bvitg Bundesverband Gesundheits-IT

gmds Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V.

GDD Gesellschaft für Datenschutz und Datensicherheit e.V.

TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.

Workshop: Anonymisierung und Pseudonymisierung in Patientenversorgung und Forschung

Agenda

10:00 bis 11:00: Rechtliche Grundlagen

11:00 bis 12:30: Anforderungen der Datenverarbeiter

12:30 bis 13:30: Mittagspause

13:30 bis 15:00: Wann kann ich Daten als anonym ansehen?

15:50 bis 15:20: Kaffeepause

15:20 bis 17:30: Umsetzungskonzepte zur Anonymisierung

Rechtliche Grundlagen der Pseudonymisierung/Anonymisierung

Dr. Bernd Schütze

Berlin, 23. Mai 2016

HEALTHCARE SOLUTIONS

Was ist „pseudonym“, was „anonym“?

Begriffsbestimmung: Pseudonym, Anonym

§3 Abs. 6a: Pseudonymisieren ist das Ersetzen des Namens und anderer Identifikationsmerkmale durch ein Kennzeichen zu dem Zweck, die **Bestimmung des Betroffenen auszuschließen** oder **wesentlich zu erschweren**	Art. 4 Abs. 5: "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
§3 Abs. 6: „Anonymisieren ist das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse **nicht mehr** oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder **bestimmbaren natürlichen Person zugeordnet werden können**	- Keine Regelungen in der Verordnung, nicht einmal die Begriffsbestimmung - Ausschliesslich kurze Berücksichtigung in Erwägungsgrund 26

Genauer: was ist „anonym“ gemäß DS-GVO?

Anonyme Daten = Keine Re-Identifikation möglich

- EU DS-GVO
  - ErwGr. 26: Die Grundsätze des Datenschutzes sollten daher nicht für anonyme Informationen gelten …Diese Verordnung betrifft somit nicht die Verarbeitung solcher anonymer Daten
  - Personenbezug Art. 4: …als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt …identifiziert werden kann
    - ErwGr. 26: „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
    - Indirekte Identifizierung = Pseudonym

Begrifflichkeit „anonym“: BDSG vs. DS-GVO

Relativer und absoluter Personenbezug

- BDSG
  - Anonymisieren ist Verändern, dass … Einzelangaben … nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand … einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden kann
  - In Deutschland etablierte sich daraus der „absolute“ (= nicht mehr) und der „relative“ (= unverhältnismäßig großer Aufwand) Personenbezug bzgl. Anonymität
- EU DS-GVO
  - Personenbezug Art. 4 i.V.m. ErwGr 26: ist eine natürliche Person direkt oder indirekt identifizierbar = personenbezogene Daten
  - Artikel-29-Datenschutzgruppe: keine relative Anonymität („keine Möglichkeit zur Re-Identifizierung“, WP 216)
  - ➢ Unter Berücksichtigung der Zusammensetzung des künftigen Datenschutz-Ausschusses und der Tatsache, dass die entsprechenden Regelungen der RL 95/46/EG denen der DS-GVO entsprechen…
  - ➢ DS-GVO spricht wohl von absoluter Anonymität

Was folgt daraus?

Folgerungen

- BDSG
  - Pseudonymisieren ist Ersetzen
  - Anonymisieren ist Verändern
  - Beides laut §3 Abs. 4 BDSG „Verarbeitung“ bzw. „Nutzung“
  - ➢ §4 Abs. 1 BDSG: Erhebung, Verarbeitung und Nutzung personenbezogener Daten sind **nur zulässig**, soweit **dieses Gesetz** oder eine **andere Rechtsvorschrift** dies **erlaubt** oder anordnet oder der **Betroffene eingewilligt** hat.
- EU DS-GVO
  - Alles, was mit personenbezogenen Daten geschieht, ist Verarbeitung
  - Insbesondere auch die Pseudonymisierung oder Anonymisierung
  - Artt. 6,9: Verarbeitung nur mit Erlaubnistatbestand zulässig
- Im Folgenden Betrachtung bzgl. Verarbeitung von Gesundheitsdaten, genetischen Daten

Erlaubnistatbestand gefordert

Rechtsgrundlage erforderlich

→ Eine Pseudonymisierung oder Anonymisierung benötigt einen Erlaubnistatbestand

- ➢ Gesetzlicher Erlaubnistatbestand
- ➢ Betroffener willigt ein

– Mögliche heutige gesetzliche Erlaubnistatbestände

- Forschung mit Gesundheitsdaten
  (z.B. §28 Abs. 6 Ziff. 4 BDSG, Regelungen der Landeskrankenhausgesetze)
- Qualitätssicherung mit Gesundheitsdaten
  (z.B. SGB, Vorgaben der Landeskrankenhausgesetze)

☞ Problem: in BDSG und Landesdatenschutzgesetzen häufig Hinweis „erlaubt nur, wenn gesetzliche Schweigepflicht nicht verletzt wird“

☞ Entbindung Schweigepflicht erforderlich

Erlaubnistatbestand gefordert, aber woher nehmen?

Rechtsgrundlage erforderlich: was gilt unter der DS-GVO ab 25. Mai 2018?

→ Eine Pseudonymisierung oder Anonymisierung benötigt einen Erlaubnistatbestand

- ➢ Gesetzlicher Erlaubnistatbestand
- ➢ Betroffener willigt ein

– Mögliche Erlaubnistatbestände unter der EU DS-GVO

- ???
- Nationale Regeln erforderlich

Abgrenzung: Pseudonym vs. Anonym

Wann sind Daten anonym, wann pseudonym?

- Wann sind Daten als „anonym“ anzusehen, wann als pseudonym?
  - Definition „Pseudonym“ der EU DS-GVO beinhaltet, was Stand heute in Deutschland als „faktisch anonym“ angesehen wird
  - Artikel-29-Datenschutzgruppe
    - „Ein häufiger Irrtum liegt in der Annahme, dass pseudonymisierte Daten mit anonymisierten Daten gleichzusetzen seien“
    - „Pseudonymisierung verringert die Verknüpfbarkeit eines Datenbestands mit der wahren Identität einer betroffenen Person und stellt somit eine sinnvolle Sicherheitsmaßnahme, aber kein Anonymisierungsverfahren dar“
    - Häufiger Fehler: „Annahme, dass ein pseudonymisierter Datenbestand anonymisiert ist…“
    - „…Ergebnis der Anonymisierung … so dauerhaft sein sollte wie eine Löschung … es darf nicht möglich sein, die personenbezogenen Daten weiter zu verarbeiten“

Quelle: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_de.pdf

Abgrenzung: Pseudonym vs. Anonym

Re-Identifikationspotential medizinischer Daten

- Medizinische Daten können in sich das Potential der Re-Identifikationsmöglichkeit beinhalten
  - Genetische Daten
  - Bilddaten, die eine 3D-Rekonstruktion identifizierender Merkmale erlauben
  - Medizinische Daten selbst, abhängig von der Gruppengröße (z.B.: „Der“ männliche Patient mit Brustkrebs)
- Diese Daten sind im Sinne der DS-GVO als pseudonym anzusehen, eine Anonymisierung erscheint kaum möglich

→ Ausführliche Besprechung der daraus resultierenden Herausforderungen nach der Mittagspause

Offene Fragen

Fragen, auf die Antworten gefunden werden müssen

→ Welche Rechtsgrundlage benötigt Deutschland für Anonymisierung/Pseudonymisierung
  - Forschung
  - Qualitätssicherung
  - Routineversorgung (z.B. Wartung von Informationssystemen)
  - …
  - ➢ Nationaler Erlaubnistatbestand zur Zweckänderung eine Herangehensweise?

→ Wann sind Daten eines Patienten als anonym anzusehen, wann „nur" als pseudonym?
  - Merke: pseudonyme Daten = personenbezogene Daten (ErwGr. 26 DS-GVO)

→ Wie gehen wir mit nationalen Regelungen um, wenn wir innereuropäisch international arbeiten wollen?
  - Kann uns der Datenschutz-Ausschuss unterstützen?
    (Stichwort: Kohärenz-Verfahren und einheitliche Auslegung der DS-GVO)

UNIVERSITÄT LEIPZIG

Anonymisierung und Pseudonymisierung: Anforderungen der Datenverarbeiter (Entwickler)

Alfred Winter

imise.
Institut für Medizinische Informatik, Statistik und Epidemiologie

Unstrukturierte Daten aus Arztbriefen extrahieren

- Klinische Forschung benötigt viele phänotypisierte Patienten/Individuen
- Zu jedem Patienten gibt es viele phänotypische Informationen …

  … aber zu oft nur unstrukturiert auf Papier, in TIFF- oder PDF-Dokumenten.
- Natural Language Proceessing (NLP) kann Daten extrahieren …

  … aber die besten Verfahren funktionieren nur bei englischen Texten.

Unstrukturierte Daten aus Arztbriefen extrahieren

- *Projekt*: Entwicklung von NLP-Verfahren zur Extraktion der Medikation aus Arztbriefen.

  Testdaten: je 1.000 Arztbriefe aus drei Universitäts-Klinika.
- *Problem*: Kann man Arztbriefe anonymisieren (siehe heute Nachmittag!)?
- *Lösung*: Arztbriefe von Verstorbenen?

Warnsystem zur Beatmung auf der ITS entwickeln

- Viele Patienten auf der ITS könnten von einer lungen-protektiven Beatmung profiteren …

  … wenn die Beatmung rechtzeitig eingeleitet würde.
- Ein Warnsystem ist denkbar, das die großen Datenmengen des PDMS intelligent und zeitnah auswertet.

Warnsystem zur Beatmung auf der ITS entwickeln

- *Projekt*: Entwicklung eines Warnsystems für die Einleitung lungen-protektiver Beatmung.

  Testdaten: entsprechende Phänotypen aus drei Universitäts-Klinika.
- *Problem*: Wie erklärt ein Patient seine Einwilligung gegenüber einem Verbund aus drei Universitätsklinika?
- *Lösung*: Gründung eines Vereins?

Patientenidentifikation für Record-Linkage in Krebsregistern testen

- In Krebsregistern mit Pseudonymen müssen Datensätze zu demselben Patienten aus unterschiedlichen Quellen zusammengeführt werden.
- Welche Pseudonymisierungsverfahren liefern die wenigsten Fehler beim Record-Linkage?

Patientenidentifikation für Record-Linkage in Krebsregistern testen

- *Publikation in der MIBE*: Test eines Psydonymisierungsverfahrens gegen einen Testdatensatz als Goldstandard.
- *Gutachterfrage*: Können die Daten des Goldstandards publiziert werden?
- *Antwort der Autoren*: Nein, Datenschutz!

Patienten die Kontrolle über die Nutzung ihrer Daten in einem Forschungskonsortium einräumen

Access Control <-> Data Usage Control

- Patient kann jederzeit sehen, wer welche seiner Daten wofür nutzt (Transparenz).
- Patient kann jederzeit die Zustimmung zur Nutzung seiner Daten zurückziehen/geben (Dynamic Consent)

Patienten die Kontrolle über die Nutzung ihrer Daten in einem Forschungskonsortium einräumen

- Erhöht die Partizipation der Patienten (Data Usage Control mit Transparenz und Dynamic Consent) die Rechtssicherheit bei der Verwendung der Patientendaten für Forschung und Entwicklung?

Anonymisierung und Pseudonymisierung in Patientenversorgung und Forschung

Christoph Isele

23. Mai 2016

Primäre Nutzung von Gesundheitsdaten

- Z.B. BDSG §4: Einwilligung, Gesetze, eigene Geschäftszwecke, etc.
- Verarbeitung von Klardaten

- Elektronische Krankenakte
  - Fachdokumentation, Aufbewahrung, Nachfragen
  - Kommunikation mit Weiterbehandelnden
- Praxis-, Abteilungs-, Klinikumsorganisation
  - Terminplanung, Abrechnung
  - Behandlungssteuerung

Sekundäre Nutzung in der Klinik

- **Gesetzliche externe Qualitätssicherung**
  durch Gesetz abgedeckt, Pseudonymisierung, Treuhandstelle
- **Interne Qualitätssicherung, …**
  Aggregierte Ausgabe, Wahrung eines begründeten eigenen Interesse
  Datenschutzrechtliche Auswertung „Wer hat aus was zugegriffen“
- **Testsystem, Schulungssystem**
  realistische Testdaten durch Pseudonymisierung von Routinedaten
- **Klinische Forschung auf eigenen Routinedaten**
  Feasibility, Rekrutierung von Probanten für Studien
  Übernahme von Daten aus Krankenakte in Studien oder Register
  Pseudonymisierung, Treuhänder, Reduktion des Datenumfangs
- **Lehre, Ausbildung**
  Aufbereitung von Fallstudien, Übungsmaterial
  Pseudonymisierung, (Anonymisierung), Reduktion des Datenumfangs

Sekundäre Nutzung in der Industrie

- Wartung, Fehlersuche:
  - In der Praxis weitgehend ohne Zugriff auf Patientendaten
  - Übertragung von „Messdaten“ aus der Klinik ins Labor beim Hersteller zur Fehlerananlyse
  - Anonymisierung/De-Identifikation von DICOM Bildern,
  - Predictive Maintenance, Best Practice Hinweise, Benchmark auf der Basis von „Maschiniendaten“
  - Anonymisierung bei einfachen Datensätzen

Sekundäre Nutzung in der Industrie

- Produktverbesserung
  - Statistische Auswertungen der Log Datei
  - Analyse von Fehlerprotokollen
  - Optimierung von Programmen

  - Beispiel: Herzschrittmacher
  - (In Zukunft) entscheidungsunterstützende Software

Sekundäre Nutzung in der Industrie

- Produktneuentwicklung
  - Datenanalysen, UI Design
  - Datengetriebene Anwendungen zur Entscheidungs- und/oder Ablaufunterstützung

  - Beispiele:
    - Optimierung einer Regel zum frühen Erkennen von Sepsis;
    - Automatische Extraktion von Merkmalen aus Bildern
    - Automatische Extraktion von Schlüsselworten und Informationen aus Freitexten

„Beispiel für eine Architektur“

- Daten werden aus dem Primärsystem für eine weitere Nutzung „übertragen“ und können dabei aufbereitet werden
- Kopie der Quelldaten in ein Repository/DWH
  - Strukturierte Daten
  - Unstrukturierte Daten
  - Bilder, Kurven, Messreihen, *omics
- Pseudonymisierung auf dem Weg ins DWH
- Anonyme Data Marts aus DWH für Wartung, Produktverbesserung, -neuentwicklung

UNIVERSITÄT LEIPZIG

Anforderungen der Datenverarbeiter: Diskussion

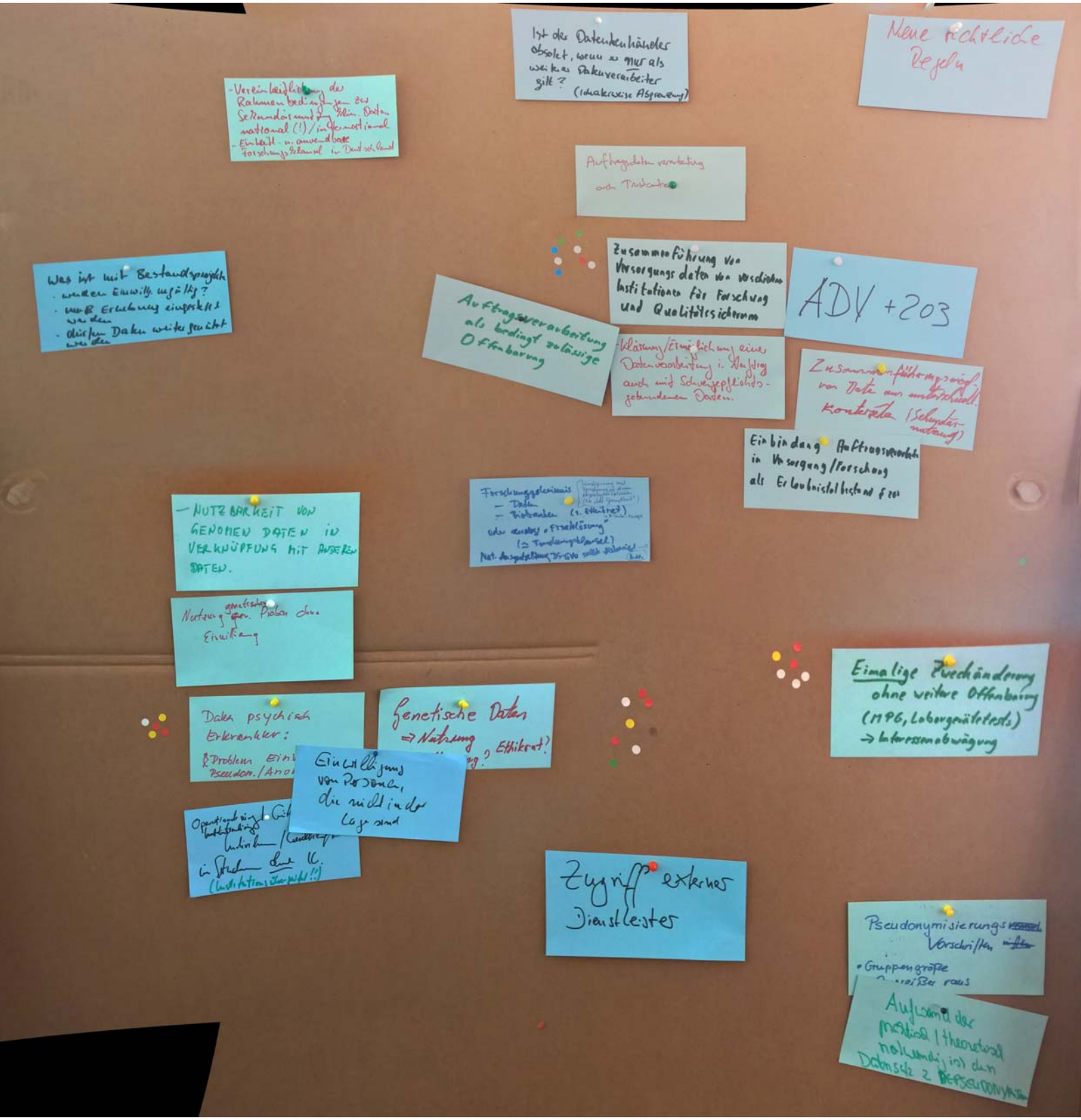
Alfred Winter

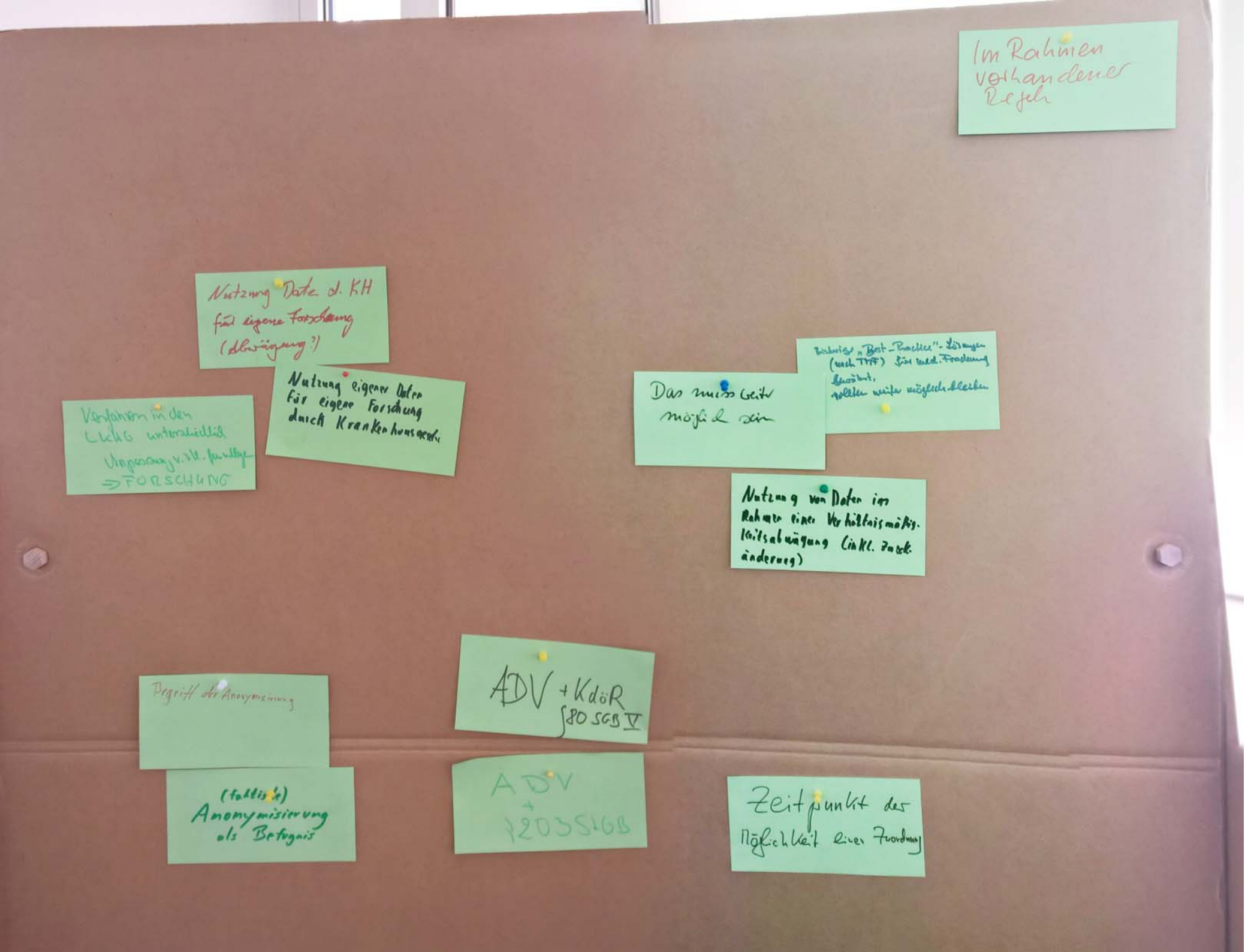
imise.
Institut für Medizinische Informatik, Statistik und Epidemiologie

- Möchten Sie weitere Problemfelder ergänzen, für die neue (rechtliche) Lösungen erforderlich sind?
- Wo besteht Handlungsbedarf,
  um nicht anonymisierte/anonymisierbare Daten nutzen zu können?
  - Maßnahmen im Rahmen geltender rechtlicher Möglichkeiten? (grüne Kärtchen)
  - Wo sind neue rechtliche Regeln erforderlich? (rote Kärtchen)
- Wer kann/muss einen Beitrag leisten neue Regeln zu entwickeln? Werden Konzepte wie von der TMF oder ob neue Gesetze/Verordnungen benötigt?

Weitere Problemfelder

- Continuity of Care
- Multi-Centes-Studien
- Zusammenführung von Registern
- Gerätetests in Lasermedizin
- MPG


Ist der Datenschutzbegriff obsolet, wenn es nur als weiches Bankverarbeiter gilt? (deskriptive Abgrenzung)
Neue rechtliche Regeln
- Vereinheitlichung der Rahmenbedingungen des Sekundärmarkts Klin. Daten national (!) / international
- Einheitl. u. anwendbare Forschungsklausel in Deutschland
Auftragsdatenverarbeitung von Treuhandstellen
Was ist mit Bestandsproben
- weiteren Einwilg. ungültig?
- weitere Erhebung eingeschränkt werden
- diesen Daten weiter geforscht werden
Zusammenführung von Versorgungsdaten von verschiedenen Institutionen für Forschung und Qualitätssicherung
ADV +203
Auftragsverarbeitung als bedingt zulässige Offenbarung
Klärung/Ermöglichung einer Datenverarbeitung i. Auftrag auch mit Schweigepflicht-gebundenen Daten.
Zusammenführung von Daten aus unterschiedl. Kontexten (Sekundärnutzung)
Einbindung Auftragsverarbeiter in Versorgung/Forschung als Erlaubnistatbestand § 203
- NUTZBARKEIT VON GENOMEN DATEN IN VERKNÜPFUNG MIT ANDEREN DATEN.
Forschungsgeheimnis
- Daten
- Proben (s. Ethikrat)
Nutzung genetischer Proben ohne Einwilligung
Daten psychisch Erkrankter:
§ Problem Einw. Pseudon./Anon.
Genetische Daten
⇒ Nutzung ... ? Ethikrat?
Einmalige Zweckänderung ohne weitere Offenbarung (MPG, Laborgeräteteststs)
→ Interessenabwägung
Einwilligung von Probanden, die nicht in der Lage sind
Zugriff externer Dienstleister
Pseudonymisierungs Vorschriften
• Gruppengröße
Aufklärung der medizin./theoretisch notwendig ist dem Datensatz z. Re-Pseudonymis.


Im Rahmen vorhandener Regeln
Nutzung Date. d. KH für eigene Forschung (Abwägung?)
Nutzung eigener Daten für eigene Forschung durch Krankenhausarzt
Das muss Gesetz möglich sein
Nutzung von Daten im Rahmen einer Verhältnismäßigkeitsabwägung (inkl. Zweckänderung)
Begriff der Anonymisierung
(faktische) Anonymisierung als Betrugnis
ADV + KdöR §80 SGB V
ADV + 1203S168
Zeitpunkt der Möglichkeit einer Trennung

Die EU-DSGVO und die „anonymen Daten“

Kann man und wann kann man Daten nach der EU-DSGVO als „anonym“ ansehen?

23.05.2016

Gerald Spyra,
LL.M.
Kanzlei Spyra

gerald.spyra@kanzlei-spyra.de

Vorstellung meiner Person

Gerald Spyra, LL.M.

- **Rechtsanwalt**
- **Hohe Affinität für die Informationssicherheit**
- **Spezialisiert auf:**
  - **den Informations- / Datenschutz,**
  - **das Software-Medizinprodukterecht**
  - **die IT-Forensik**
- **Externer betrieblicher Datenschutzbeauftragter**

gerald.spyra@kanzlei-spyra.de

Vorbemerkung

- **Von vielen Seiten hört man, dass es mit Geltung der EU-Datenschutzgrundverordnung (VO) keine anonymen Daten mehr geben soll!?**

- **Dieses insbesondere deshalb, weil die „Anonymisierung“ bzw. der Begriff „anonym“ nicht mehr in den Regelungen der VO erwähnt ist.**

- **Kurz und bündig:**
  „Weil die „Anonymisierung“ nicht in den Regelungen enthalten ist, gibt es sie auch nicht mehr“!

- **Aber so leicht ist es wiederum auch nicht…**

gerald.spyra@kanzlei-spyra.de

Die Geltung der EU-DSGVO

- Die VO enthält in ihren Regelungen **nur Vorgaben** dazu, **wann** sie **Anwendung** findet.

- Sie **findet** immer dann **Anwendung**, wenn Daten verarbeitet werden (sollen), die einen **Personenbezug ermöglichen** bzw. eine Person mittels dieser Daten „**identifizierbar**“ ist / wird.

- Im **Umkehrschluss** muss dieses aber auch bedeuten, dass die **Regelungen** der VO **keine Anwendung** finden, wenn die **Daten keinen Personenbezug** (mehr) zulassen.

- In diesem Punkt ähneln die Regelungen der VO auch der **Datenschutz-Richtlinie 95/46/EG**.

gerald.spyra@kanzlei-spyra.de

Die Anonymität in der EU-DSGVO

- **Dass es aber weiterhin „anonyme“ Daten gibt / geben soll, sagt bspw. Erwägungsgrund (EG) 26.**

- **EG 26 ist der maßgebliche Erwägungsgrund u.a. zur Regelung von Artikel 4 Nr. 1 der VO, in dem „personenbezogene Daten“ definiert werden…**

- **N.B.: Leider ist dieser EG „nur“ ein Erwägungsgrund (Auslegungshinweis) und keine eigenständige Regelung im Verordnungstext!**

gerald.spyra@kanzlei-spyra.de

EG 26 (Teil 1)

- **So heißt es in EG 26:**

„Die Grundsätze des Datenschutzes sollten daher nicht für anonyme Informationen gelten,

d.h. für Informationen, die sich nicht auf eine identifizierte oder identifizierbare natürliche Person beziehen,

ODER

personenbezogene Daten, die in einer Weise ANONYMISIERT worden sind, dass die betroffene Person nicht oder nicht mehr identifiziert werden kann.“

- **Nach EG 26 existieren deshalb „zwei Möglichkeiten“ für anonyme Daten…**

gerald.spyra@kanzlei-spyra.de

EG 26 – die „zwei Möglichkeiten“

- EG 26 zeigt „zwei Möglichkeiten“ auf, wann Daten anonym sind / werden können.
  - 1. Daten sind von „Natur“ aus anonym
  - 2. Personenbezogene Daten werden so verarbeitet (Art. 4 Nr. 2 = Erlaubnistatbestand notwendig), dass sie „anonym“ werden und bleiben…

- Das bedeutet mithin, dass wann immer eine Person (für wen auch immer) mittels entsprechender Informationen identifizierbar ist, die Daten nicht mehr anonym sein können.

- Und wann eine Person (re-) identifizierbar ist bzw. (re-) identifiziert werden kann, sagt uns auch EG 26…

gerald.spyra@kanzlei-spyra.de

Erwägungsgrund 26 – (Re-)Identifizierbarkeit

- ***„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“.***

- **1. Voraussetzung:**
- **Der Verantwortliche oder (jede) andere Person = KEINER darf mehr in der Lage sein, eine Person mittels der Daten zu (re-) identifizieren.**

- **2. Voraussetzung:**
- **Es gilt ferner alle zur Verfügung stehenden Mittel zu berücksichtigen (und ihre „Kombination“), die nach allgemeinen Ermessen, wahrscheinlich genutzt werden (können), um eine Person zu identifizieren….**

gerald.spyra@kanzlei-spyra.de

Erwägungsgrund 26 – Mittel

- ***„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.“***

- **Alle objektiven Faktoren müssen deshalb herangezogen werden, wozu bspw. die zu Verfügung stehenden Ressourcen wie:**
  - **Kosten / Zeit oder**
  - **technologische Entwicklungen (zum Zeitpunkt der Verarbeitung) gehören.**

- **Damit lässt sich nun auch das Verhältnis „anonymer“ zu „pseudonymer“ Daten darstellen…**

gerald.spyra@kanzlei-spyra.de

Verhältnis anonymer vs. pseudonymer Daten

- **Pseudonyme Daten können niemals anonyme Daten sein (siehe auch WP 216 der Art. 29 Gruppe)!**
- **Diese, von der Art. 29 Datenschutzgruppe vertretene Ansicht, wird höchstwahrscheinlich auch beim EU-Datenschutzausschuss, der die Art. 29 Gruppe ersetzt (und gleich besetzt sein wird) fortbestehen.**
- **Sobald deshalb (bei wem auch immer) eine Zuordnungsregelung existiert, fehlt es an der „Anonymität" von Daten (Vgl. Art. 4 Nr. 5).**
- **Die „faktische Anonymität" dürfte es damit nicht mehr geben.**
- **Daher lassen sich folgende Schlüsse ziehen....**

gerald.spyra@kanzlei-spyra.de

Zusammenfassung – „Anonyme Daten“

- Es gibt auch **mit Geltung** der **VO** weiterhin noch **anonyme Daten**.
- Anonyme Daten zeichnet aus, dass **KEINER** mehr aus ihnen eine **Identifizierung** einer Person (mit seinen Mitteln) vornehmen kann.
- Ferner zeichnet anonyme Daten aus, dass sie entweder
  - von „**Natur**“ aus anonym sind oder
  - durch eine **entsprechende Verarbeitung** anonym werden (und bleiben).
- Daten aus dem **medizinischen Umfeld** und die zur **Verfügung stehenden Identifizierungsmöglichkeiten**, lassen die **erste Alternative** (von „Natur“ aus anonym) jedoch immer **weniger wahrscheinlich** werden…

gerald.spyra@kanzlei-spyra.de

Begebenheiten, die die „Anonymität“ von Daten in Frage stellen können… (1)

- Viele der im Gesundheitsbereich verwendeten Daten sind so „verdichtet“, dass sie oftmals von sich aus eine (Re-) Identifizierung ermöglichen.

- Ferner gilt es folgende Aspekte zu berücksichtigen:
  - Einzelne atypische Vorkommnisse lassen eine Person, direkt identifizierbar werden (z. B. der Mann mit Brustkrebs).
  - Moderne Technik gewinnt aus vermeintlich als anonym angesehenen Daten personenbeziehbare Informationen (z. B. Möglichkeit der 3D- Konstruktion eines Gesichts aus Röntgenbildern).
  - Immer mehr „smarte“ Geräte des Alltags „tracken“ Nutzer und ihre Tätigkeiten (Metadaten)
  - „BigData“ eröffnet Möglichkeit, Daten praktisch „unendlich“ miteinander bzw. anderen Daten zu kombinieren;

gerald.spyra@kanzlei-spyra.de

Begebenheiten, die die „Anonymität“ von Daten in Frage stellen können… (2)

- Und noch mehr:
  - Der „genetische Fingerabdruck“ (70 SNPs reichen meistens aus – oftmals auch weniger);
  - Der Aufbau öffentlicher / privater Gen- / Biobanken, ihre Vernetzung und der Austausch von Daten / gemeinsame Verarbeitungsmöglichkeiten ermöglichen (Re-) Identifizierbarkeit;
  - Die steigende „Mitteilungsbedürftigkeit“ über „social networks“ ermöglicht, die preisgegeben Daten mit medizinischen Daten in Kontext zu setzen;
  - …

- Aufgrund der vielen Unsicherheitsfaktoren bei „anonymen“ Daten sollte man deshalb prüfen, ob man die Daten nicht entsprechend der zweiten Alternative „bearbeiten“ kann, um sie „anonym“ nutzen zu können.

gerald.spyra@kanzlei-spyra.de

„Anonymisierung“ als Verarbeitung

- Die „Anonymisierung“ / „Pseudonymisierung“ von Daten stellt wie bisher einen Verarbeitungsvorgang (Art. 4 Nr. 2) dar.

- Daraus folgt, dass auch zur „Anonymisierung“ ein Ermächtigungsgrund dem Verantwortlichen zur Verfügung stehen muss.

- Dem Verbot mit Erlaubnisvorbehalt folgend, kann dieser entweder eine (wirksame) Einwilligung des Betroffenen oder eine gesetzliche Regelung sein.

gerald.spyra@kanzlei-spyra.de

Legitimation durch VO / Rechtsvorschrift

- Die Regelungen der VO sind abschließend (VO geht deutschen Regelungen vor / verdrängt sie).
- Es existiert kein expliziter Legitimationstatbestand für die „Anonymisierung“ in der VO.
- Jedoch existieren „Nationale Öffnungsklauseln“ für die Verarbeitung von Gesundheitsdaten.
- Damit ist der bzw. die deutschen Gesetzgeber gefordert, neue Regelungen zu schaffen bzw. vorhandene Regelungen wie § 28 Abs. 6, Abs. 7 BDSG beizubehalten.
- Ansonsten existiert die Möglichkeit der Einwilligung des Betroffenen, die jedoch auch nicht ohne Probleme ist…

gerald.spyra@kanzlei-spyra.de

Legitimation durch Einwilligung

- Weiterhin besteht die Möglichkeit, die Einwilligung des Betroffenen in die „Verarbeitung“ seiner Daten einzuholen.

- Diese muss jedoch den Anforderungen der VO nach insbesondere informiert, freiwillig sein und sich auf einen bzw. mehrere festgelegte Zwecke beziehen.

- Daher besteht ein Problem der Einwilligung beim „broad consent“ (wann ist es noch „bestimmt“ genug?).

- Ein weiteres Problem existiert bei der Einwilligung bei der Verarbeitung bestimmter Daten....

gerald.spyra@kanzlei-spyra.de

Herausforderung bei der Einwilligung bei „besonderen“ Daten

- Soll die **Einwilligung** in die Verarbeitung von Daten gegeben werden, die **nicht nur den Einwilligenden betreffen**, treten naturgemäß weitere Fragestellungen auf.

- Kann der **Betroffene** etwa in die Verarbeitung dieser Daten **einwilligen**, wenn sie bspw. auch **seine Vorfahren**, **Nachfahren**, **Familienangehörigen** betreffen?

- Lässt sich hierfür eine **gesetzliche Legitimation** schaffen (**Nationale Öffnungsklausel für genetische Daten**)?

- Und selbst wenn man eine Legitimation hat, muss man sich dann noch mit der **Frage** auseinandersetzen, welchen **Wert** als „**anonymisiert anzusehende Daten**“ für die **moderne Forschung** überhaupt noch haben können?

gerald.spyra@kanzlei-spyra.de

Nutzen von „anonymisierten“ Daten

- **Selbst wenn man Daten rechtskonform „anonymisiert“ hat, bleibt stets die Frage, ob sich die „anonymisierten“ Daten dann noch für die mit der jeweiligen Forschung beabsichtigten Zwecke überhaupt eignen.**

- **Die zu ergreifenden technischen und organisatorischen Maßnahmen, um einen Personenbezug auszuschließen, dürften die Nutzbarkeit von „anonymisierten Daten“ deutlich limitieren / reduzieren.**

- **Und daher gibt es viel zu diskutieren…**

gerald.spyra@kanzlei-spyra.de

Gibt es noch Fragen?

Gerald Spyra, LL.M.
Rechtsanwalt,
externer Datenschutzbeauftragter

www.recht-technisch.de

gerald.spyra@kanzlei-spyra.de

Kanzlei Spyra
Kaiserstr. 7
51688 Wipperfürth

Vielen Dank für Ihr Interesse!

gerald.spyra@kanzlei-spyra.de

Anonymisierungsverfahren

Dr. M Sariyar / Dr. J Drepper
TMF

Outline

- Relevante Grundbegriffe
- Privacy-Kriterien und Risikomodellierung
- Anonymisierungsverfahren

Relevante Grundbegriffe

Anonymisierungsbegriff

- **ISO 29100:2011:** “Anonymization is the **process** by which personally **identifiable** information (PII) is **irreversibly** altered in such a way that a PII principal can no longer be **identified** directly or indirectly, either by the PII controller alone or in collaboration with any other party.”

- **Wiederholung**: Das **Re-Identifizierungsrisiko** in den Daten soll reduziert und dennoch die **Nützlichkeit** der Daten für vielfältige Analysen erhalten werden

Gegen welche Risiken wird gesichert?

- Reidentification
  - Singling out: einen Datensatz, der zu einem Inviduum gehört, isoliert
  - Record Linkage: Datensätze als zu einem Individuum gehörig klassifiziert
- Attribute disclosure
  - Attribute Linkage: Sensitive Werte für Individuen geschlussfolgert
  - Probabilistische Inferenz: Erhöhung der Wahrscheinlichkeit für die Schlussfolgerung über sensitive Werte
- Membership disclosure
  - Table Linkage: Schlussfolgere die Präsenz eines Individuums

Arten von Attributen

1. Global-eindeutige (z.B. Sozialversicherungs-Nr.) und direkte Identifikatoren (z.B. Name)
   => Unbedingt Löschen zum Erreichen von Anonymität

2. **Quasi-Identifikatoren** (z.B. PLZ, Alter, Geschlecht) => QIDs

3. Sensitive Attribute (z.B. Krankheitsstatus)

4. Non-Sensitive Attribute

Quasi-Identifier

OECD-Definition für Quasi-Identifier:

“Variable values or combinations of variable values within a dataset that are not structural uniques but might be empirically unique and therefore in principle uniquely identify a population unit.”

Arten von Attributen

Irrelevant		QIDs		SensAttr
ID	Geschlecht	Geburtsdatum	PLZ	ICD-10 Code
6	M	1980	10117	Q90.1
8	F	1966	10117	F31.1
1	M	1979	10118	F31.0
9	M	1988	11067	F31.9
11	F	1965	11910	G50.1
4	F	1983	11934	F34.8
10	M	1973	12002	F34.8
3	F	1967	12033	F31.9
2	M	1989	12200	F31.1
5	F	1959	12200	G50.1
12	M	1976	13011	Q90.1
7	M	1975	13135	Q90.0

Privacy-Kriterien und Risikomodellierung

Häufig genannte syntaktische Privacy-Kriterien für personen-beziehbare strukturierte Daten

- ***k*-Anonymity**: Datensätze mit gleichen Werten für die QIDs tauchen mindestens k mal auf (Äquivalenzklasse) => Re-Identifikationsrisiko wird auf maximal 1/k festgelegt!

- **Distinctive *l*-Diversity**: Es gibt mindestens l verschiedene Ausprägungen des sensitiven Attributs in einer Äquivalenzklasse

- **Alternativen zu syntaktischen Kriterien**
  - Risk-based models (häufig in der Literatur zu Statistical Disclosure Control)
  - Semantic privacy models (e.g., differential privacy)

k-anonymity und l-diversity

PLZ	Alter	Krankheit
476**	2*	COPD
476**	2*	COPD
476**	2*	COPD
4790*	≥40	AIDS
4790*	≥40	COPD
4790*	≥40	Krebs
476**	3*	COPD
476**	3*	Krebs
476**	3*	Krebs

keine l-Diversität

mind. 2-Diversität

Anonymisierungsverfahren

Einige Anonymisierungsverfahren für Tabellen

- **Generalisierung und Suppression** (Details in QIDs verstecken)
  - Ersetze Werte in der höheren Ebene einer Generelisierungshierarchie
  - Full-domain oder lokale (subtree, cell) Generalisierung
  - Suppression

- **Perturbation**: z.B.
  - Additive Noise (z.B. Randomization),
  - Data swapping
  - Microaggregation: teile Datensatz in homogene Cluster der Länge k und ersetze alle Attributwerte im Cluster durch einen Wert (Mittelwert, Modalwert, etc.)

Rahmen für Anonymisierung

- Komplexe Aufgabe, da
  - Viele Verfahren existieren, die miteinander kombiniert werden können
  - Methoden oft geeignet zu parametrisieren sind

- Kontext ist zu berücksichtigen:
  - Anwendung (Nutzer, Typen & Prozess. von Daten, gewünschte Analysen, Release-Mechanismus, etc.)
  - Risiken sind zu modellieren (unterschiedliche Möglichkeiten)
  - Nutzen ist zu bewerten (unterschiedliche Möglichkeiten)
    - General purpose metric: z.B. information loss
    - Special purpose metric: z.B. für logistische Regression

Aktivitäten der TMF: Ressourcen und Unterstützung

- 2013: Weiterentwicklung des OpenAnonymizer zum TMF-ANON-Tool
  - unterstützt k-Anonymisierung u. l-Diversifizierung

- 2015: TMF-Workshop zu Anonymisierungstools:
  - ANON: a flexible tool for achieving k-anonymous and l-diverse tables
  - ARX: Comprehensive Tool for Anonymizing Biomedical Data
  - MuArgus: Software to produce safe microdata
  - sdcMicro and sdcMicroGUI: R-packages for SDC

  (Nachbericht und Folien s. www.tmf-ev.de/news/1706)

- 2016: TMF erarbeitet aktuell ein Schulungskonzept
  - Erste Evaluationsschulung am 18.5.2016 durchgeführt
  - Folgetermin voraussichtlich am 7.7.2016 (bereits ausgebucht)

Umsetzungskonzepte und Möglichkeiten der Pseudonymisierung

TMF Workshop
„Anonymisierung und Pseudonymisierung“

Martin Bialke

Berlin, 23. Mai 2016

Community Medicine

Universitätsmedizin Greifswald

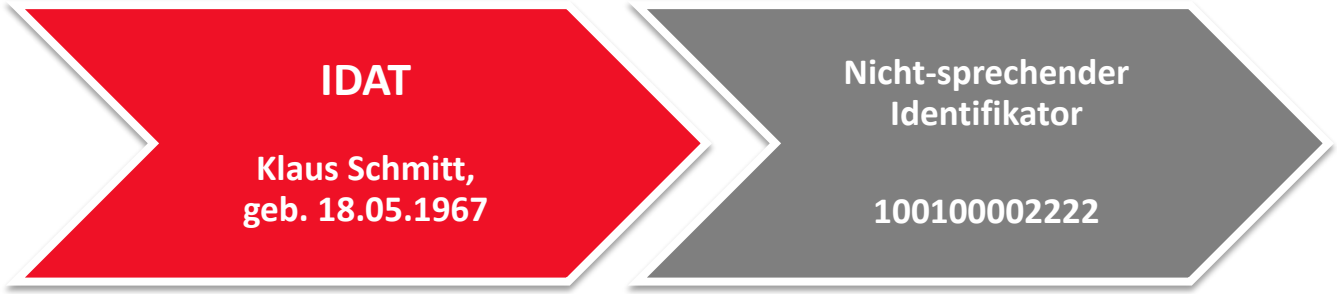
Agenda

- Stufen der Pseudonymisierung
- Methoden der Pseudonymisierung
- Anwendungsbeispiel
- Diskussion

Stufenweise Pseudonymisierung (1/3)

Pseudonym 1. Stufe

- Erstellungsprozess sollte Matching-Verfahren und Doppler-Ausschluss umfassen



Stufenweise Pseudonymisierung (2/3)

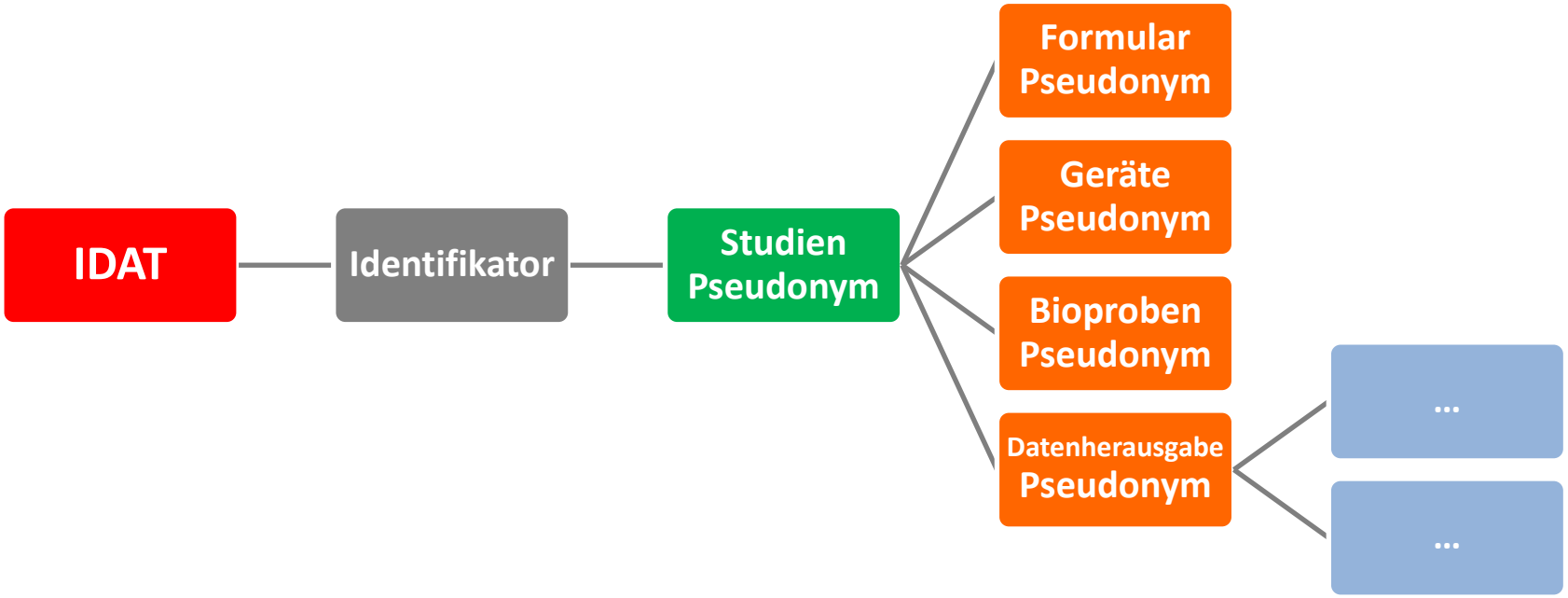
Pseudonym 2. Stufe

- wird anstelle der IDAT im Forschungskontext genutzt
- Zuordnung/Zuordnungsvorschrift darf nur autorisiertem Personal (Datentreuhänder) bekannt sein

Stufenweise Pseudonymisierung (3/3)

Pseudonym N-ter Stufe

- Steigender Aufwand zur Re-Identifizierung je Stufe (zumindest für unautorisierte Personen)
- Ideal: unterschiedliche Pseudonyme je Anwendungskontext



Methoden der Pseudonymisierung (1/3)

Pseudonymerzeugung mittels Schlüssel

- Pseudonym wird durch kryptografischen Algorithmus und festen Schlüssel erzeugt
- erzeugtes Pseudonym ist abhängig vom Eingabewert (z.B. PID), dadurch Rückrechnung mittels Schlüssel möglich (De-Pseud.)
- *Folge: erhöhte Aufwände bei Zwang zu Schlüsselwechsel (Beispiel: Mitarbeiterwechsel)*

Vertreter

- Kombination von PID-Generator (Patientenliste) und Pseudonymisierungsdienst (PSD) der TMF

Quelle: K. Pommerening et al. **Leitfaden zum Datenschutz in medizinischen Forschungsprojekten Generische Lösungen der TMF – Version 2** , Berlin 2014

Methoden der Pseudonymisierung (2/3)

Pseudonymerzeugung mittels Hash-Verfahren

- „Verschlüsselung“ des Eingabewertes z.B. mittels SHA-3
- Sicherheit des Pseudonyms abhängig von Komplexität des Eingabewerts (+ Salt) und Hash-Algorithmus
- *Aber:*
  - *Gefahr von Kollisionen (versch. PIDs, gleiche Pseudonyme)*
  - *wird Hash-Verfahren bekannt, können Eingabewerte ggf. rechnerisch ermittelt werden (De-Pseudonymisierung)*

Vertreter

- Zahlreiche Bibliotheken und Online Plattformen

Methoden der Pseudonymisierung (3/3)

Pseudonymgenerierung und Mapping

- Erzeugung von Pseudonym mittels Generatoralgorithmus und Alphabet
- Zuordnung von Eingabewert und Pseudonym über kontextspezifische Mapping-Tabelle -> Pseudonym unabhängig vom Eingabewert
- *Vorteil: bei Löschung der Mapping-Informationen kann Eingabewert in keinem Fall wiederhergestellt werden (Anonymisierung)*

Vertreter: gPAS a generic pseudonym administration service

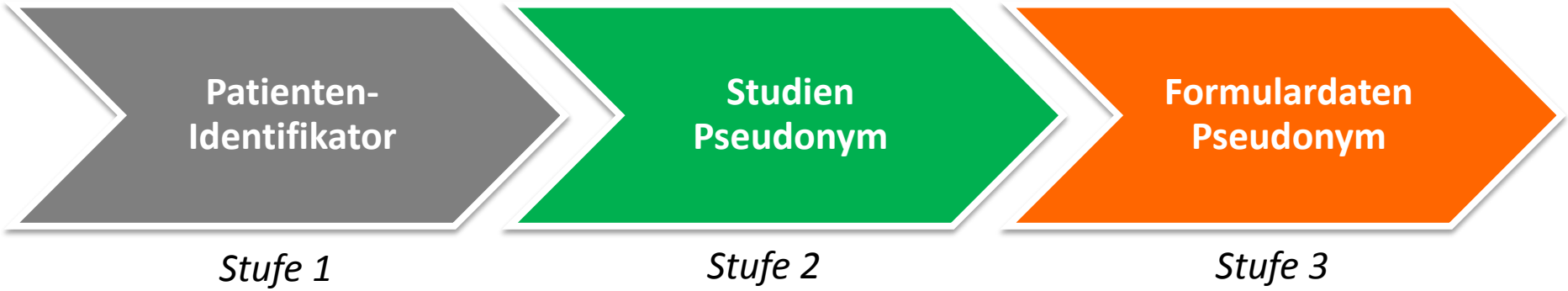
Anwendungsbeispiel (1/8)

Szenario

- Föderierte Studie mit zentraler Datenhaltung
- Verzicht auf Record Linkage
  -> Standortinterner Patienten-Identifikator vorhanden
- Bioproben und Daten aus eCRF, Laborgeräten
- Sekundärnutzung: Auf Antrag Herausgabe erneut pseudonymisierter Daten an Forschungsvorhaben

Anwendungsbeispiel (2/8)

Ziel: spezifische Pseudonymisierung je Datentyp bzw. Anwendungszweck (Datenherausgabe)



Anwendungsbeispiel (3/8) – gPAS Kurz & Knapp

- Pseudonyme generieren und verwalten
- Verarbeitung beliebiger Zeichenfolgen
- Mehrfach-Pseudonymisierung
- De-Pseudonymisierung und Anonymisierung
- Flexibel konfigurierbar
- Integration von Altpseudonymen
- Anzeige von Pseudonymhierarchien

Anwendungsbeispiel (4/8)

Einrichtung von Pseudonymdomänen

gPAS
a generic pseudonym administration service

Community Medicine

Pseudonym Administration

Domain Management | Pseudonym Management | Batch Processing | Statistics

Domain Name	Parent Domain Name	Comment	Alphabet	Check Digit Generator	Properties	Pseudonyms
device	mosaic_study	psns for device data	Symbol32	NoCheckDigits	PSN_LENGTH=10 PSN_PREFIX=dev_	1
forms	mosaic_study	psns for forms data	Numbers	NoCheckDigits	PSN_LENGTH=10 PSN_PREFIX=frm_	1
mosaic_study		Demo Study	Hex	NoCheckDigits	PSN_LENGTH=10 PSN_PREFIX=msc_	2
old_data	mosaic_study	imported psns	Symbol32	HammingCode	PSN_LENGTH=16 PSN_PREFIX=old_	1
samples	mosaic_study	psns for bio samples	Symbol31	NoCheckDigits	PSN_LENGTH=12	1
secondary_use	mosaic_study	psns for secondardy use (use and access process)	Numbers	VerhoeffGumm	PSN_LENGTH=10 PSN_PREFIX=ua_	1

Refresh | Delete Domain

Create A New Domain

Domain Name	Parent Domain Name	Comment	Alphabet	Check Digit Generator	Properties
Domain Name	Parent Domain Name		define custom alphabet	select check digit generator	MAX_DETECTED_ERRORS PSN_LENGTH PSN_PREFIX PSN_SUFFIX INCLUDE_PREFIX_IN_CHECK_DIGIT_CALCULATION

Create Domain | Reset

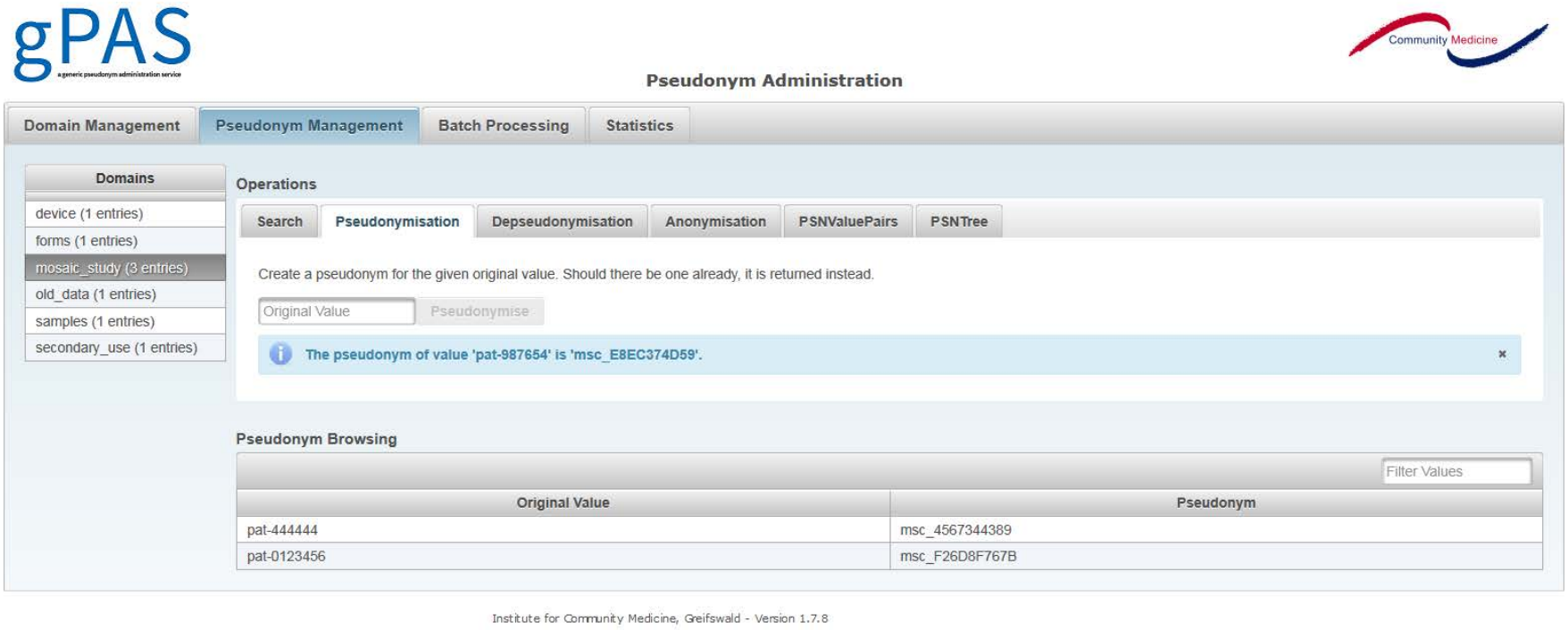
Institute for Community Medicine, Greifswald - Version 1.7.8

Community Medicine

Universitätsmedizin Greifswald

Anwendungsbeispiel (5/8)

Pseudonymisierung und De-Pseudonymisierung

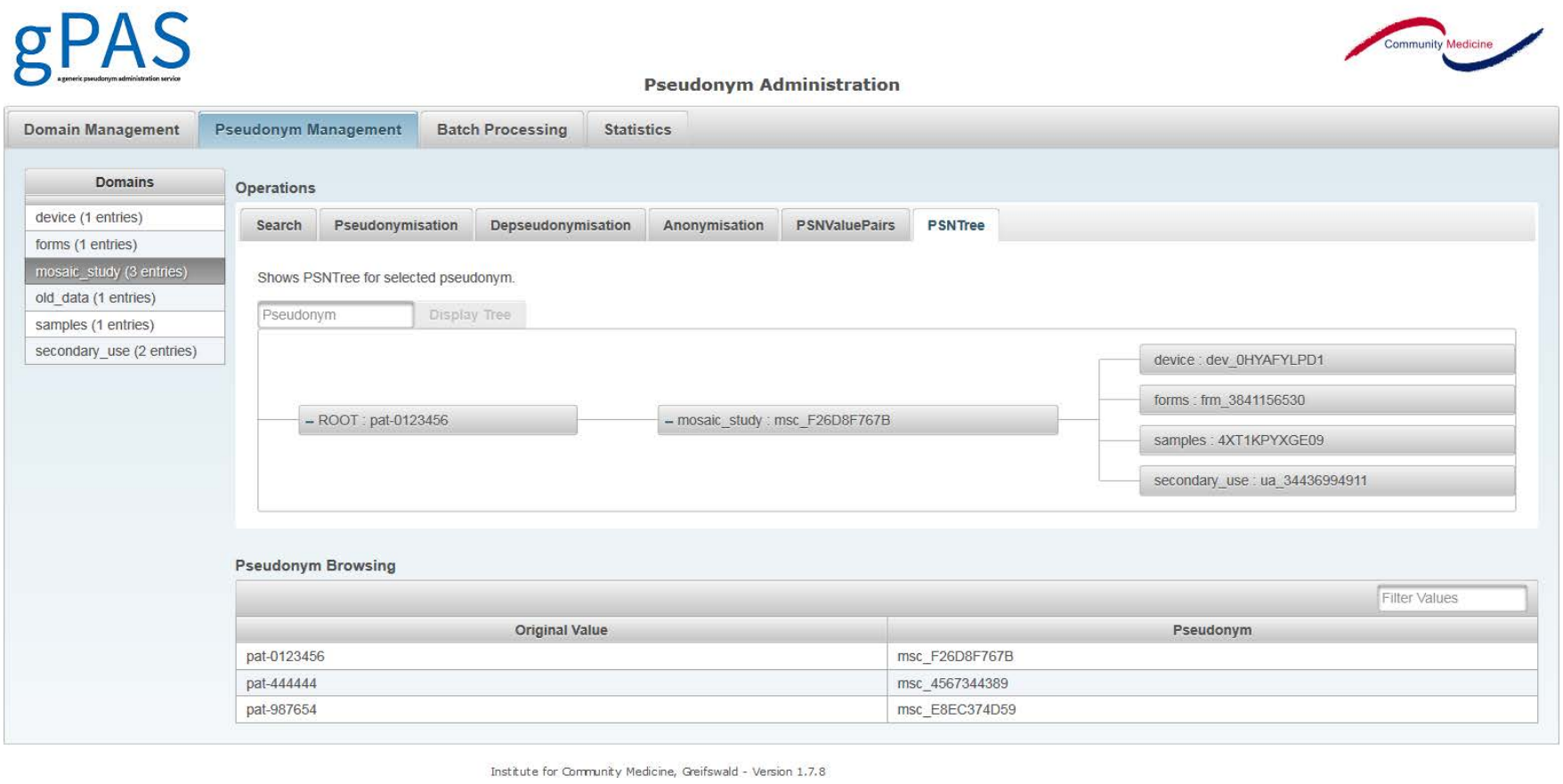


Community Medicine

Universitätsmedizin Greifswald

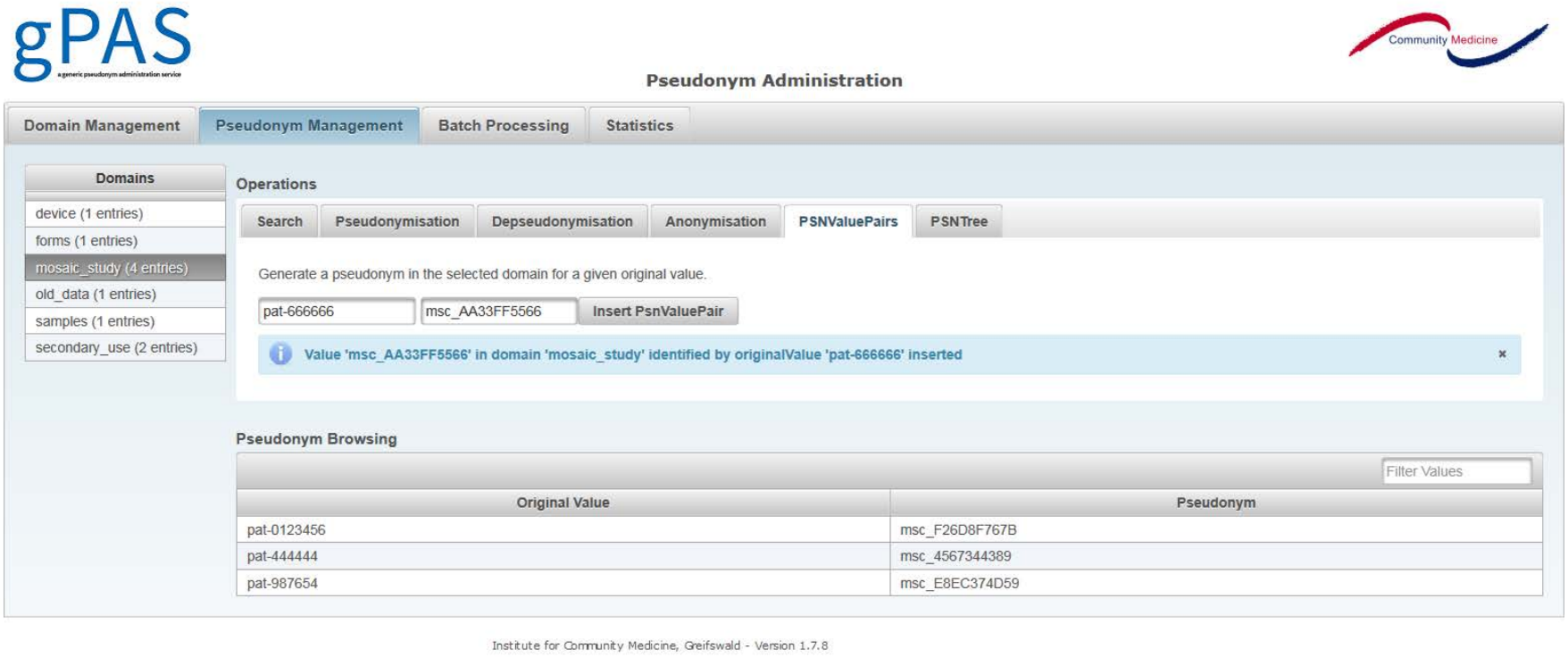
Anwendungsbeispiel (6/8)

Hierarchische Darstellung



Anwendungsbeispiel (7/8)

Integration vorhandener Pseudonyme

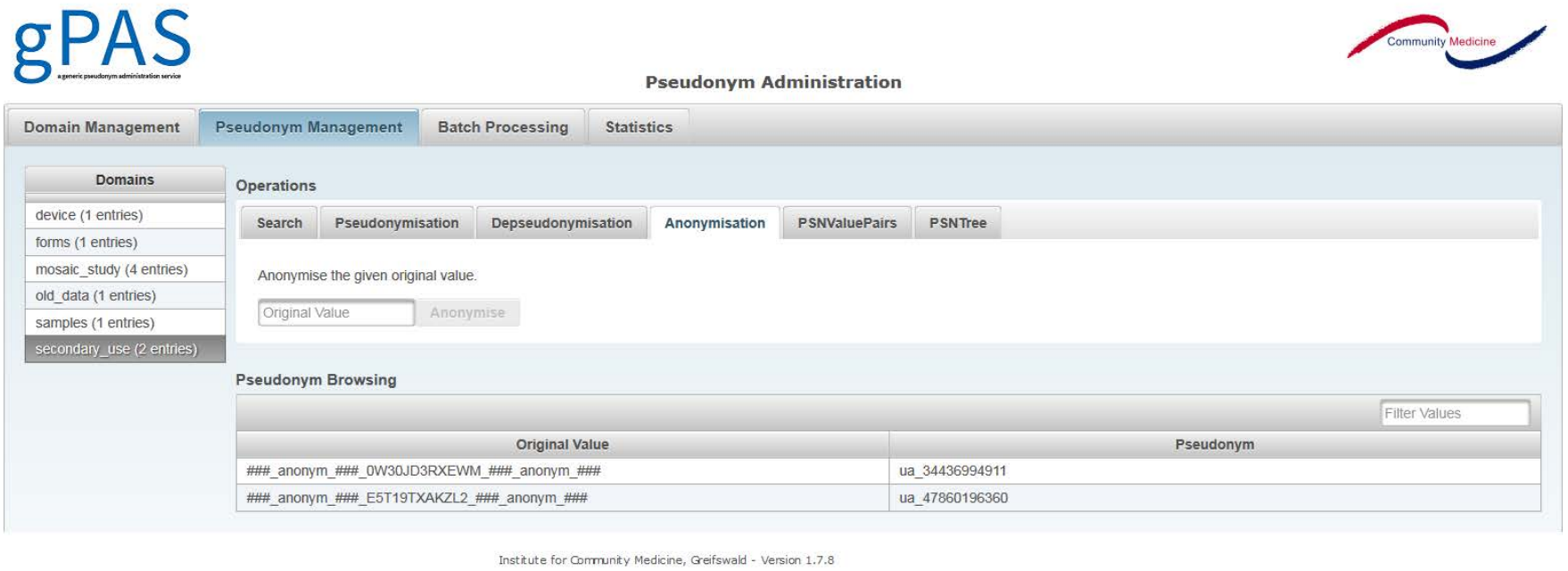


Community Medicine

Universitätsmedizin Greifswald

Anwendungsbeispiel (8/8)

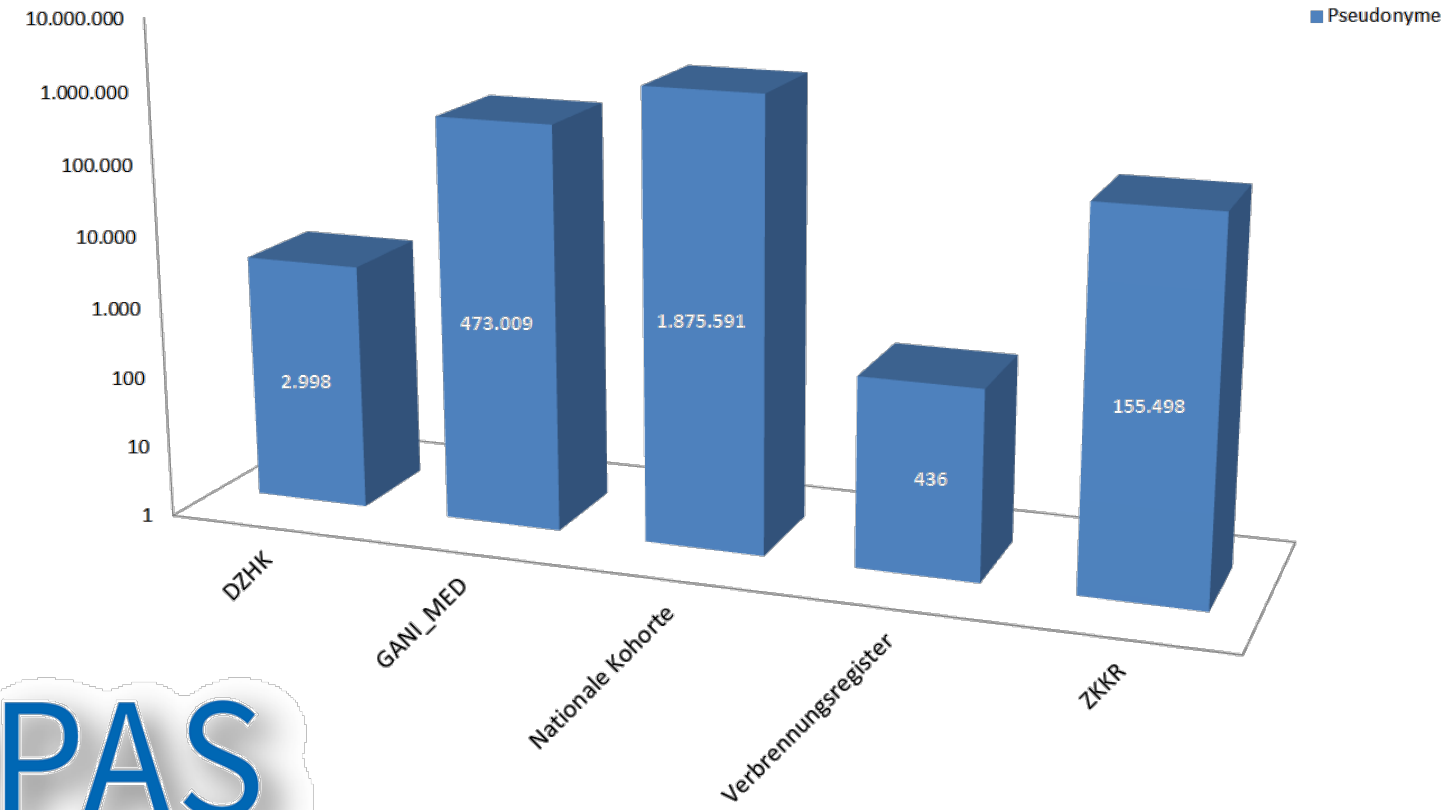
Irreversible Löschung von Zuordnungen (Anonymisierung)



Community Medicine

Universitätsmedizin Greifswald

Blick in die Praxis



gPAS
a generic pseudonym administration service

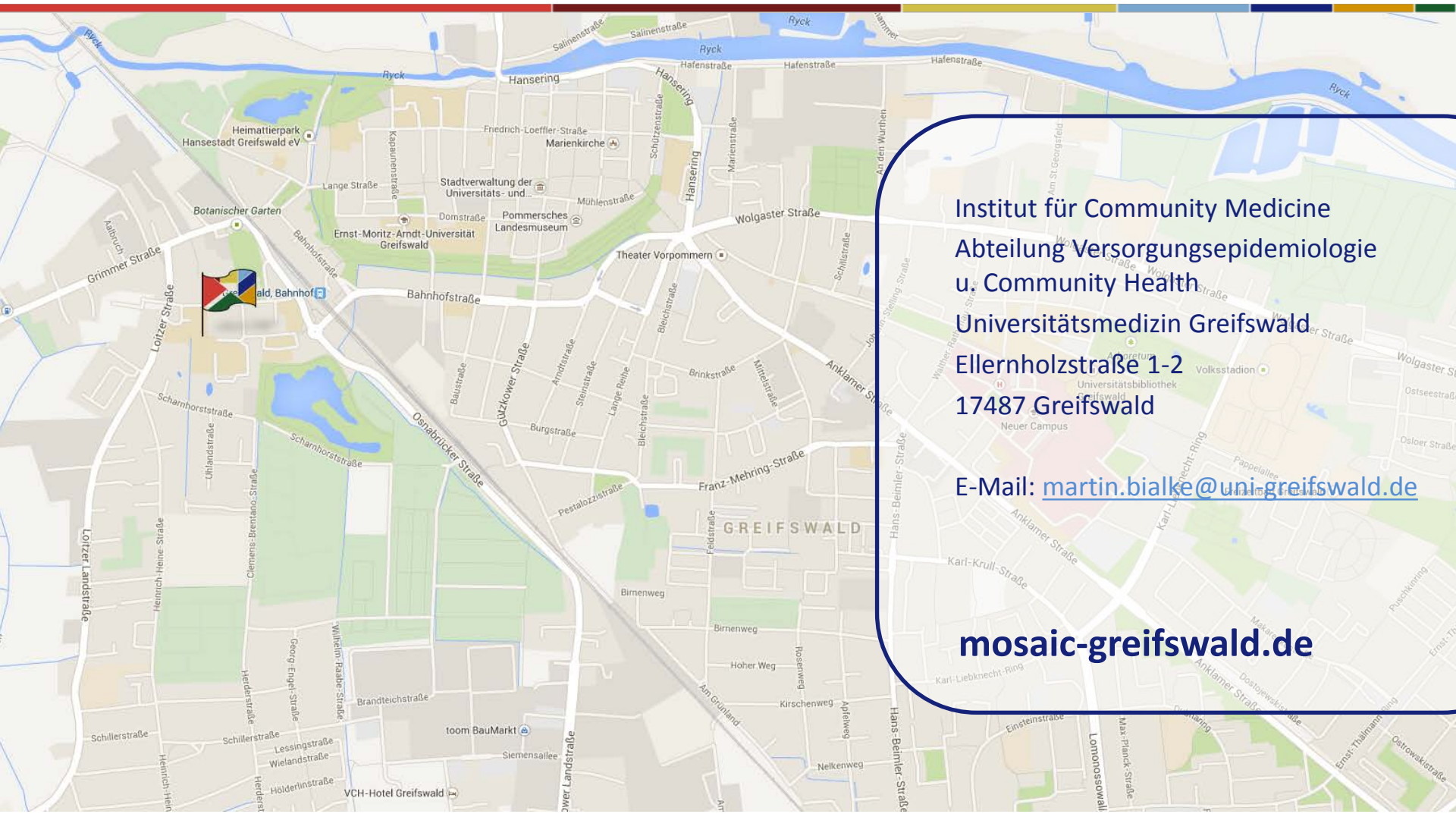
2.507.532 Pseudonyme

Stand: 03.05.2016

Diskussion

- Zählt der Datentreuhänder trotz informationeller Gewaltenteilung nach EU-DSGVO zu den Datenverarbeitern?
- Wie können kleinere Forschungsprojekte bei der Pseudonymisierung von Forschungsdaten unterstützt werden?

Vielen Dank für Ihre Aufmerksamkeit



Institut für Community Medicine

Abteilung Versorgungsepidemiologie u. Community Health

Universitätsmedizin Greifswald

Ellernholzstraße 1-2

17487 Greifswald

E-Mail: martin.bialke@uni-greifswald.de

mosaic-greifswald.de