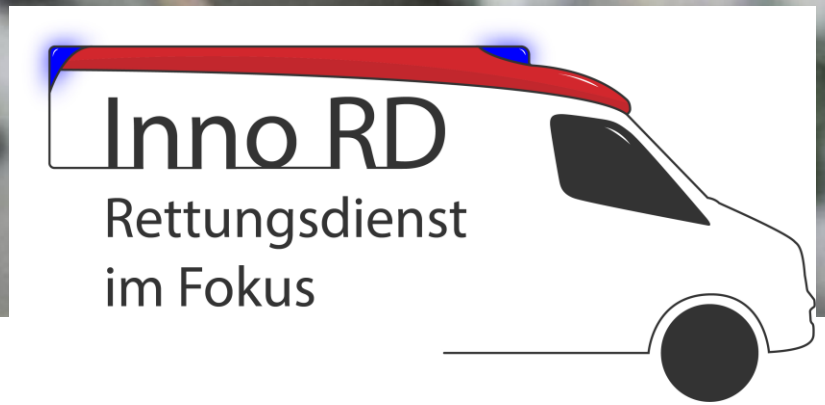
Fakultät VI – Medizin und Gesundheitswissenschaften
Departement für Versorgungsforschung
Abteilung Medizinische Informatik



Pseudonymisierungs-Infrastruktur im Projekt Inno_RD
Sitzung der TMF-AG Datenschutz 01.02.2019

Hauke Fischer

Projekte

ENQuIRE

- Notaufnahme - daten
- Krankenkassen- daten
- Patienten- befragung

- Rettungsdienst - daten
- Krankenkassen- daten
- Befragungsdaten

NDEED

- Notaufnahmedaten
- Abrechnungsdaten der Kassenärztlichen Vereinigungen

CARL VON OSSIETZKY universität OLDENBURG

OFFIS INSTITUT FÜR INFORMATIK

Das Projekt Inno_RD

- Zielsetzung:
  - Verbesserung der Versorgung von Rettungsdienstpatienten in Deutschland
  - Verknüpfung von Daten zur ambulanten, stationären und rettungsdienstlichen Versorgung
- Herangehensweise:
  - Untersuchung des gesamten Behandlungsverlaufes
  - Verknüpfung von Rettungsdienstdaten mit Krankenkassendaten und Befragungsdaten
  - Zusätzliche Informationen: http://rettungsdienst-im-fokus.ovgu.de/
  - Förderkennzeichen: 01VSF17032

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

UNIVERSITÄTSKLINIKUM MAGDEBURG A.Ö.R.

BKK Dachverband

Deutsches Rotes Kreuz

Rahmenbedingungen

- Teilnehmer:
  - 11 Betriebskrankenkassen (Audi BKK, BKK Pfalz, BKK VerbundPlus, BKK ZF & Partner, BMW BKK, Bosch BKK, Daimler BKK, Die Schwenninger Krankenkasse, mhplus Betriebskrankenkasse, pronova BKK, Siemens-Betriebskrankenkasse) [Datenlieferer]
  - Bayerisches Rotes Kreuz [Datenlieferer]
  - Rettungsdienst Heidenheim-Ulm gGmbh [Datenlieferer]
  - Otto-von-Guericke-Universität Magdeburg [Auswertestelle]
- Vorliegend: Sekundärdaten mit einem eindeutigen Schlüssel, der jeweiligen Krankenversichertennummer des Patienten

Problemstellung

- Grundanforderung: Übertragung der Daten zur Auswertestelle unter Nutzung eines zentralen Pseudonymisierungsdienstes
- Herausforderung: Viele Teilnehmer mit restriktiver IT – Aufbau von komplexen Schnittstellen und eingehenden Verbindungen schwierig
- Deshalb: Lokal ausführbare Pseudonymisierung auf Basis von CSV-Dateien mit Minimal-Schnittstelle
- Addons:
  - Verschlüsselung von Dateien
  - Proxy-Verbindungen
  - Datei-Versand per FTP
  - Plausibilitätsprüfung

Pseudonymisierungsansatz I

- Daten werden zerlegt in identifizierende Daten (IDAT) und medizinische Daten (MDAT)
- Einsatz eines unabhängigen Pseudonymisierungsdienstes (PSD)
- Zweistufige Pseudonymisierung:
  - Aus IDAT wird lokal ein permanenter Identifikator (PID) berechnet, so dass der PSD keinen Zugang zu den IDAT erhält.
  - Der PID wird beim Pseudonymisierungsdienst hinterlegt und für ihn ein permanentes Pseudonym (PSN) erstellt, welches für die Datenzusammenführung bei der Auswertstelle verwendet wird.

Pseudonymisierungsansatz II

- MDAT werden in dem Prozess vom Pseudonymisierungsdienst nicht verarbeitet und nur verschlüsselt an ihm vorbei gereicht
- Zweistufiger Ansatz verhindert Rückverfolgbarkeit von Pseudonym zu IDAT bzw. kapselt dies in der Zuordnungsliste beim PSD

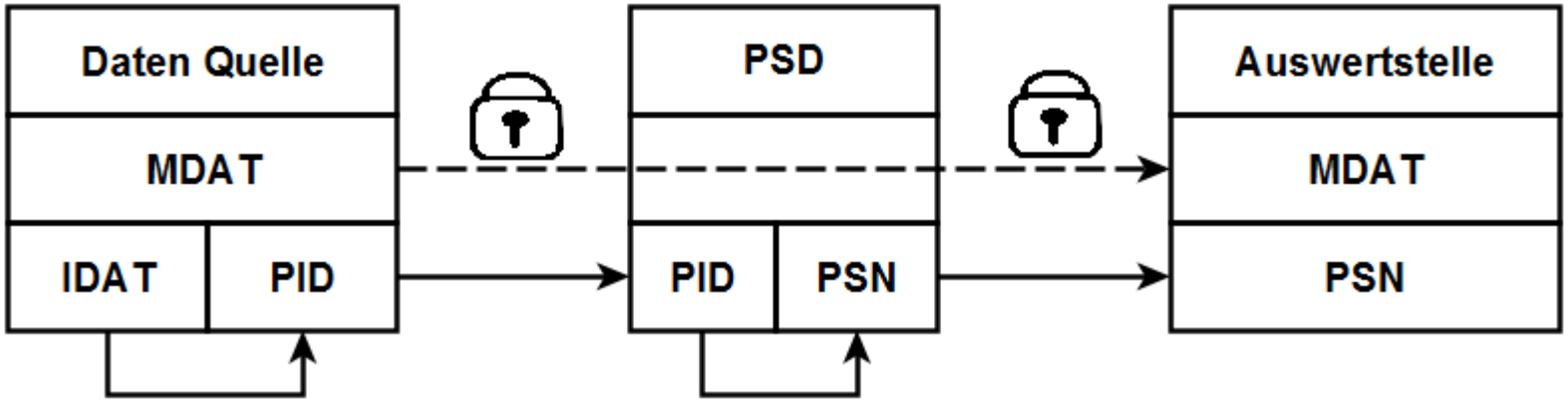


Abbildung in Anlehnung an : Pommerening K, Müller T (2014) Leitfaden zum Datenschutz in medizinischen Forschungsprojekten: Generische Lösungen der TMF 2.0, MWV Med. Wiss. Verl.-Ges, Berlin.

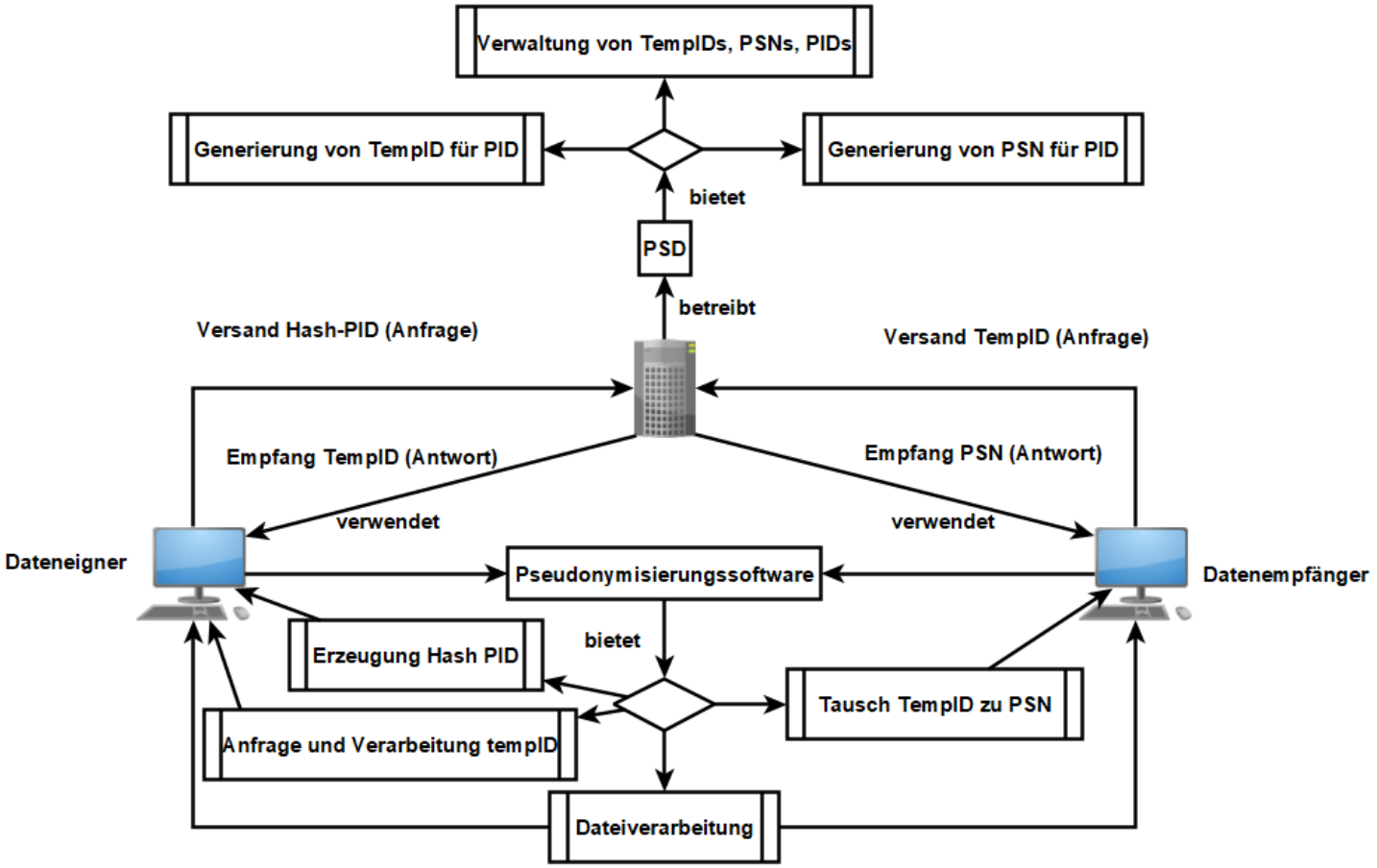
Technische Umsetzung I

- Client-Server-Architektur
- Client ist eine Desktop-Anwendung und nimmt erste Stufe der Pseudonymisierung vor
  - IDAT bleibt während des ganzen Prozess lokal beim Dateneigner
  - Nur Datenquelle hat Zuordnung von IDAT zu PID
- Server nimmt als Pseudonymisierungsdienst die zweite Stufe der Pseudonymisierung vor
  - Nur Pseudonymisierungsdienst hat Zuordnung von PID zu PSN
  - Zusätzliche Vergabe von temporären IDs („Tickets“) für den Datentransfer
  - Temporäre IDs dienen zur Kapselung der PID vom Datenempfänger und der PSN von der Datenquelle

Technische Umsetzung II

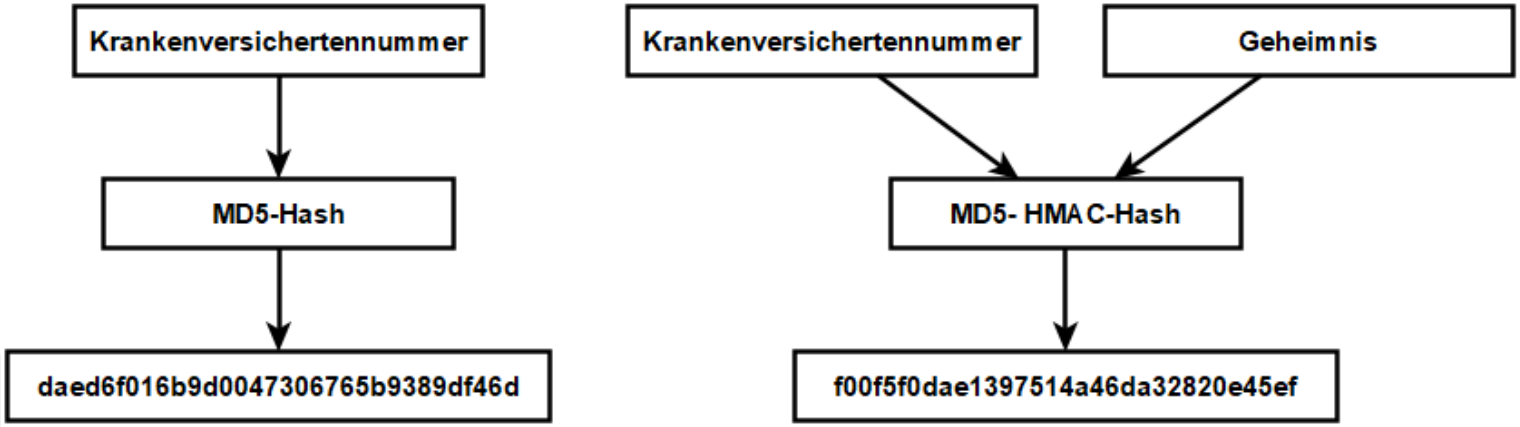
- Client und Server in Java implementiert
- Server-Anwendung läuft auf gängigen Java-Applikations-Servern (Hier: Tomcat 9.xx)
- Datenbank: PostgreSQL
- Alle benötigten Programme für den Betrieb des Dienstes sind frei verfügbar.
- Läuft auf Windows und Linux
  - Aktueller Produktiv-Server unter RedHat (virtueller Server) im Einsatz
  - Entwicklung unter Windows

Infrastruktur



PID-Erzeugung

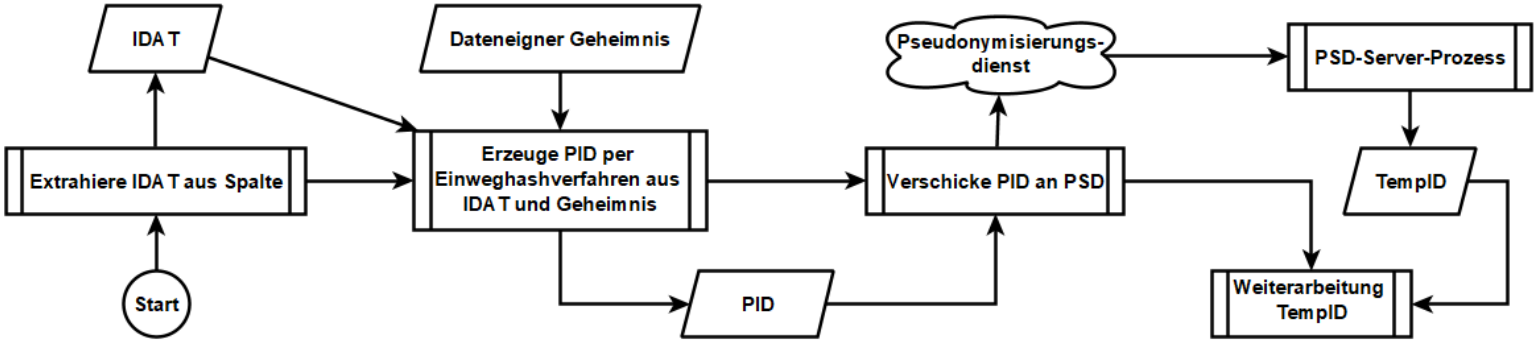
- Hashverfahren mit Pepper-Zusatz
  - Hashbildung auf Basis von zwei Zeichenketten
  - Pepper-Zusatz Geheimnis, welches von den Datenquellen vereinbart wurde und Pseudonymisierungsdienst und der Auswertungsstelle nicht bekannt ist
  - Lokale Überprüfung des Geheimnisses durch Prüfsummenverfahren (Vermeidung von Tipp-Fehlern)



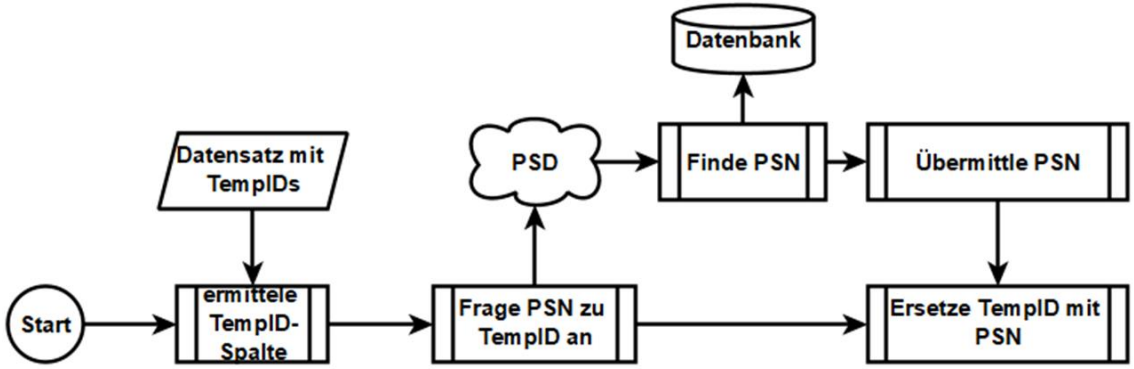
Hauptprozesse der Client-Anwendung

Inno RD
Rettungsdienst
im Fokus

- Pseudonymisierung erste Stufe



- Pseudonymisierung zweite Stufe

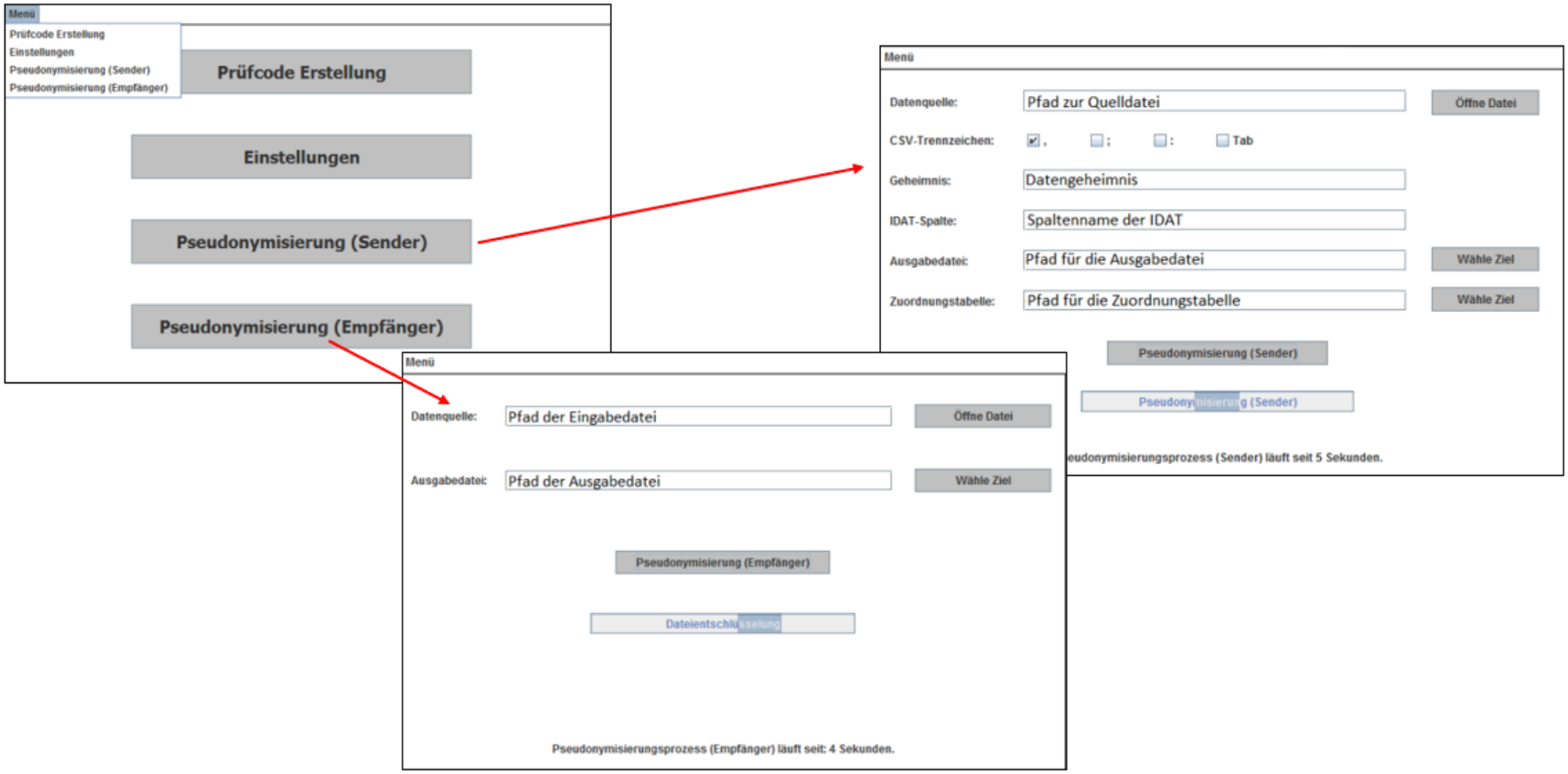


Sicherheit und Datenqualitätssicherung

- MDAT werden für den Versand verschlüsselt
  - Hybride Verschlüsselung mit symmetrischen 256 Bit AES-Schlüssel und 2048 Bit RSA-Schlüssel
- Die Kommunikation von Client und Server erfolgt mittels Transportverschlüsselung über das HTTP-Protokoll (TLS-v1.2)
- Bei der Kommunikation wird die Datenintegrität mit Prüfsummen sichergestellt.
- Erstellung eines Prozess-Logs für den lokalen Pseudonymisierungs-prozess
- IDAT-Spalte kann aktuell geprüft werden:
  - Luhn-Algorithmus (Krankenversichertennummer)
  - Reguläre Ausdrücke (In Inno_RD: Überprüfung ob ein Spaltenfeld „leer“ ist“)
  - Verknüpfung der Prüfungen auf „UND“ oder „ODER“-Basis möglich

Grafische Benutzerschnittstelle

Inno RD
Rettungsdienst
im Fokus



Fazit

- Hier verwendeter Pseudonymisierungsprozess sehr einfach umsetzbar bei der Existenz eines projektweit eindeutigen Identifikators
- Fokus auf möglichst breite Kontrolle der Datenlieferer über die Daten und den lokalen Prozess (Datensparsamkeit)
- Pseudonymisierungsdienst arbeitet passiv mit minimalen Informationen
- Generischer Ansatz, welcher leicht bei ähnlich gelagerten Projekten eingesetzt werden kann.
- Fehlt eindeutiger Identifikator müssen aufwendigere und komplexere Methoden für die Zusammenführung der Daten verwendet werden.

Inno RD
Rettungsdienst
im Fokus

Danke für die Aufmerksamkeit!