen Personen die Ergreifung geeigneter Maßnahmen zum Schutz der betroffenen Person.

Daher muss zwingend das Risiko, welches durch eine Re-Identifikation für die betroffene Person existieren kann, beurteilt werden. Dazu werden Risiken am besten einerseits in Kategorien eingeteilt, welche eine Zuordnung der Risiken in individuelle und Individuen übergreifende Risiken erlaubt, z. B. sein[19]:

- Strukturelle Risiken, beispielsweise gesellschaftlich-politische Risiken (wie z. B. die Informationsmacht, die gegenüber einem Individuum gewonnen wird) oder wirtschaftliche Risiken;
- Individuelle Risiken, wie z. B. die Erhöhung individueller Verletzlichkeit für Straftaten, da jemand erfährt, wo betroffene Personen angreifbar sind;
- Risiken für Gesellschaft und Individuum, z. B. durch Bildung von Persönlichkeitsprofilen oder Fremdbestimmung oder auch die Enttäuschung von Vertraulichkeitserwartungen.

Weiterhin müssen Risiken hinsichtlich der Bedeutung erfasst werden, d. h. eine Quantifizierung vorgenommen werden. Naturgemäß wird das jeweilige Risiko nur abgeschätzt werden können, sodass das Risiko entsprechend einem zuvor definierten Skalenniveau eingeteilt werden kann, z. B.:

Bewertung	Kriterien	Geschätzte Kosten
Katastrophal	Keine Kontrolle möglich	> 1 Mill. €
Kritisch	Gravierende Mängel / Schäden	≤ 1. Mill. €
Mittelmäßige Auswirkungen	Beträchtliche Abweichungen vom Soll	50 – 100.000 €
Geringe Auswirkungen	Geringe Folgen	< 50.000 €
Vernachlässigbare Auswirkungen	Unbedeutende Folgen	Keine

Entsprechend ErwGr. 26 DS-GVO wird eine objektive Bewertung bzgl. des Risikos verlangt. Die deutsche Datenschutzkonferenz veröffentlichte ein Kurzpapier „Risiko für die Rechte und Freiheiten natürlicher Personen“[20], welches auch Hinweise bzgl. der Schweregradbeurteilung aus Sicht der deutschen Aufsichtsbehörden enthält.

8.5 Aufbau und Struktur einer Verfahrensbeschreibung

Entsprechend Art. 30 DS-GVO müssen Verarbeitungstätigkeiten in einem Verzeichnis geführt werden. Dies gilt selbstverständlich auch für die Verarbeitung im Rahmen einer Pseudonymisierung oder Anonymisierung. Darüber hinaus enthält Art. 32 Abs. 3 DS-GVO eine implizite Aufforderung, dass die ergriffenen technischen und organisatorischen Maßnahmen, zu denen sowohl die

[19] Stefan Drackert (2014) Die Risiken der Verarbeitung personenbezogener Daten - Eine Untersuchung zu den Grundlagen des Datenschutzrechts. Duncker & Humblot GmbH. ISBN '978-3-428-1 4730-4

[20] Konferenz der unabhängigen Datenschutzbehörden des Bundes und der Länder (Datenschutzkonferenz - DSK): Kurzpapier 18 „Risiko für die Rechte und Freiheiten natürlicher Personen“. [Online, zitiert am 2018-05-18]; Verfügbar unter https://www.ldi.nrw.de/mainmenu_Aktuelles/submenu_EU-Datenschutzreform/Inhalt/EU-Datenschutzreform/KP_18_Risiko.pdf

Pseudonymisierung als auch die Anonymisierung gehören, einer Nachweisfähigkeit genügen müssen. Somit ist es erforderlich, dass die Verfahren hinreichend genau beschrieben werden.

Eine entsprechende Beschreibung der Durchführung einer Pseudonymisierung bzw. Anonymisierung sollte folgende Informationen enthalten[21, 22]:

- Beschreibung der Verarbeitung, für welche die Daten erhoben wurden;
- Beschreibung der Verarbeitung, für welche die Anonymisierung oder Pseudonymisierung benötigt wird;
- ID des für die Verarbeitung der personenbezogenen Daten Verantwortlichen;
- ID des für die Verarbeitung der anonymisierten oder pseudonymisierten Daten Verantwortlichen;
- Beschreibung des Verfahrens, mit welchem eine Anonymisierung oder Pseudonymisierung durchgeführt wird;
- Beschreibung, welche Daten für die Anonymisierung oder Pseudonymisierung ausgewählt wurden sowie eine Begründung, warum diese Daten relevant bzgl. einer Identifikationsmöglichkeit waren, andere nicht;
- ID der Person oder des automatisierten IT-Systems, welches die Anonymisierung oder Pseudonymisierung durchführt;
- Bei einer Anonymisierung der Nachweis der Anonymität, d. h. der Nachweis der Nicht-Beziehbarkeit der verarbeiteten Daten auf eine identifizierte oder identifizierbare natürliche Person;
- Umgang mit ggf. zur Anonymisierung oder Pseudonymisierung genutzten kryptographischen Schlüssel oder einer entsprechenden Verknüpfungstabelle; hierzu gehört insbesondere auch
    - eine Beschreibung dessen, was geschieht, wenn die Organisation ihren Betrieb hinsichtlich der Anonymisierungs- oder Pseudonymisierungsaktivitäten einstellt,
    - eine Beschreibung, in welchen Bereichen und für welche Anwendungen die kryptographischen Schlüssel oder die entsprechenden Verknüpfungstabelle verwendet werden
    - eine Beschreibung des Gültigkeitszeitraum (aus welchem sich letztlich auch der späteste Zeitpunkt zur Validierung der durchgeführten Pseudonymisierung oder Anonymisierung ergibt),
    - eine Beschreibung der Möglichkeiten und Verfahren zur Verknüpfung mit Alt-Daten oder neu hinzugekommenen Daten, sofern die Möglichkeit vorhanden ist;

[21] Siehe auch DIN EN ISO 25237:2017: Medizinische Informatik – Pseudonymisierung. Erhältlich z. B. online beim Beuth-Verlag unter https://www.beuth.de/de/norm/din-en-iso-25237/258588981

[22] Beispiele bzgl. Policy findet man z. B.

- NHS Business Services Authority: Pseudonymisation and anonymisation of data .policy. [Online, zitiert am 2018-06-20]; Verfügbar unter https://www.nhsbsa.nhs.uk/sites/default/files/2017-05/anonymisation-of-data-policy.pdf
- Mayo Clinic: De-identification and Re-identification of Protected Health Information. [Online, zitiert am 2018-06-20]; Verfügbar unter https://www.mayoclinic.org/documents/deidentification-jax-pdf/doc-20079518
- Sanofi: Clinical Trial Data Sharing Data De-Identification Guidelines. [Online, zitiert am 2018-06-20]; Verfügbar unter https://www.clinicalstudydatarequest.com/Documents/Sanofi-DeIdentification-Guide.pdf

- Ausführliche Beschreibung, unter welchen Umständen die Pseudonymisierung durch wen auf welche Art umkehrbar ist und welche Berechtigung hierzu von wem erforderlich ist;
- Festlegung der Beschränkungen, denen der Empfänger der anonymisierten oder pseudonymisierten Daten unterliegt, z. B. vertragliche Regelungen oder die vereinbarten Verarbeitungsgrundsätze zu informationsbezogenen Aktionen mit diesen Daten, insbesondere bzgl. Weiterleitung und Aufbewahrung wie beispielsweise:
    - Der Empfänger darf die Daten nicht öffentlich zugänglich machen.
    - Der Empfänger muss die Daten vor unberechtigtem Zugriff schützen.
    - Der Empfänger darf die Daten nur intern nutzen, um entpersonalisierte Daten zu erzeugen und erst diese dürfen öffentlich zugänglich gemacht oder an Kunden veräußert werden.
    - Der Empfänger muss die Daten zerstören, wenn die Verarbeitung hinsichtlich der vereinbarten Zwecke beendet wurde und kein weiterer rechtlicher Aufbewahrungsgrund für die Daten mehr existiert.

9 Checkliste

9.1 Organisatorische Anforderungen

Anforderung	Erfüllt	Nicht erfüllt
Es werden alle Daten pseudonymisiert/anonymisiert verarbeitet oder es existiert eine Begründung, warum eine pseudonyme Verarbeitung nicht möglich ist		
Es ist gewährleistet, dass eine dezidierte Analyse zur Beurteilung der Kritikalität der Daten und des Aufwandes zur Anonymisierung bzw. Pseudonymisierung durchgeführt wird und die Ergebnisse dokumentiert sind, sodass jederzeit eine Überprüfung der Ergebnisse erfolgen kann		
Die Pseudonymisierung/Anonymisierung erfolgt im jeweiligen Quellsystem		
Es ist gewährleistet, dass personenbezogenen Daten vor einer Pseudonymisierung/Anonymisierung ausschließlich zu Zwecken der Prüfung auf Inplausibilitäten oder Doppelungen zwischengespeichert werden		
Es ist gewährleistet, dass die Prüfung auf Inplausibilitäten und Dopplungen im Vorfeld einer Pseudonymisierung/ Anonymisierung grundsätzlich automatisiert erfolgt		

9.2 Vorgaben für das Verfahren

Anforderung	Erfüllt	Nicht erfüllt
Es ist sichergestellt, dass das eingesetzte Identifikationsschutzverfahren auch Freitextfelder, Kommentarfelder, Anlagen, etc. im Hinblick auf die Ersetzung von personenbezogenen Daten berücksichtigt		
Es ist gewährleistet, dass der Aufwand zur De-Pseudonymisierung bzw. De-Anonymisierung für jedermann unverhältnismäßig hoch (d. h. faktisch ausgeschlossen) ist		
Es ist gewährleistet, dass der eingesetzte Verfremdungsprozess so gestaltet ist, dass er das notwendige Identifikations-Schutzverfahren wirksam umsetzt		
Es ist beim eingesetzten Verfremdungsprozess sichergestellt, dass mittels der verfügbaren Datenfelder eine Rückführbarkeit nur auf Gruppen mit mindestens 5 Personen möglich ist		
Es ist gewährleistet, dass die verwendete Verfremdungsmethode vertrauenswürdig, sicher implementiert und dokumentiert sowie allen beteiligten Parteien bekannt und funktional überprüfbar ist		

9.3 Nichtangabe

Anforderung	Erfüllt	Nicht erfüllt
Daten, die einen Personenbezug ermöglichen, werden gelöscht		

9.4 Maskierung/Ersetzung

Anforderung	Erfüllt	Nicht erfüllt
Daten, die einen Personenbezug ermöglichen, werden durch andere konstante oder sich ändernde Werte ersetzt		

9.5 Mischung/Shuffeling

Anforderung	Erfüllt	Nicht erfüllt
Es ist sichergestellt, dass die Mischung unter Änderung aller Datensätze und Datenfelder erfolgt		
Daten, die schon für sich eindeutig personenbeziehbar sind und durch eine Mischung nicht veränderbar sind, werden zusätzlich mit einer anderen Verfremdungsmethode bearbeitet		

9.6 Varianzmethode

Anforderung	Erfüllt	Nicht erfüllt
Die jeweilige Erhöhung oder Verringerung basiert auf Zufallswerten		

9.7 Kryptografische Methoden

9.7.1 Verschlüsselung

Anforderung	Erfüllt	Nicht erfüllt
Es ist gewährleistet, dass die Erzeugung des Schlüssels bzw. Schlüsselmaterials ein sicherer Prozess ist		
Es ist gewährleistet, dass der Erzeugung des Schlüssels bzw. Schlüsselmaterials eine qualitativ hochwertige Zufallszahlenquelle zugrunde liegt		
Es ist sichergestellt, dass der Schlüssel bzw. das Schlüsselmaterial derart erzeugt wird, dass diese weder vorhersagbar sind noch erraten werden können		
Es ist gewährleistet, dass die Vertraulichkeit des Schlüssels bzw. des Schlüsselmaterials während des vollständigen Lebenszyklus der verarbeiteten personenbezogenen Daten gewährleistet ist		
Es ist sichergestellt, dass der Zugriff auf den Schlüssel bzw. das Schlüsselmaterial auf ein absolutes Minimum vertrauenswürdiger Anwender beschränkt ist		
Es werden ausschließlich Standard-Verschlüsselungs-Algorithmen entsprechend den Empfehlungen des BSI bzw. der Bundesnetzagentur verwendet		
Es ist sichergestellt, dass das verwendete Verfahren eine hinreichende Stärke sowie keinerlei bekannte Schwächen aufweist		
Es ist sichergestellt, dass der für die Verschlüsselungsalgorithmen verwendete Schlüssel von ausreichender Qualität ist		

Anforderung	Erfüllt	Nicht erfüllt
Es ist gewährleistet, dass der Schlüssel geheim gehalten wird		
Es liegt ein Konzept zum Schlüsselmanagement vor und dieses enthält Informationen zum Schlüsseltausch, zur Feststellung von und Vorgehensweisen bei Kompromittierung		

9.7.2 Hash-Funktionen

Anforderung	Erfüllt	Nicht erfüllt
Es ist sichergestellt, dass ausschließlich Standard-Hash-Funktionen verwendet werden, für die es keine bekannten Schwachstellen gibt		
Es ist sichergestellt, dass bei Verwendung von Hash-Funktionen ein Salt benutzt wird		
Es ist sichergestellt, dass der Salt derart erzeugt wird, dass dieser weder vorhersagbar ist noch erraten werden kann		
Es ist sichergestellt, dass der Zugriff auf den Salt auf ein absolutes Minimum vertrauenswürdiger Anwender beschränkt ist		
Es ist sichergestellt, dass der Salt von ausreichender Qualität ist (= Mindestentropie von 100 Bit)		
Es ist gewährleistet, dass der Salt geheim gehalten wird		

10 Abkürzungen

Abs.	Absatz
Art.	Artikel (Einzahl)
Artt.	Artikel (Plural)
BDSG a.F.	Bundesdatenschutzgesetz, in der Fassung gültig bis einschließlich 24. Mai 2018
BDSG n.F.	Bundesdatenschutzgesetz, in der Fassung gültig ab dem 25. Mai 2018
DICOM	Digital Imaging and Communications in Medicine
DIN	Deutsche Institut für Normung e. V.
DSFA	Datenschutz-Folgenabschätzung
DSG	Datenschutzgesetz
DS-GVO	Datenschutz-Grundverordnung
EDV	Elektronische Datenverarbeitung
EN	Europäische Norm
ErwGr.	Erwägungsgrund/Erwägungsgründe
EU	Europäische Union
GMDS	Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V.
HIPAA	Health Insurance Portability and Accountability Act
i.d.R.	in der Regel
i.S.d.	Im Sinne des/der
ISO	International Organization for Standardization
i.S.v.	im Sinne von
IT	Informationstechnik, informationstechnisches…
ICD	Internationale Klassifikation der Krankheiten (International Classification of Diseases)
ID	Identifikator/Identifier, Identifikationsnummer
Kap.	Kapitel
LDSG	Landesdatenschutzgesetz
lit.	littera (lat. „Buchstabe“)
PIN	Persönliche Identifikationsnummer
Nr.	Nummer
S.	Satz
StudyUID	Study Unique Identifiers (Untersuchungs-UID)
TAN	Transaktionsnummer
Ziff.	Ziffer

11 Glossar

Aggregierte Daten	Zusammenfassung von Einzelwerten zu größeren Einheiten; ein Rückschluss auf die Einzeldaten ist i.d.R. nicht mehr möglich
Anonymisierte Daten	Daten, die als Ergebnis eines Anonymisierungsprozesses erzeugt wurden (Quelle: DIN EN ISO 25237)
Anonymisierung	Prozess, bei dem personenbezogene Daten entweder vom für die Verarbeitung der Daten Verantwortlichen allein oder in Zusammenarbeit mit einer anderen Partei (Auftragsverarbeiter) unumkehrbar so verändert werden, dass sich die betroffene Person danach weder direkt noch indirekt identifizieren lässt (Quelle: DIN EN ISO 25237)
Auftragsverarbeiter	„'Auftragsverarbeiter' eine natürliche oder juristische Person, Behörde, Einrichtung oder andere Stelle, die personenbezogene Daten im Auftrag des Verantwortlichen verarbeitet" (Quelle: Art. 4 Ziff. 8 DS-GVO)
Automatische Verarbeitung	Verarbeitung unter Nutzung von EDV; also z. B. Word- oder Excel-Datei, aber auch KIS, RIS, PACS, unabhängig ob Client-Server-Lösung oder Stand-alone PC, Tablet oder anderweitige Hardware genutzt wird
Betroffener/betroffene Person	Genau genommen „betroffene Person", in der Literatur aber häufig als "Betroffener" aufgeführt; „'Personenbezogene Daten' alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (im Folgenden „betroffene Person") beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind, identifiziert werden kann" (Quelle: Art. 4 Ziff. 1 DS-GVO)
Data linkage	Abgleich und Zusammenführung von Daten aus mehreren Datenbanken (Quelle: DIN EN ISO 25237)
Datenverknüpfung	Siehe „Data linkage"
De-Anonymisierung	(Gezielte) Aufhebung einer zuvor durchgeführten Anonymisierung von Daten
Einwilligung	„'Einwilligung' der betroffenen Person jede freiwillig für den bestimmten Fall, in informierter Weise und unmissverständlich abgegebene Willensbekundung in Form einer Erklärung oder einer sonstigen eindeutigen bestätigenden Handlung, mit der die betroffene Person zu verstehen gibt, dass sie mit der Verarbeitung der sie betreffenden personenbezogenen Daten einverstanden ist" (Quelle: Art. 4 Ziff. 11 DS-GVO)
Entpersonalisierung	allgemeine Benennung für jeden Prozess der Reduktion der Zuordnung zwischen einer Menge von zur Identifizierung geeigneten Daten und der betroffenen Person (Quelle: DIN EN ISO 25237)
Genetische Daten	„'Genetische Daten' personenbezogene Daten zu den ererbten oder erworbenen genetischen Eigenschaften einer natürlichen Person, die eindeutige Informationen über die Physiologie oder die Gesundheit dieser natürlichen Person liefern und insbesondere aus der Analyse einer biologischen Probe der betreffenden natürlichen Person gewonnen wurden" (Quelle: Art. 4 Ziff. 13 DS-GVO)

Gesundheitsdaten	„'Gesundheitsdaten' personenbezogene Daten, die sich auf die körperliche oder geistige Gesundheit einer natürlichen Person, einschließlich der Erbringung von Gesundheitsdienstleistungen, beziehen und aus denen Informationen über deren Gesundheitszustand hervorgehen“ (Quelle: Art. 4 Ziff. 15 DS-GVO)
Hashfunktion	Eindeutige Prüfsumme mit festgelegter Länge (Hash-Wert) einer Datei oder eines Datensatzes
identifizierbare Person	Jemand, der direkt oder indirekt identifiziert werden kann, insbesondere über die Referenz zu einer Identifizierungsnummer oder zu einem oder mehreren Kennzeichen, die bezüglich seiner körperlichen, physiologischen, geistigen, ökonomischen, kulturellen oder sozialen Identität spezifisch sind (Quelle: DIN EN ISO 25237)
Identifizierung	Prozess der Nutzung von behaupteten oder beobachteten Attributen einer juristischen Person mit dem Ziel, diese aus den anderen juristischen Personen in einer Reihe von Identitäten herauszufinden (Quelle: DIN EN ISO 25237)
Normadressat	Rechtssubjekt (z. B. natürliche Person, juristische Person, Personenvereinigung), an die sich die Regelung eines Gesetzes (= einer Norm) richtet
Personenbezeichner	Informationen, deren Zweck darin besteht, eine Person innerhalb eines bestimmten Kontexts eindeutig zu identifizieren (Quelle: DIN EN ISO 25237)
Personenidentifizierung	Prozess der Aufstellung einer Verbindung zwischen einem Informationsobjekt und einer physischen Person (Quelle: DIN EN ISO 25237)
Pseudonym	Personenbezeichner, 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 (Quelle: DIN EN ISO 25237)
Pseudonymisierung	„'Pseudonymisierung' die Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden“ (Quelle: Art. 4 Ziff. 5 DS-GVO)
Re-Identifikation	(Gezielte) Aufhebung einer zuvor durchgeführten Pseudonymisierung oder Anonymisierung von Daten
Salt	Zufällig gewählte Zeichenfolge, die an einen gegebenen Klartext vor der Verwendung als Eingabe einer Hashfunktion angehängt wird
Unumkehrbarkeit	Situation, in der es für einen beliebigen Übergang von identifizierbar zu pseudonym rechentechnisch unmöglich ist, vom Pseudonym auf den ursprünglichen Bezeichner zu schließen (Quelle: DIN EN ISO 25237)

Verantwortlicher	„’Verantwortlicher’ die natürliche oder juristische Person, Behörde, Einrichtung oder andere Stelle, die allein oder gemeinsam mit anderen über die Zwecke und Mittel der Verarbeitung von personenbezogenen Daten entscheidet; sind die Zwecke und Mittel dieser Verarbeitung durch das Unionsrecht oder das Recht der Mitgliedstaaten vorgegeben, so kann der Verantwortliche beziehungsweise können die bestimmten Kriterien seiner Benennung nach dem Unionsrecht oder dem Recht der Mitgliedstaaten vorgesehen werden“ (Quelle: Art. 4 Ziff. 7 DS-GVO)
Verarbeitung	„’Verarbeitung’ jeden mit oder ohne Hilfe automatisierter Verfahren ausgeführten Vorgang oder jede solche Vorgangsreihe im Zusammenhang mit personenbezogenen Daten wie das Erheben, das Erfassen, die Organisation, das Ordnen, die Speicherung, die Anpassung oder Veränderung, das Auslesen, das Abfragen, die Verwendung, die Offenlegung durch Übermittlung, Verbreitung oder eine andere Form der Bereitstellung, den Abgleich oder die Verknüpfung, die Einschränkung, das Löschen oder die Vernichtung“ (Quelle: Art. 4 Ziff. 2 DS-GVO)
Verknüpfung von Informationsobjekten	Prozess, der die Aufstellung einer logischen Verbindung zwischen verschiedenen Informationsobjekten ermöglicht (Quelle: DIN EN ISO 25237)

12 Literatur

12.1 Bücher

- Berberich O. Trusted Web 4.0 – Konzepte einer digitalen Gesellschaft: Konzepte der Dezentralisierung und Anonymisierung. Springer Vieweg, 1. Auflage 2016. ISBN 978-3-662-49189-8
- Khaled El Emam / Luk Arbuckle. Anonymizing Health Data. O'Reilly Media, Inc., 1. Auflage 2014. ISBN 9781449363079

12.2 Online

- DICOM Standard – Supplement 142 Clinical Trial De-identification Profiles (Stand 2009)). [Online, zitiert am 2018-05-16]; Verfügbar unter ftp://medical.nema.org/medical/dicom/final/sup142_ft.pdf
- Garfinkel S. (2015) De-Identification of Personal Information. [Online, zitiert am 2018-04-24]; Verfügbar unter https://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf
- Hochfellner et al. (2012) FDZ-Methodenreporte: Datenschutz am Forschungsdatenzentrum. (Hrsg.: Bundesagentur für Arbeit) [Online, zitiert am 2018-04-24]; Verfügbar unter http://doku.iab.de/fdz/reporte/ 2012/MR_06-12.pdf
- Integrating the Healthcare Enterprise (IHE), Domain „IT Infrastructure" (ITI): nalysis of Optimal De-Identification Algorithms for Family Planning Data Elements (Stand 2016-12-02). [Online, zitiert am 2018-05-16]; Verfügbar unter http://www.ihe.net/uploadedFiles/Documents/ITI/IHE_ITI_WP_Analysis-of-DeID-Algorithms-for-FP-Data_Elements.pdf
- Integrating the Healthcare Enterprise (IHE), Domain „IT Infrastructure" (ITI): Algorithm Mapping Spreadsheet (for use with De-Identification Handbook) (Stand 2014-06-06). [Online, zitiert am 2018-05-16]; Verfügbar unter http://www.ihe.net/uploadedFiles/Documents/ITI/IHE_ITI_Handbook_De-Identification-Mapping_Rev1.1_2014-06-06.xlsx
- Integrating the Healthcare Enterprise (IHE), Domain „Quality, Research and Public Health" (QRPH): Pseudonymization White Paper (Stand: 2008-12-22). [Online, zitiert am 2018-05-16]; Verfügbar unter ftp://ftp.ihe.net/Quality/2009_2010_YR_3/Planning/White%20papers%20yr%203/Pseudonymisation-WP.doc
- Oracle.(2013) Data Masking Best Practice. [Online, zitiert am 2018-04-24]; Verfügbar unter http://www.oracle.com/us/products/database/data-masking-best-practices-161213.pdf
- Statistisches Bundesamt (Destatis)
    - Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten - Band 4 der Reihe Statistik und Wissenschaft. [Online, zitiert am 2018-04-24]; Verfügbar unter https://www.destatis.de/DE/Publikationen/StatistikWissenschaft/Band4_AnonymisierungMikrodaten.html
    - Verfahren zur Anonymisierung von Einzeldaten - Band 16 der Reihe Statistik und Wissenschaft. [Online, zitiert am 2018-04-24]; Verfügbar unter https://www.destatis.de/DE/Publikationen/StatistikWissenschaft/Band16_AnonymisierungEinzeldaten.html
    - Methoden der Geheimhaltung wirtschaftsstatistischer Einzeldaten und ihre Schutzwirkung - Band 18 der Reihe Statistik und Wissenschaft. [Online, zitiert am 2018-04-24]; Verfügbar unter

https://www.destatis.de/DE/Publikationen/StatistikWissenschaft/Band18_MethodenGeheimhaltung.html

- TMF - Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.
  - Anonymisierung. [Online, zitiert am 2018-04-24]; Verfügbar unter https://www.toolpool-gesundheitsforschung.de/produkte/?term=anonymisierung
  - Pseudonymisierung. [Online, zitiert am 2018-04-24]; Verfügbar unter https://www.toolpool-gesundheitsforschung.de/produkte/?term=pseudonymisierung
- U.S. Department of Health & Human Services: Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. [Online, zitiert am 2018-04-24]; Verfügbar unter https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html

12.3 Zeitschriften

- Akgün et al. (2015) Privacy preserving processing of genomic data: A survey. Journal of Biomedical Informatics, http://dx.doi.org/10.1016/j.jbi.2015.05.022
- Brisch K, Pieper F. (2015) Das Kriterium der "Bestimmbarkeit" bei Big Data-Analyseverfahren - Anonymisierung, Vernunft und rechtliche Absicherung bei Datenübermittlungen. CR: 724-729
- El Emam, et al. (2011) A Systematic Review of Re-Identification Attacks on Health Data. PLoS One 6(12): e28071, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3229505/
- Gal et al. (2014) A data recipient centered de-identification method to retain statistical Attributes. J Biomed Inform, http://dx.doi.org/10.1016/j.jbi.2014.01.001
- Geschonneck A, Meyer J, Scheben B. (2011) Anonymisierung im Rahmen der forensischen Datenanalyse. BB:2677-2680
- Hammer V, Knopp M. (2015) Datenschutzinstrumente Anonymisierung, Pseudonyme und Verschlüsselung. DuD: 503509
- Härting N. (2013) Anonymität und Pseudonymität im Datenschutzrecht. NJW: 2065-2071
- He et al. (2015) CRFs based de-identification of medical records. J Biomed Inform: S39-S46, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4988860/pdf/nihms807198.pdf
- Höhne H. (2008) Anonymisierungsverfahren für Paneldaten. Wirt Sozialstat Archiv: 259–275
- Jiang et al. (2017) De-identification of medical records using conditional random fields and long short-term memory networks. J Biomed Inform: S43-S53, https://www.sciencedirect.com/science/article/pii/S1532046417302228
- Karg M. (2015) Anonymität, Pseudonyme und Personenbezug revisited. DuD: 520-526
- Kim S, Lee H, Chung YD (2016) Privacy-preserving Data Cube for Electronic Medical Records: An Experimental Evaluation. International Journal of Medical Informatics, http://dx.doi.org/10.1016/j.ijmedinf.2016.09.008
- Knopp M. (2015) Pseudonym – Grauzone zwischen Anonymisierung und Personenbezug. DuD: 527-530
- Kushida, et al. (2012) trategies for De-identification and Anonymization of Electronic Health Record Data for Use in Multicenter Research Studies. Med Care: S82–S101
- Loukides et al. (2014) Disassociation for electronic health record privacy. Journal of Biomedical Informatics: 46-61
- Lu Y, Sinnott RO, Verspoor K. (2017) A Semantic- ased K- nonymity Scheme for Health Record Linkage. Integrating and Connecting Care: 84-90

- Malin B, Sweeney L. (2004) How (not) to protect genomic data privacy in a distributed network: using trail re-identification to evaluate and design anonymity protection systems. Journal of Biomedical Informatics: 179-192
- Malin B. (2010) Secure construction of k-unlinkable patient records from distributed providers. Artificial Intelligence in Medicine: 29–41
- Malin et al. (2011) Identifiability in biobanks: models, measures, and mitigation strategies. Hum Genet: 383–392
- Marnau N. (2016) Anonymisierung, Pseudonymisierung und Transparenz für Big Data - Technische Herausforderungen und Regelungen in der Datenschutz-Grundverordnung. DuD: 428-433
- Neamatullah et al. (2008) Automated de-identification of free-text medical records. BMC Med Inform Decis Mak, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2526997/pdf/1472-6947-8-32.pdf
- Neubauer et al. (2010) Pseudonymisierung für die datenschutzkonforme Speicherung medizinischer Daten. Elektrotechnik & Informationstechnik: 135–142
- Poulis et al. (2017) Anonymizing datasets with demographics and diagnosis codes in the presence of utility constraints. Journal of Biomedical Informatics: 76-96
- Raisaro et al. (2017) Addressing Beacon re-identification attacks: quantification and mitigation of privacy risks. JAMIA: 1-8, doi: 10.1093/jamia/ocw167
- Rosemann M, Vorgrimler D, Lenz R. (2004) Erste Ergebnisse faktischer Anonymisierung wirtschaftsstatistischer Einzeldaten. Allgemeines Statistisches Archiv: 73–99
- RoßnagelA. (2018) Pseudonymisierung personenbezogener Daten. ZD: 243-247
- Shringarpure S, Bustamante C. (2015) Privacy Risks from Genomic Data-Sharing Beacons. The American Journal of Human Genetics: 631–646
- Wallace SE. (2016) What Does Anonymization Mean? DataSHIELD and the Need for Consensus on Anonymization Terminology. Biopreservation and Biobanking: 224-230
- Wójtowicz M, Cebulla M. (2017) Anonymisierung nach der DSGVO. PinG: 186-192