Umsetzungskonzepte und Möglichkeiten der Pseudonymisierung

TMF Workshop
„Anonymisierung und Pseudonymisierung“

Martin Bialke

Berlin, 23. Mai 2016

Community Medicine

Universitätsmedizin Greifswald

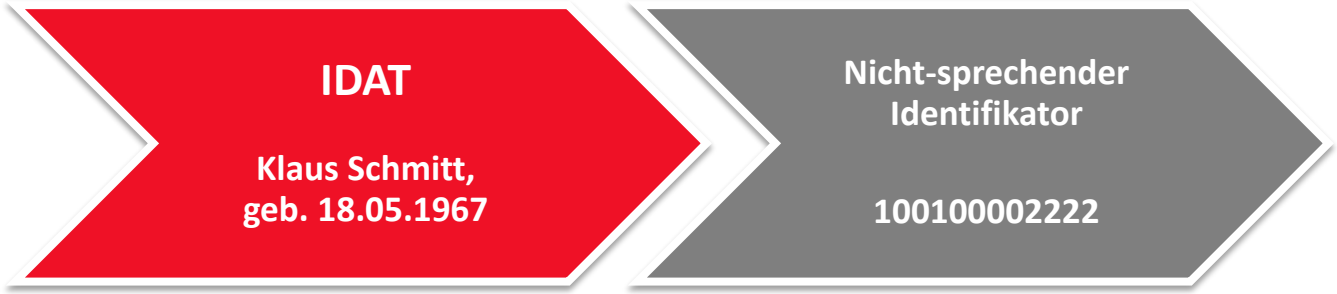
Agenda

- 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)

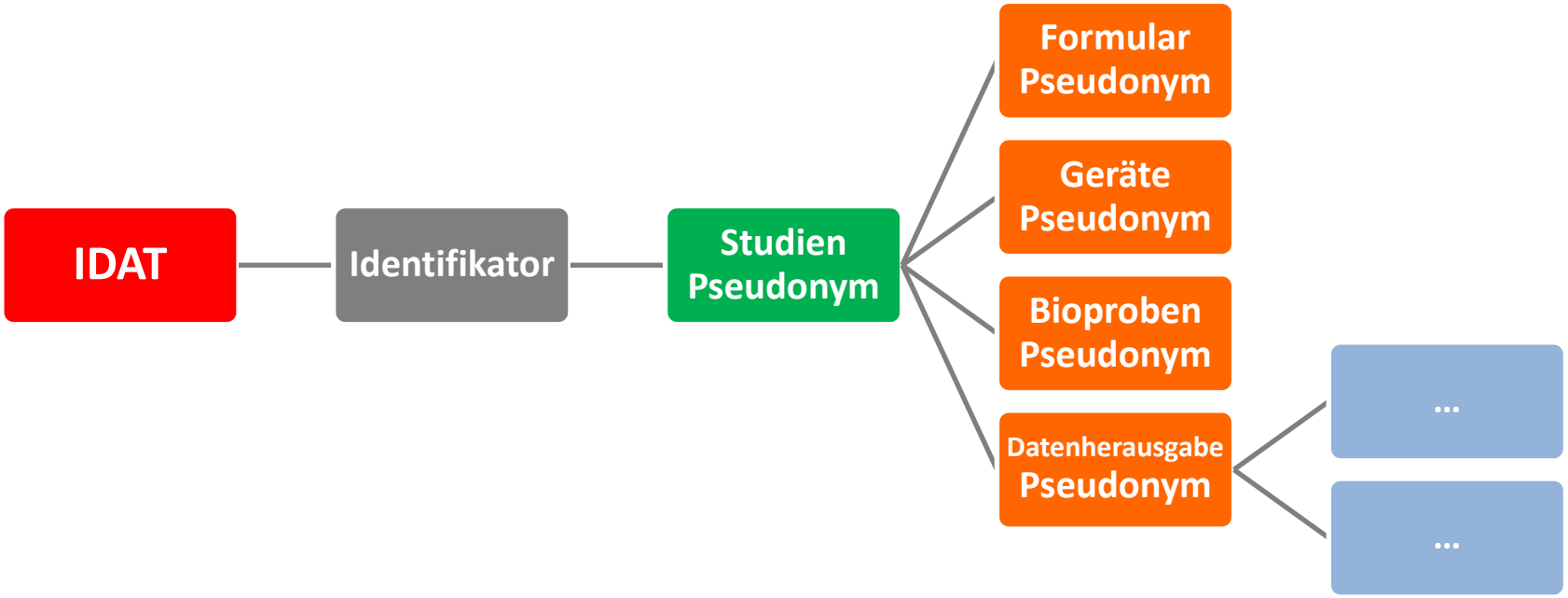
Pseudonym 2. Stufe

- 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

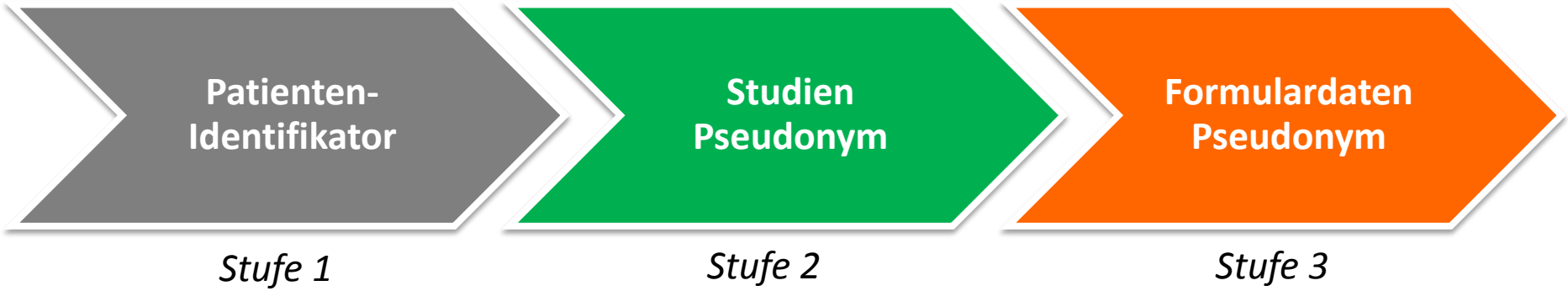
Anwendungsbeispiel (1/8)

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

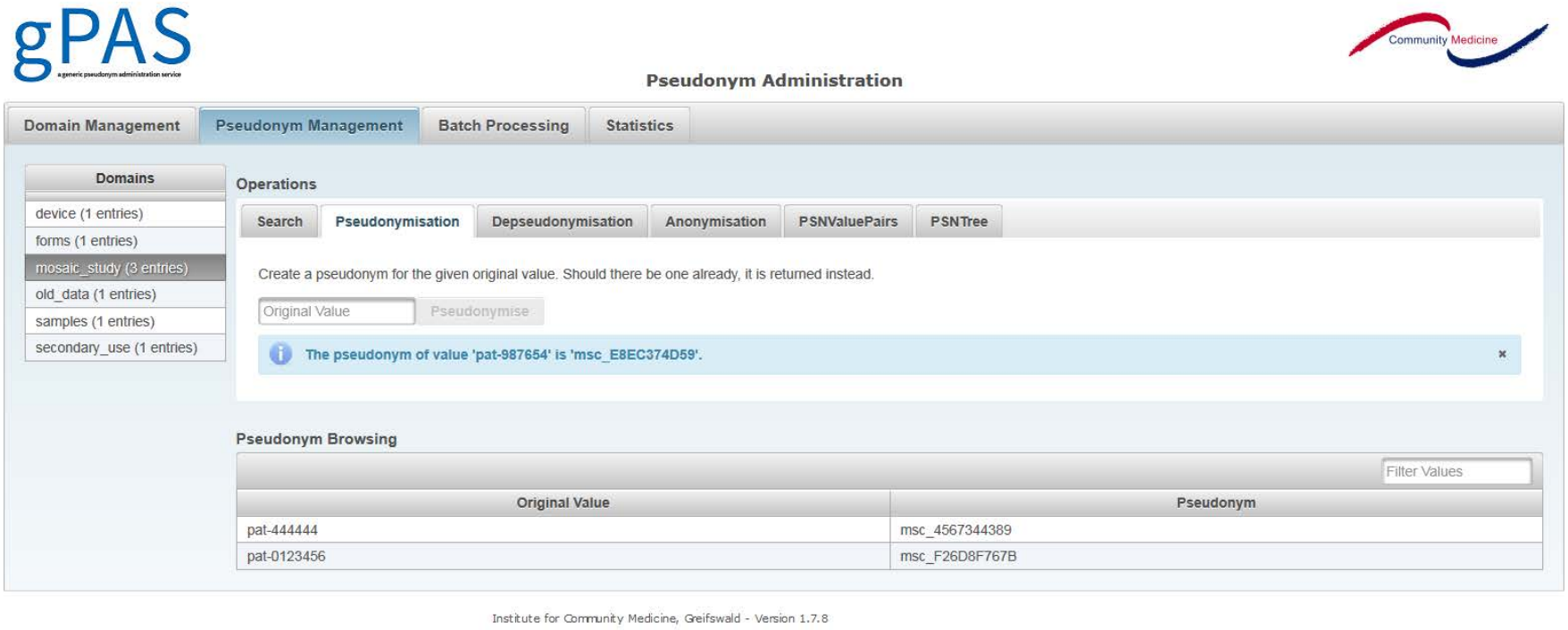
Institute for Community Medicine, Greifswald - Version 1.7.8

Community Medicine

Universitätsmedizin Greifswald

Anwendungsbeispiel (5/8)

Pseudonymisierung und De-Pseudonymisierung

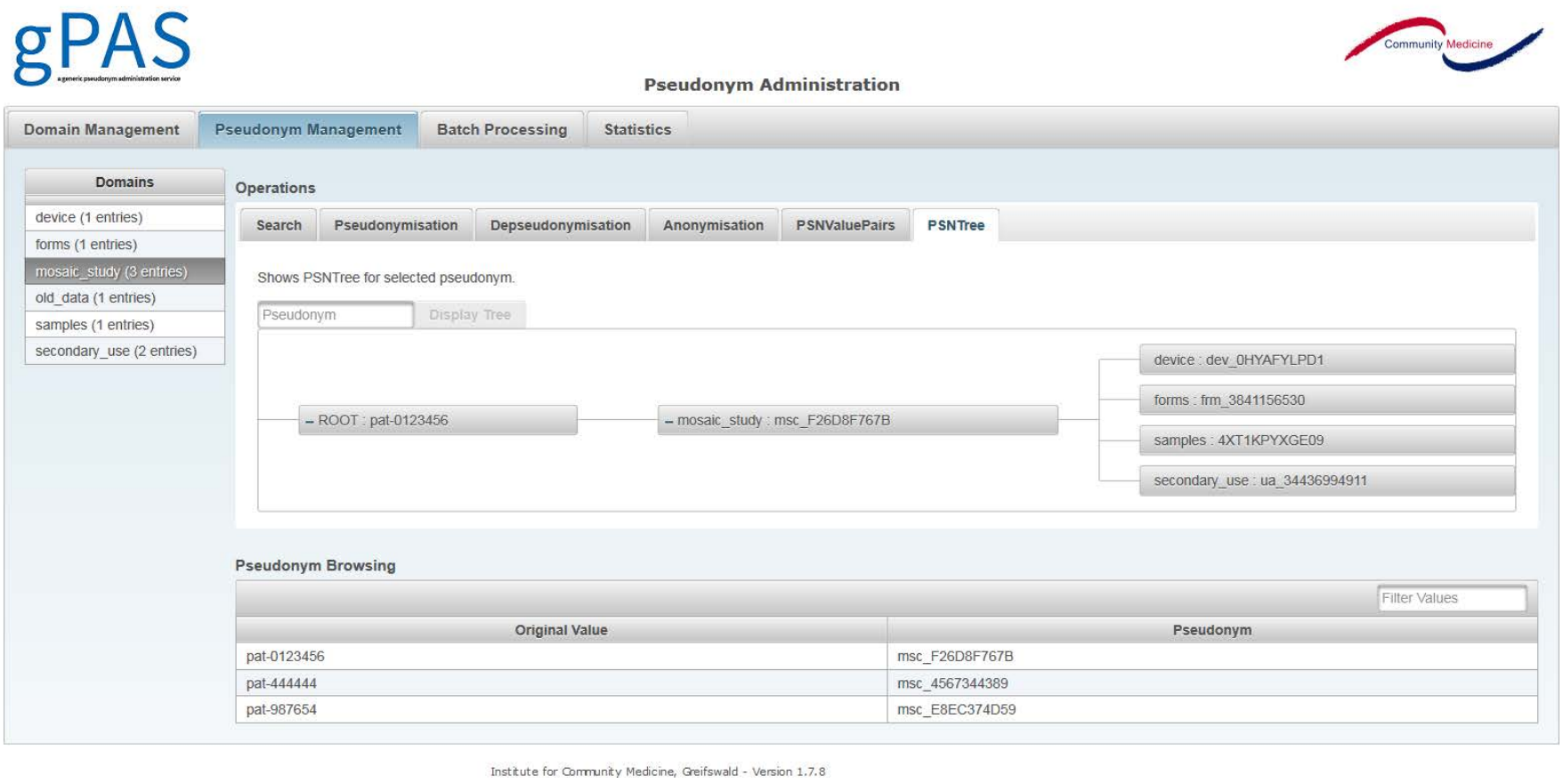


Community Medicine

Universitätsmedizin Greifswald

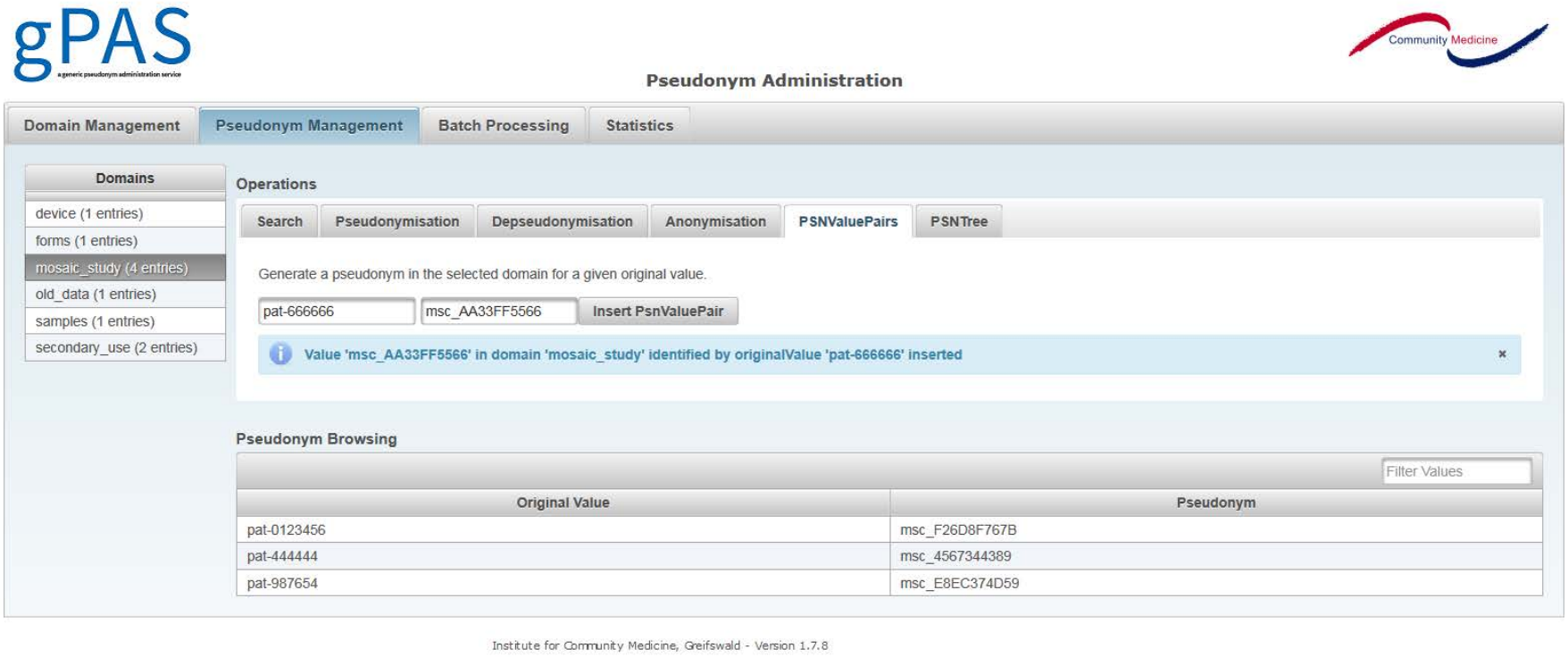
Anwendungsbeispiel (6/8)

Hierarchische Darstellung



Anwendungsbeispiel (7/8)

Integration vorhandener Pseudonyme

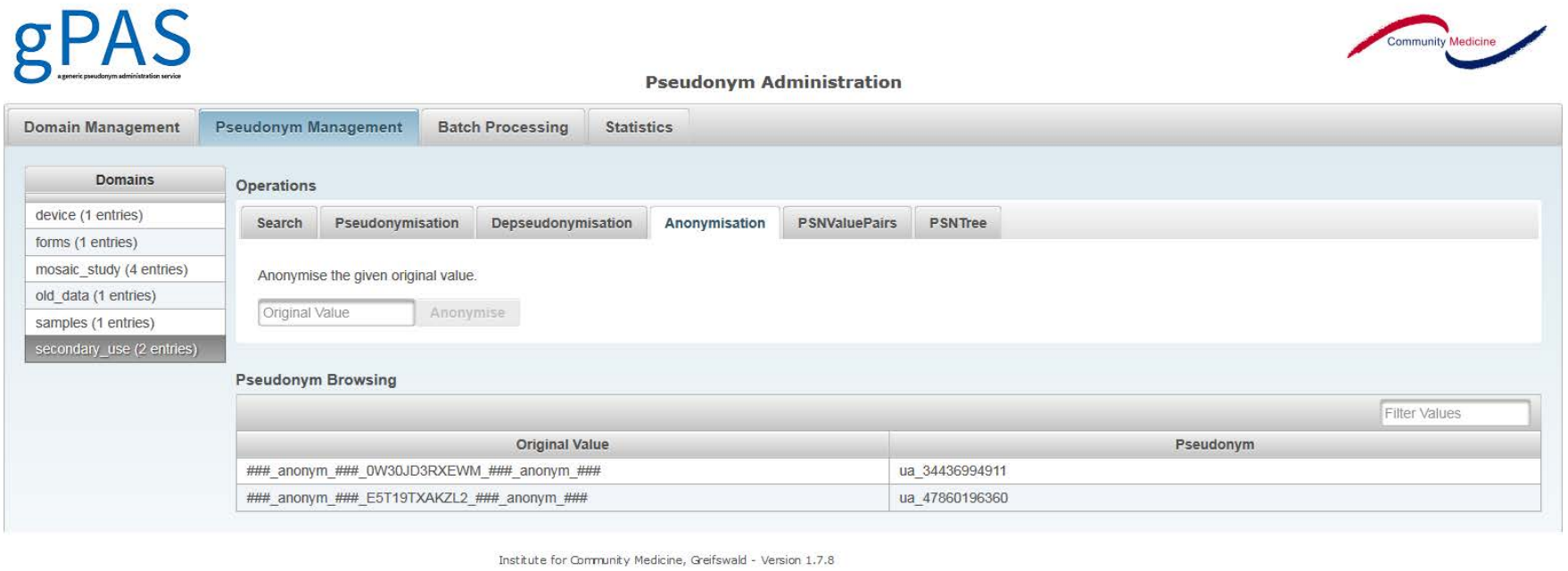


Community Medicine

Universitätsmedizin Greifswald

Anwendungsbeispiel (8/8)

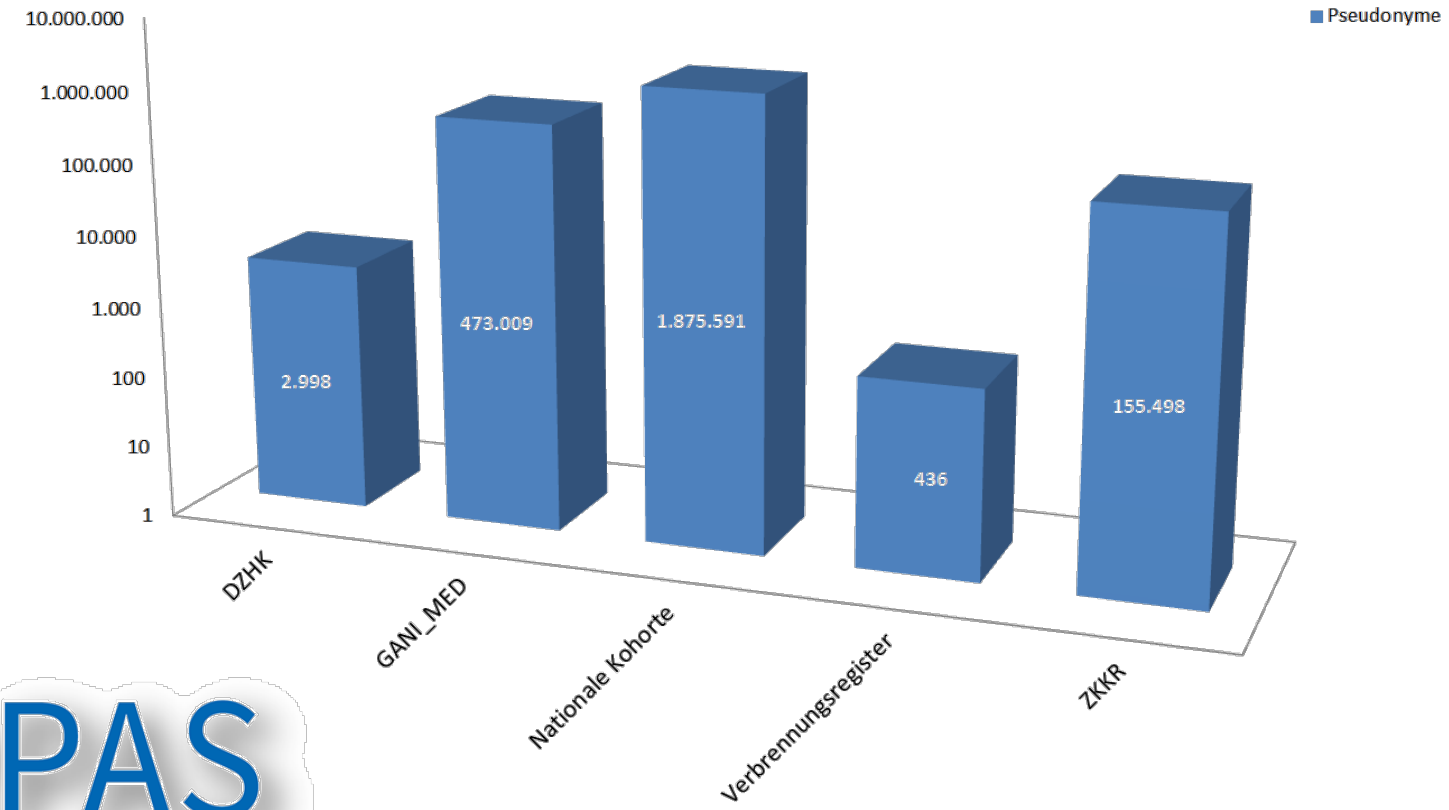
Irreversible Löschung von Zuordnungen (Anonymisierung)



Community Medicine

Universitätsmedizin Greifswald

Blick in die Praxis



gPAS
a generic pseudonym administration service

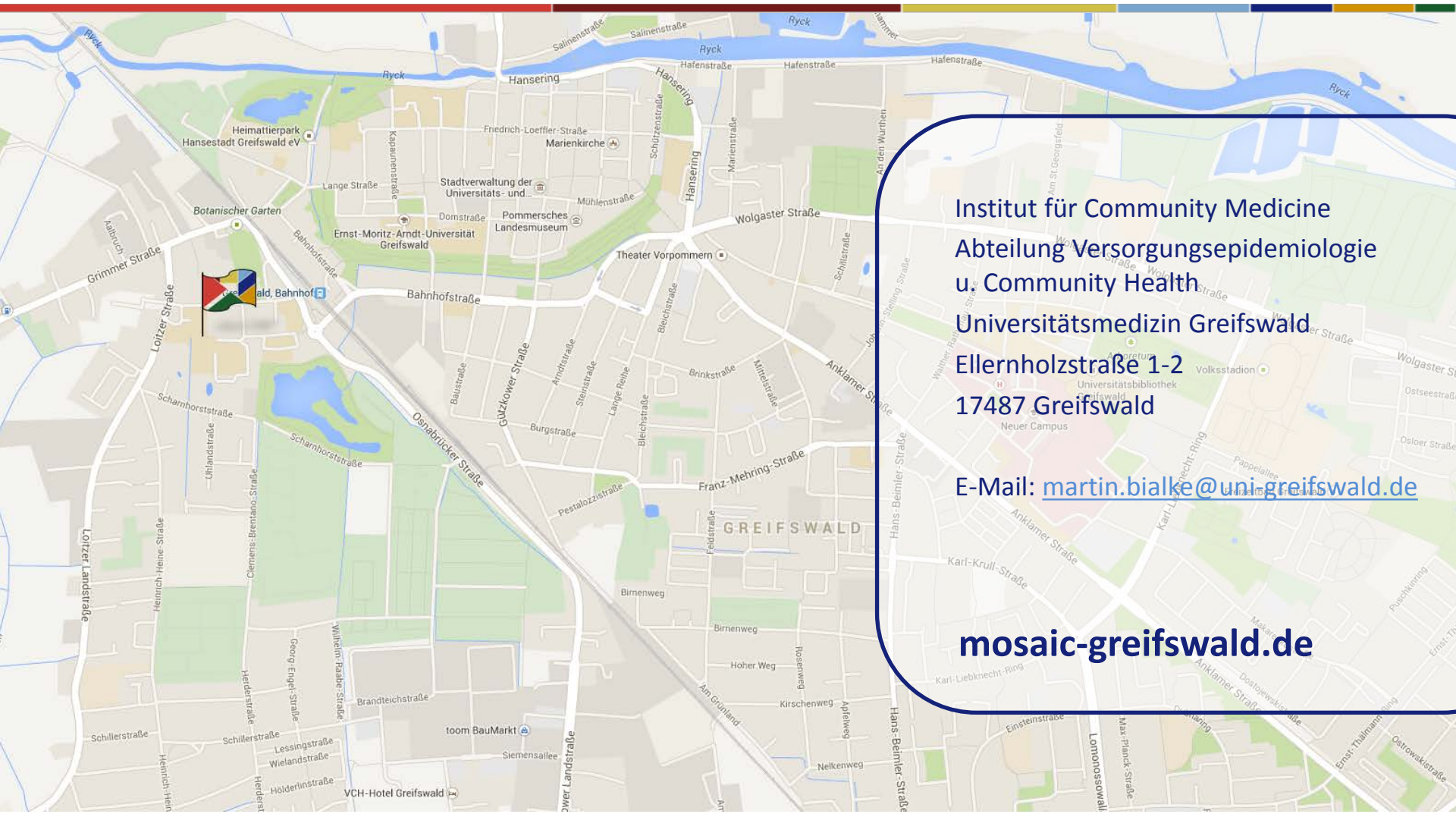
2.507.532 Pseudonyme

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