Datenschutz - Pseudonymisierung, Anonymisierung

Was ist schon anonym?

DI Dr. Michael Prinz
Med. Universität Wien
ITSC, IT4Science

Inhalt

- Basis DSGVO
- Begriffsabklärung Anonymisierung/Pseudonymisierung
- Beispiel Anonymisierung eines Datensatzes
- Stufen der Anonymisierung
- Beispiel Pseudonymisierung eines Datensatzes
- Vergleich Anonyme Daten – Pseudonyme Daten
- Folgerungen und Gefahren
- Anwendungsgebiete

Zitate

Tobias Hann (Geschäftsführer Mostly):

Beim Anonymisieren von Personendaten sinke einerseits die Datenqualität, andererseits seien sie mit wenig Aufwand immer noch identifizierbar.

Ein Datenschatz für die Finanzbranche, Der Standard, 12.5.2022

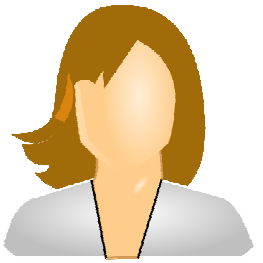
NN (aus Antrag an Ethikkommission):

Die Daten werden pseudoanonymisiert.

Datenschutzgrundverordnung (DSGVO)

- regelt die Verarbeitung **personenbezogener Daten**
- seit 2018 in Kraft
- gilt für den gesamten EU-Raum
- wurde auch in den Ländern des EWR übernommen
- hohe Bußgelder möglich
- wird ergänzt durch nationale zus. Datenschutzregelungen (z.B. FOG in Österreich)

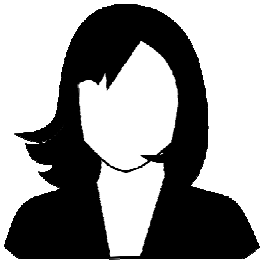
Datenschutzgrundverordnung (DSGVO)



personenbezogene Daten (Art. 4 DSGVO):

… „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, … identifiziert werden kann.

Datenschutzgrundverordnung (DSGVO)



Pseudonymisierung (Art. 4 DSGVO):

… „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;

Datenschutzgrundverordnung (DSGVO)

anonyme Daten (Art. 29 Arbeitsgruppe):

… Demnach müssen im Zuge der Anonymisierung von Daten hinreichend viele Elemente entfernt werden, sodass eine Identifizierung der betroffenen Person **ausgeschlossen** ist. …

Ein wichtiger Faktor ist, dass die Verarbeitung **unumkehrbar** sein muss.

Anonymisierung Schritt 1

Nach-namen	Vor-namen	Geb.dat.	Sex	Pat-ID	Anamnese	Diagnose	Lab 1	Lab 2
Müller	Fritz	24.3.1950	m	321411	Bauchschmerzen	Tumor	1.5	2.5
Huber	Hans	17.11.1973	m	765421	Kopfschmerzen	Hoher Blutdruck	2.5	2.5

ID	Geb.dat.	Sex	Pat-ID	Anamnese	Diagnose	Lab 1	Lab 2
A56Gv3	24.3.1950	m	321411	Bauchschmerzen	Tumor	1.5	2.5
bG55HI	17.11.1973	m	765421	Kopfschmerzen	Hoher Blutdruck	2.5	2.5

ID	Geb.dat.	Sex	Anamnese	Diagnose	Lab 1	Lab 2
A56Gv3	24.3.1950	m	Bauchschmerzen	Tumor	1.5	2.5
bG55HI	17.11.1973	m	Kopfschmerzen	Hoher Blutdruck	2.5	2.5

Anonymisierung Schritt 2

Nach-namen	Vor-namen	Geb.dat.	Sex	Pat-ID	Anamnese	Diagnose	Lab 1	Lab 2
Müller	Fritz	24.3.1950	m	321411	Bauchschmerzen	Tumor	1.5	2.5
Huber	Hans	17.11.1973	m	765421	Kopfschmerzen	Hoher Blutdruck	2.5	2.5

ID	Geb.jahr	Sex	Anamnese	Diagnose	Lab 1	Lab 2
A56Gv3	1950	m	Bauchschmerzen	Tumor	1.5	2.5
bG55HI	1973	m	Kopfschmerzen	Hoher Blutdruck	2.5	2.5

ID	Alter	Sex	Anamnese	Diagnose	Lab 1	Lab 2
A56Gv3	72	m	Bauchschmerzen	Tumor	1.5	2.5
bG55HI	49	m	Kopfschmerzen	Hoher Blutdruck	2.5	2.5

Anonymisierung Schritt 3

Nach-namen	Vor-namen	Geb.dat.	Sex	Pat-ID	Anamnese	Diagnose	Lab 1	Lab 2
Müller	Fritz	24.3.1950	m	321411	Bauchschmerzen	Tumor	1.5	2.5
Huber	Hans	17.11.1973	m	765421	Kopfschmerzen	Hoher Blutdruck	2.5	2.5

ID	Alter	Sex	Anamnese		Lab 1	Lab 2
A56Gv3	40-80	m	Schmerzen		1-3	2.5
bG55HI	40-80	m	Schmerzen		1-3	2.5

usw.

Anonymisierungsstufen

	Absolute Anonymisierung	Faktische Anonymisierung	Formale Anonymisierung
Reidentifikation	**ausgeschlossen**	**unwahrscheinlich** mit aktuellem Stand der Technik und verhältnismäßigem Aufwand	**wahrscheinlich** nur direkt identifizierende Merkmale werden entfernt
Sicherheit	**sicher**	**Risikoabschätzung**	**unsicher** nicht DSGVO konform

Faktische Anonymisierung

Einschätzen des Reidentifikationsrisikos nach Five Safes Modell (nach Desai*)

1. **Sichere Projekte**: Ist die Verwendung der Daten angemessen?
2. **Sichere Personen**: Kann den Forschern vertraut werden, dass sie die Daten angemessen verwenden?
3. **Sichere Daten**: Gibt es ein Identifikationsrisiko in den Daten selbst?
4. **Sichere Umgebung**: Limitiert die Bereitstellungsinfrastruktur unautorisierten Zugang zu den Daten?
5. **Sichere Ergebnisse**: Führen die Ergebnisse, die aus den Daten gewonnen werden, zu einem Identifikationsrisiko?

* *Tanvi Desai, Felix Ritchie and Richard Welpton, 'Five Safes: Designing data access for research' (2016) http://csrm.cass.anu.edu.au/sites/default/files/rsss/Ritchie_5safes.pdf*

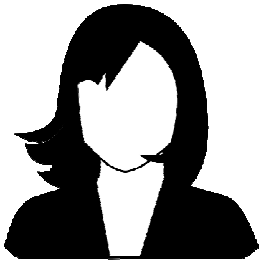
Pseudonymisierung Schritt 1

Nach-namen	Vor-namen	Geb.dat.	Sex	Pat-ID	Anamnese	Diagnose	Lab 1	Lab 2
Müller	Fritz	24.3.1950	m	321411	Bauchschmerzen	Tumor	1.5	2.5
Huber	Hans	17.11.1973	m	765421	Kopfschmerzen	Hoher Blutdruck	2.5	2.5

ID	Geb.dat.	Sex	Pat-ID	Anamnese	Diagnose	Lab 1	Lab 2
001-321	24.3.1950	m	321411	Bauchschmerzen	Tumor	1.5	2.5
001-453	17.11.1973	m	765421	Kopfschmerzen	Hoher Blutdruck	2.5	2.5

ID	Geb.dat.	Sex	Anamnese	Diagnose	Lab 1	Lab 2
001-321	24.3.1950	m	Bauchschmerzen	Tumor	1.5	2.5
001-453	17.11.1973	m	Kopfschmerzen	Hoher Blutdruck	2.5	2.5

Pseudonymisierung Schritt 2

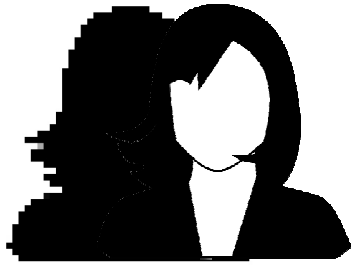


Nach-namen	Vor-namen	Geb.dat.	Sex	Pat-ID	Anamnese	Diagnose	Lab 1	Lab 2
Müller	Fritz	24.3.1950	m	321411	Bauchschmerzen	Tumor	1.5	2.5
Huber	Hans	17.11.1973	m	765421	Kopfschmerzen	Hoher Blutdruck	2.5	2.5

ID	Geb.jahr	Sex	Anamnese	Diagnose	Lab 1	Lab 2
001-321	1950	m	Bauchschmerzen	Tumor	1.5	2.5
001-453	1973	m	Kopfschmerzen	Hoher Blutdruck	2.5	2.5

ID	Alter	Sex	Anamnese	Diagnose	Lab 1	Lab 2
001-321	72	m	Bauchschmerzen	Tumor	1.5	2.5
001-453	49	m	Kopfschmerzen	Hoher Blutdruck	2.5	2.5

Vergleich Anonymisierung/Pseudonymisierung



Nach-namen	Vor-namen	Geb.dat.	Sex	Pat-ID	Anamnese	Diagnose	Lab 1	Lab 2
Müller	Fritz	24.3.1950	m	321411	Bauchschmerzen	Tumor	1.5	2.5
Huber	Hans	17.11.1973	m	765421	Kopfschmerzen	Hoher Blutdruck	2.5	2.5

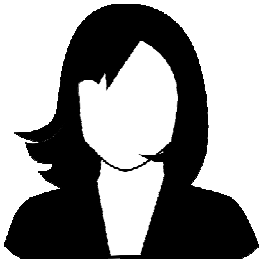
anonym

ID	Alter	Sex	Anamnese		Lab 1	Lab 2
A56Gv3	40-80	m	Schmerzen		1-3	2.5
bG55HI	40-80	m	Schmerzen		1-3	2.5

pseudonym

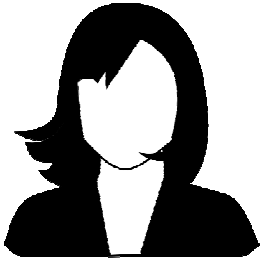
ID	Alter	Sex	Anamnese	Diagnose	Lab 1	Lab 2
001-321	72	m	Bauchschmerzen	Tumor	1.5	2.5
001-453	49	m	Kopfschmerzen	Hoher Blutdruck	2.5	2.5

wesentliche Unterschiede



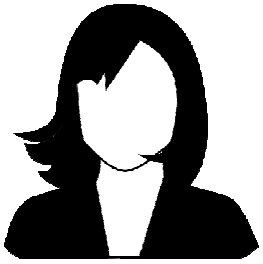
	Anonyme Daten	Pseudonyme Daten
Aufwand	hoch	niedrig
	muss an Dateninhalten und Datenumfang angepasst werden	keine Anpassung notwendig
Datenqualität	sinkt	bleibt gleich

wesentliche Unterschiede



	Anonyme Daten	Pseudonyme Daten
DSGVO anzuwenden	X	✓
Einwilligung notwendig	X	✓ Wenn nicht vorhanden, andere ges. Abdeckung notwendig
Weitergabe außerhalb EU/EWR	✓	Abkommen zwischen EU und Zielland oder Einzelabkommen zwischen Forschungspartnern muss vorhanden sein
Reidentifikation	X	✓

Gefahren



	Anonyme Daten	Pseudonyme Daten
Missbräuchliche Reidentifikation	ausgeschlossen bis gering	besteht
Empfänger	oft unbekannt	bekannt
Kontrolle über Daten	meist keine	Vereinbarungen mit Empfänger
	Anonymität kann nicht für immer gewährleistet werden	
	Keine Judikatur, welche Anonymisierung ausreichend ist	

Anwendungsgebiete Anonyme Daten

- Lehrfallsammlungen
- Datensammlungen für Forschung
- Sammlung von AI-Trainingsdaten
- Testdaten für Entwicklung von Software
- Erstellung von personenunabhängigen Statistiken

➢ Personen sind unerheblich

Anwendungsgebiete Pseudonyme Daten

- Klinische Studien
- Prospektive Forschungsprojekte
- personenbezogene Statistiken/Auswertungen

➢ Aggregation von Daten notwendig

➢ Reidentifikation muss möglich sein

Zitate Auflösung

Tobias Hann (Geschäftsführer Mostly):

Beim Anonymisieren von Personendaten sinke einerseits die Datenqualität, andererseits seien sie mit wenig Aufwand immer noch identifizierbar.

Ein Datenschatz für die Finanzbranche, Der Standard, 12.5.2022

Datenqualität sinkt ✓

Personen in anonymen Daten sind per Definition nicht mit wenig Aufwand identifizierbar. ✗

Zitate Auflösung

NN (aus Antrag an Ethikkommission):

Die Daten werden pseudoanonymisiert.

Daten sind entweder anonym oder pseudonym. Der Begriff pseudoanonym existiert nicht. ✕

Kontakt

Vielen Dank für Ihre Aufmerksamkeit!

https://www.meduniwien.ac.at/daten-clearingstelle

Michael.Prinz@meduniwien.ac.at