Karlsruher Institut für Technologie

ZAR
Zentrum für Angewandte Rechtswissenschaft

Personenbezogene Daten, Pseudonymisierung & Anonymisierung

Dr. Christoph Werner

Vorlesung Datenschutzrecht
WS 2024/2025, UE 5/15

28.11.2024

wooclap.com Code: CEFDGQ

Agenda

Wiederholung

1. Übersicht zu den Datenkategorien
2. Pseudonymisierte Daten
3. Anonyme/anonymisierte Daten
4. Beispielsfälle
5. Synthetische Daten

Wiederholung:
Wesentliche Lerninhalte der letzten UE

- Unterschiedliche staatliche Datenerhebungsmaßnahmen (öffentlicher Datenschutz) betreffen unterschiedliche Grundrechte (RiSb, Fernmeldegeheimnis, Computergrundrecht)
- Eröffnung sachlicher Anwendungsbereich der DSGVO (Art. 2 DSGVO): Verarbeitung personenbezogener Daten. Wichtigste Ausnahmen: „Haushaltsausnahme“, Gefahrenabwehr & Polizei, elektronische Kommunikation
- Räumlicher Anwendungsbereich (Niederlassung, Marktortprinzip oder Verhaltensbeobachtung in EU)
- Datenübermittlung ins Ausland **entweder auf Basis eines Angemessenheitsbeschlusses oder geeignete Garantien (Art. 45, 46 DSGVO)**
- Mit den USA besondere Herausforderungen im Rahmen des Angemessenheitsbeschlusses (aktuell EU-US Data Privacy Framework)

Wiederholung

Angemessenheitsbeschluss vs. geeignete Garantien (Art. 45, 46 DSGVO)

EU

DSGVO

USA (Empfängerland)

modifiziertes US-Datenschutzrecht, welches insb. Sicherheitsbehörden reguliert

DSGVO

Verantwortliche

Datenempfänger in den USA

Wiederholung

Angemessenheitsbeschluss vs. geeignete Garantien (Art. 45, 46 DSGVO)

EU

DSGVO

~~Angemessenheitsbeschluss~~

USA (Empfängerland)

~~modifiziertes US-Datenschutzrecht, welches insb. Sicherheitsbehörden reguliert~~

DSGVO

Verantwortliche

Standardvertragsklauseln (SCCs)

Datenempfänger in den USA

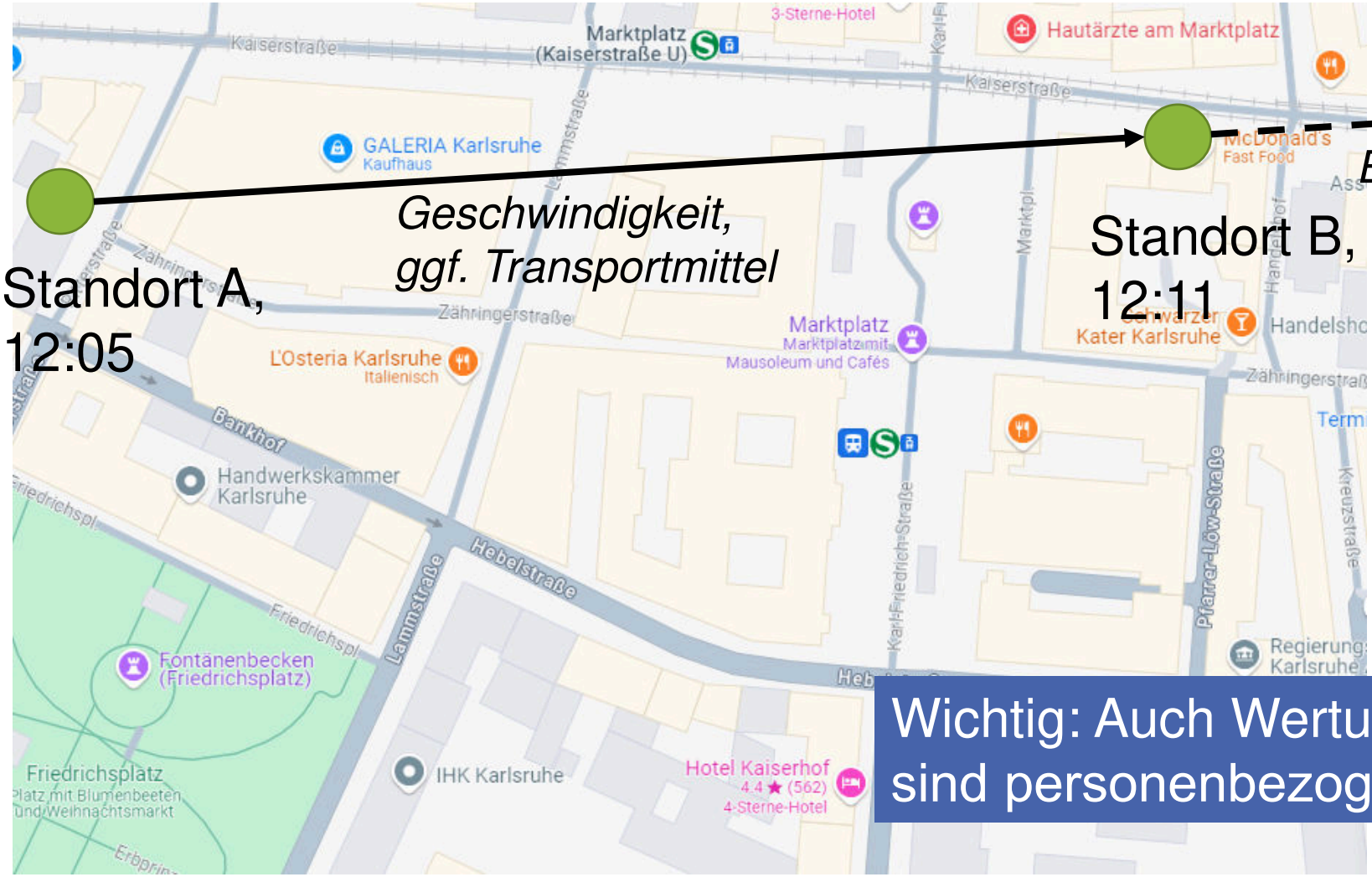
1. Übersicht

Sind „Daten“ dasselbe wie „Informationen“?

Art. 4 Nr. 1 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, 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;

Exkurs: Daten/Informationen



Wichtig: Auch Wertungen sind personenbezogene „Daten“

1. Übersicht

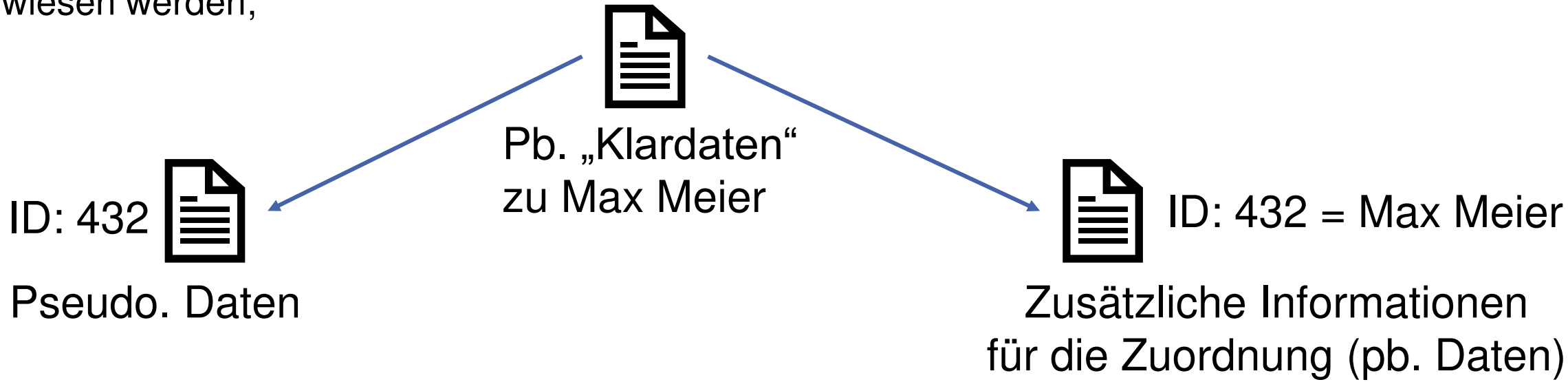
‚Personenbezogene Daten' sind **alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person** ('betroffene Person') beziehen (Art. 4 Nr. 1 DSGVO)

Bezogen auf eine identifizierte natürliche Person	**Bezogen auf eine identifizierbare natürliche Person**	**Besondere Kategorien pb. Daten („sensible Daten")**	**Kein Personenbezug**
• Name • Geburtsdatum • Adresse • E-Mail-Adresse • Telefonnummer	Ermittlung der Identität einer Person durch Verwendung von alternativer Informationen z.B. • IP-Adresse • Kfz-Kennzeichen und FIN • Personalnummer • Kontonummer • besonderen Merkmalen, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität • **pseudonymisierte Daten** und nicht ausreichend anonymisierte Daten	Daten bezogen auf • Rasse oder ethnische Herkunft • Politische Gesinnung • Religion oder philosophische Überzeugung • Gewerkschaftszugehörigkeit • Genetische und biometrische Daten • Gesundheit • Sexuelle Orientierung besonders geschützt durch Art. 9 DSGVO	Anonyme Informationen • Daten ohne Bezug auf eine identifizierte oder identifizierbare natürliche Person • Anonym erhobene Daten • Anonymisierte Daten • Voraussetzung: mit vernünftigen Mitteln keine Re-Identifizierung möglich (schwierig in Zeiten von Internet, Big Data & KI zum Mustervergleich) Maschinen- und Sachdaten

2. Pseudonymisierte Daten (Pseudo. Daten)

2. Pseudo. Daten

Definition in Art. 4 Nr. 5 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 [*z.B. eine Ordnungsziffer oder ein sonstiges Kennzeichen*] gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden;



2. Pseudo. Daten

Rechtsfolgen

Erwägungsgrund 27 S. 2 DSGVO:
Einer Pseudonymisierung unterzogene personenbezogene Daten, die durch Heranziehung zusätzlicher Informationen einer natürlichen Person zugeordnet werden könnten, sollten als Informationen über eine identifizierbare natürliche Person betrachtet werden

→ *fallen somit in den Anwendungsbereich der DSGVO*

Erwägungsgrund 28 DSGVO:
Die Anwendung der Pseudonymisierung auf personenbezogene Daten kann die Risiken für die betroffenen Personen senken und die Verantwortlichen und die Auftragsverarbeiter bei der Einhaltung ihrer Datenschutzpflichten unterstützen. Durch die ausdrückliche Einführung der „Pseudonymisierung“ in dieser Verordnung ist nicht beabsichtigt, andere Datenschutzmaßnahmen auszuschließen.

2. Pseudo. Daten

Rechtsfolgen

- Ermöglicht ggf. als ein Faktor die Datenverarbeitung nach Art. 6 Abs. 1 lit f) DSGVO [1]
- Kann (als ein Faktor) zweckändernde Weiterverarbeitung ermöglichen Art. 6 Abs. 4 lit e) DSGVO
- Ist eine Maßnahme zur Umsetzung der Datenschutzgrundsätze nach Art. 25 Abs. 1 DSGVO, insbesondere Datenminimierung - Art. 5 Abs. 1 lit c) DSGVO
- Ist auch eine Maßnahme der Datensicherheit (Art. 32 Abs. 1 lit a) DSGVO, weitere Datensicherheitsanforderungen werden dadurch aber nicht ausgeschlossen.

[1] Kühling/Buchner, 4. Aufl. 2024, DS-GVO Art. 6 Rn. 154

3. Anonyme/Anonymisierte Daten (Anon. Daten)

3. Anon. Daten

- EG 26, S. 1: Die Grundsätze des Datenschutzes sollten für alle Informationen gelten, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen. [Umkehrschluss: **nicht auf anon. Daten]**

- S. 3: Um festzustellen, ob eine natürliche Person identifizierbar ist, sollten alle Mittel berücksichtigt werden, die **von dem Verantwortlichen oder einer anderen Person** nach allgemeinem Ermessen wahrscheinlich genutzt werden, um die natürliche Person direkt oder indirekt zu identifizieren, wie beispielsweise das Aussondern.

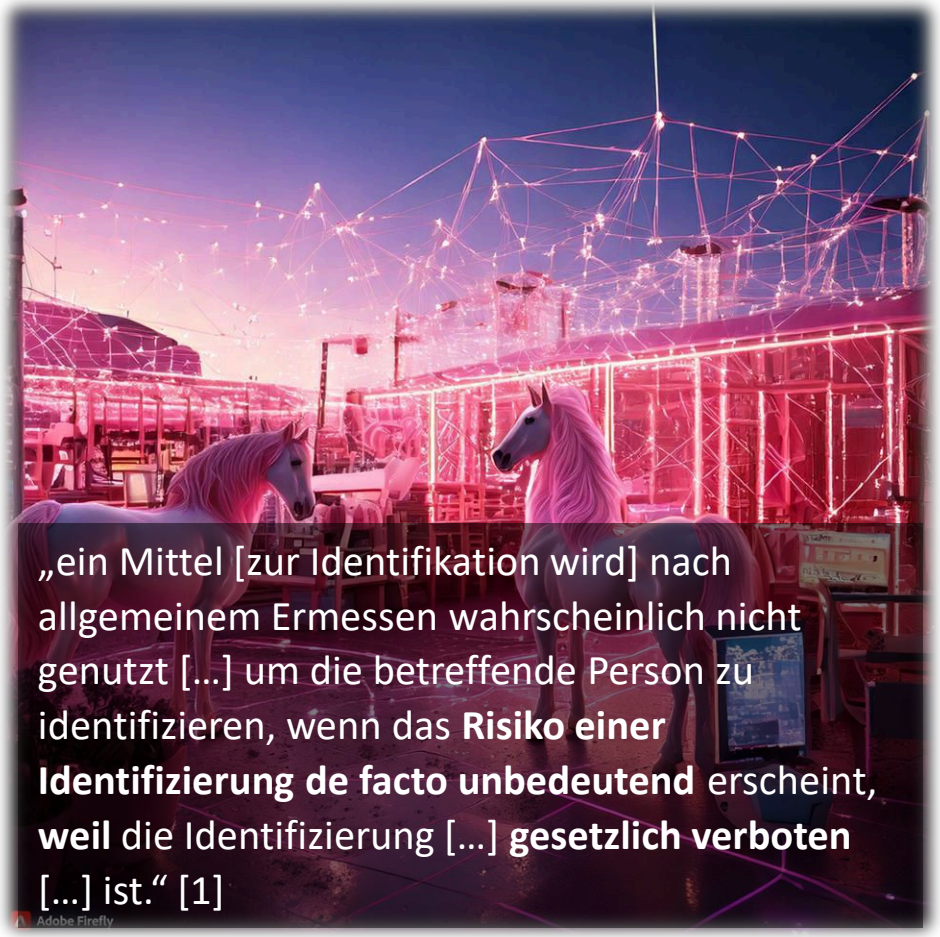
- S. 4: Bei der Feststellung, ob Mittel nach allgemeinem Ermessen wahrscheinlich zur Identifizierung der natürlichen Person genutzt werden, sollten alle objektiven Faktoren, wie die Kosten der Identifizierung und der dafür erforderliche Zeitaufwand, herangezogen werden, wobei die zum Zeitpunkt der Verarbeitung verfügbare Technologie und technologische Entwicklungen zu berücksichtigen sind

3. Anon. Daten

Was folgt aus diesen Vorgaben
(Art. 4 Nr. 1, EG 26 DSGVO)?

a) Subjektiver Ansatz (relativ)
Entscheidend ist**, ob der jeweilige „Verantwortliche“ bzw. der Empfänger** einen Personenbezug **mit legalen Mitteln** und vernünftigem Aufwand herstellen kann

- Deshalb auch pseudonymisierte Daten bei Empfänger ohne Zuordnungstabelle „anonym“
- So EuGH [1]
- **Kritik:**
  - Bzgl. Legalitätsannahme realitätsfremd
  - Datensicherheit ausgeblendet
  - Wortlaut EG 26: Verantwortliche oder „eine andere Person“



[1] EuGH, Urteil vom 4.9.2025 – C-413/23 P, Rn. 82 mit Verweis auf ältere Rspr.

3. Anon. Daten

Was folgt aus diesen Vorgaben
(Art. 4 Nr. 1, EG 26 DSGVO)?

b) Objektiver Ansatz (absolut)
Es ist bereits hinreichend, wenn **irgendeine Person** **(auch böswillige Dritte)** den Personenbezug mit vernünftigem Aufwand (und auch illegalen Mitteln) wiederherstellen könnte bzw. würde

- Einschränkung nur durch objektive Faktoren wie Kosten und (Zeit-)Aufwand, Technologie, d.h. keine irrational Handelnden

- **Kritik:**
  - für die Anonymität verbleibt nur ein sehr kleiner Anwendungsbereich
  - die Anonymität ist sehr schwer zu bestimmen

3. Anon. Daten

Gegenüberstellung der Ansätze

	Objektiver Ansatz	Subjektiver Ansatz
Berücksichtigte Faktoren	Modalitäten der „Verarbeitung“ anonymer Daten, insbesondere *Dauer;* Aufhebung der Anonymität durch *Zusammenführung mit anderen öffentlich zugänglichen Daten*?	
Kosten-Nutzen-Analyse	Kosten und (Zeit-)Aufwand der Re-Identifikation im Verhältnis zum Informationsgewinn (Sensibilität der Daten)	
Betrachteter Personenkreis	**De-Anonymisierung durch jede Person**	**De-Anonymisierung (nur) durch durch Verarbeiter**
Legalität der Mittel	wahrscheinlich genutzte, illegale Mittel (Angriffe)	Nur legale Mittel

3. Anon. Daten

Meines Erachtens objektiver Ansatz **aus folgenden Gründen** vorzugswürdig:

- Annahme der Anonymität sobald Daten vom Verarbeiter nicht mit legalen Mitteln hergestellt werden greift zu kurz
- Beachtung von illegalen Vorfällen zwingend (Datensicherheit)
- Pseudonymisierung ist gerade keine Anonymisierung (so auch der EDSA, siehe EuGH-Urteil [1])
- Rechtssicherheit durch Einordung von Daten als personenbezogen, soweit nicht objektiv anonym
- Optimale Schutzwirkung durch Erfassung möglichst aller pb. Daten

Aber trotzdem Privilegierung:

- Datenempfänger kann Auskunftsanspruch oder Löschungsanspruch (nicht) alleine erfüllen.
- Weitergabe an Dritte Verarbeitung i.S.d. Art. 6 lit f) DSGVO (bereits erleichtert, wenn pseudonym)
- Berücksichtigung bei risikobasierten Vorschriften (Insbesondere Art. 25, 32 DSGVO)

[1] EuGH, Urteil vom 4.9.2025 – C-413/23 P, Rn. 63 ff.

3. Anon. Daten

Subj. Ansatz ermöglicht zwar an sich deutlich erleichterte Datenweitergabe, aber

Defizite des subjektiven Ansatzes:

- **Verstärkte Rechtsunsicherheit durch „Zombie-Daten“**
  - Stehen pb. Daten wieder auf, greift DSGVO vollumfänglich: Rechtsgrundlage, Betroffenenrechte, Dokumentationspflichten („Backup-Plan“ erforderlich)
  - In der Literatur [1] wird vertreten, bereits der Verarbeiter der „anonymen“ Daten müsse trotzdem die **Verarbeitung beschränken bzw. sichern,** um Entfall der Anonymität zu vermeiden („DSGVO light“)
  - Ähnlich auch bei [2] ggf. Regelungsbedarf „was Datensicherheitserfordernisse und Weitergabe-Beschränkungen bei pseudonymisierten, künftig aber quasi als „relativ anonym“ zu behandelnden Daten anbelangt.“
  - Umstritten ist außerdem, ob Auftragsverarbeitungsvereinbarung (AVV) erforderlich

[1] Roßnagel, DuD 2014, 513 (519) [2] *Wendehorst*, https://background.tagesspiegel.de/digitalisierung-und-ki/briefing/der-digitale-omnibus-versuch-einer-einordnung

3. Anon. Daten

Entwurf für reformierte DSGVO (Omnibus-Verfahren [1])

Art. 4 Nr. 1: „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; Informationen über eine natürliche Person **sind nicht zwingend personenbezogene Daten für jede andere Person oder Einrichtung, nur weil eine andere Einrichtung diese natürliche Person identifizieren kann. Informationen sind für eine bestimmte Einrichtung nicht personenbezogen [anonym], wenn diese Einrichtung die natürliche Person, auf die sich die Informationen beziehen, unter Berücksichtigung der von dieser Einrichtung vernünftigerweise zu erwartenden Mittel nicht identifizieren kann.** Solche Informationen werden für diese Einrichtung [auch] nicht personenbezogen, nur weil ein späterer Empfänger über Mittel verfügt, die vernünftigerweise zur Identifizierung der natürlichen Person, auf die sich die Informationen beziehen, eingesetzt werden können.

Gesetzesentwurf, Stand 19.11.2025, S. 54, Abrufbar unter: https://digital-strategy.ec.europa.eu/en/library/digital-omnibus-regulation-proposal

3. Anon. Daten

- **Ist das Anonymisieren selbst eine Verarbeitung pb. Daten? (umstritten [1])**
  - Ursprünglich liegen personenbezogene Daten vor; diese werden nun durch Entfernen der Identifikationsmerkmale und ggf. Verrauschen der Daten (Alter: 22 wird zu Alter: 20-29) anonymisiert.

    - **Pro:**
      - Solange die Anonymisierung noch nicht abgeschlossen ist, liegen personenbezogene Daten vor
      - Art. 4 Nr. 2 DSGVO nennt auch „Löschen“ als Verarbeitung

    - **Contra:**
      - Dann auch Erlaubnistatbestand (Art. 6 Abs. 1 DSGVO) erforderlich; Wertungswiderspruch, da personenbezogene Daten gerade anonymisiert/gelöscht werden sollen;

      - Sehr weiter Anwendungsbereich der DSGVO: gilt dann von Anfang an nur nicht für von vorneherein anonyme bzw. reine Sachdaten (z.B. Geo- oder Wetterdaten)

[1] Gola in Gola/Heckmann, 3. Aufl. 2022, DS-GVO Art. 4 Rn. 52

4. Beispielsfall

EuGH [1]: hier keine pb. Daten mehr

Verantwortlicher (Online-Shop)

„Auftragsverarbeiter“ Data Analyst

- Personenbezogene Daten (Einkaufshistorie) werden pseudonymisiert

Max Meier wird zu Person432

Einkaufshistorie zu Person432

- Pseudonyme Daten werden verarbeitet
- Ergebnis: Person3572 ist sehr wahrscheinlich für die neuen Produkte X, Y, Z zu begeistern.

Ergebnis

[1] EuGH, Urteil vom 4.9.2025 – C-413/23 P, Rn. 82

4. Beispielsfall 2

- „Zum Schutz dieser Daten vor Missbrauch werden die Daten in einem vom Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) verantworteten Forschungsdatenzentrum Gesundheit (FDZ) **nicht mit Klarnamen, sondern pseudonymisiert** bereitgestellt. Als Grunddaten werden dem FDZ zunächst sämtliche Abrechnungsdaten der Krankenkassen von ihren gesetzlich Versicherten gespeichert. Hinzukommen sollen die **Daten aus der elektronischen Patientenakte** (ePA), die ab Januar 2025 flächendeckend eingeführt wird, soweit die Patienten dem nicht widersprechen.“*

- Zu den Daten gehören u.a.: **Geburtsjahr**, **Geschlecht**, **Postleitzahl**, **Krankenkasse**, **behandelnder Arzt***

- Wie „pseudonym“ ist das wohl bzw. wie sicher ist diese Pseudonymisierung?

**Weichert*, https://www.heise.de/hintergrund/Gesundheitsdatenforschung-ja-aber-bitte-mit-Datenschutz-10053620.html

Exkurs: Datenqualität vs. Anonymität

Exkurs: Datenqualität vs. Anonymität

ID	Geschlecht	Geburtsdatum	PLZ	Beruf
1	männlich	22.09.1984	76131	Professor
2	weiblich	29.05.2003	76135	Anwältin
3	männlich	10.08.2004	76359	Friseur
4	männlich	06.09.1994	76275	Bauarbeiter
5	weiblich	27.12.1999	76131	Ärztin
6	divers	05.01.1997	76131	Management

Überwiegende Identifizierbarkeit (63%), US-Studie von 2006 [1]

[1] Golle, https://crypto.stanford.edu/~pgolle/papers/census.pdf

Exkurs: Datenqualität vs. Anonymität

ID	Geschlecht	Altersgruppe	PLZ	Beruf
1	männlich	39-49	76xxx	Professor
2	weiblich	19-29	76xxx	Anwältin
3	männlich	19-29	76xxx	Friseur
4	männlich	29-39	76xxx	Bauarbeiter
5	weiblich	19-29	76xxx	Ärztin
6	divers	19-29	76xxx	Management

Exkurs: Datenqualität vs. Anoymität

Gewisse, qualitative Datensätze ((Quasi-)Identifikatoren) sind damit (nach objektivem Ansatz) per se anonymititätsfeindlich, z.B.

- Biologische Identifikatoren (Fingerabdrücke, DNA, Retina-Abbildungen)
- Psychologische Identifikatoren (längerer Browser- oder Suchmaschinenverlauf)
- Geographische Identifikatoren (kontinuierliches Standort-Tracking)

Art. 4 Nr. 1 DSGVO:

„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

5. Synthetische Daten

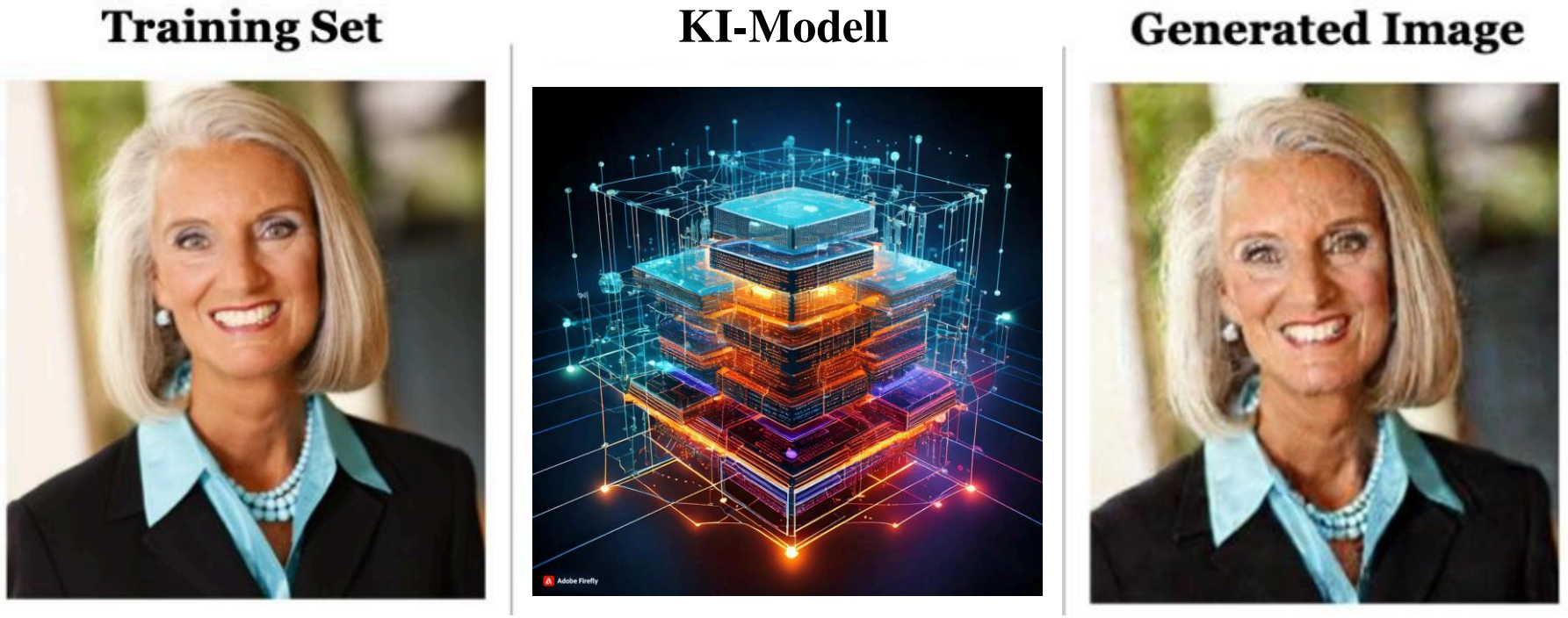
- Synthetische Daten sind realistische und möglichst allgemein verwendbare Daten; sie wurden zumeist mit KI aus realen, personenbezogenen Daten erzeugt und ähneln diesen in ihrer Struktur und Statistik möglichst stark [2, 3]

- Diese Daten können auch als anonym angesehen werden, soweit sie die dortigen Kriterien erfüllen:

  → Ist durch die KI-Erzeugung der synthetischen Daten hinreichend sichergestellt, dass keine personenbezogenen Daten aus dem ursprünglichen Trainingsdatensatz rekonstruiert werden können? [1]

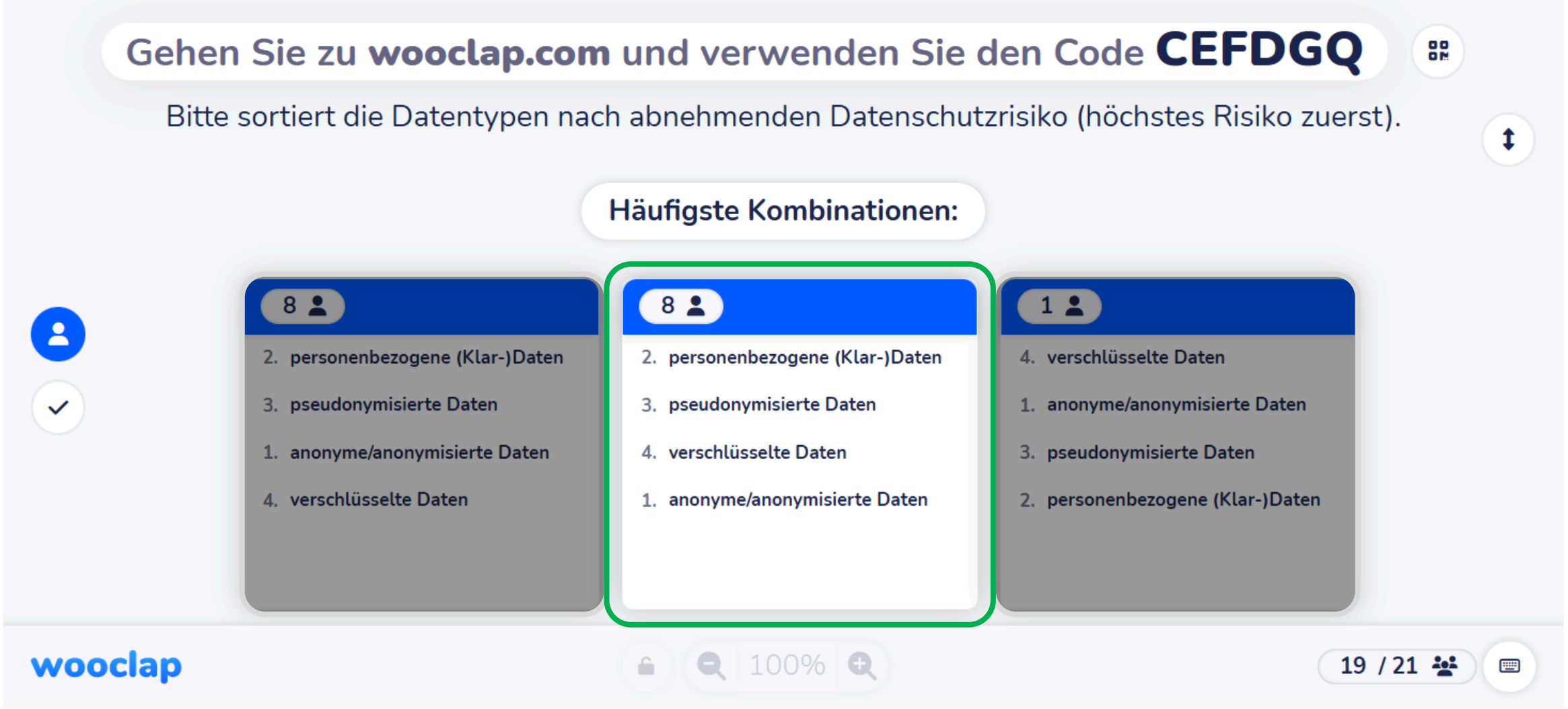
[1] Lettieri/Kipker, DuD 2024, 284 ff. [2] BeckOK DatenschutzR/Schild, 49. Ed. 1.8.2024, DS-GVO Art. 4 Rn. 27b [3] Lettieri/Kipker, DuD 2024, 284 ff.

Exkurs: Enthalten KI-Modelle pb. Daten?



Quelle: Vortragsfolien mit Fotos von Jonas Siglmüller, DGRI 2023

Risiko der unterschiedlichen Datentypen



Anhang: Wichtigste Lerninhalte

I. Personenbezogene Daten sind solche, die sich auf eine identifizierte oder identifizierbare Person beziehen

II. Vorliegen anonymer Daten sollte nach m.E. objektiv betrachtet werden, andere Entscheidung aber EuGH und Gesetzeslage (subjektive Betrachtung)

III. Streitpunkt insbesondere: Pseudonyme Daten sind und bleiben pseudonyme und damit pb. Daten (obj. Ansatz)

IV. Für synthetische Daten sind die Kriterien für anonyme Daten entsprechend anwendbar

V. In der Diskussion: KI-Modelle und personenbezogene Daten, siehe z.B. [1]

[1] Hüger, ZfDR 2024, 263