Pseudonymisierungskonzepte der TMF

Klaus Pommerening

AG DGI, 12. April 2007

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

BDSG §3 (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.*

- Pseudonymisierung ist rechtlich *nicht* äquivalent zur Anonymisierung,
  - sondern erfordert Zusatzüberlegungen und -maßnahmen;
  - z. B. nur mit Einwilligung oder gesetzlicher Regelung erlaubt!

  weil die Bestimmung des Betroffenen möglich bleibt (auch bei Einweg-Pseudonymen).

Pseudonyme

Je nach Kontext zu nutzen:

1. Inhaber-erzeugte Pseudonyme, die nur vom Inhaber zugeordnet oder aufgelöst werden können.
2. TTP-erzeugte Pseudonyme
   a) Einweg-Pseudonyme, die nicht aufgelöst werden können,
   b) reversible Pseudonyme, die eine Depseudonymisierung ermöglichen.

Grundtyp 1 von Pseudonymen

Inhaber-erzeugte Pseudonyme (Chaum ca. 1980)

- Kontrolle beim Besitzer.
- Für Sekundärnutzung von Gesundheitsdaten nicht geeignet.
- Rechtssicherheit bei Erzeugung durch blinde digitale Signatur.
  - Geeignet für E-Commerce.
  - Lüftbar im Betrugsfall.

  → Rechtssicherer pseudonymer Handel.

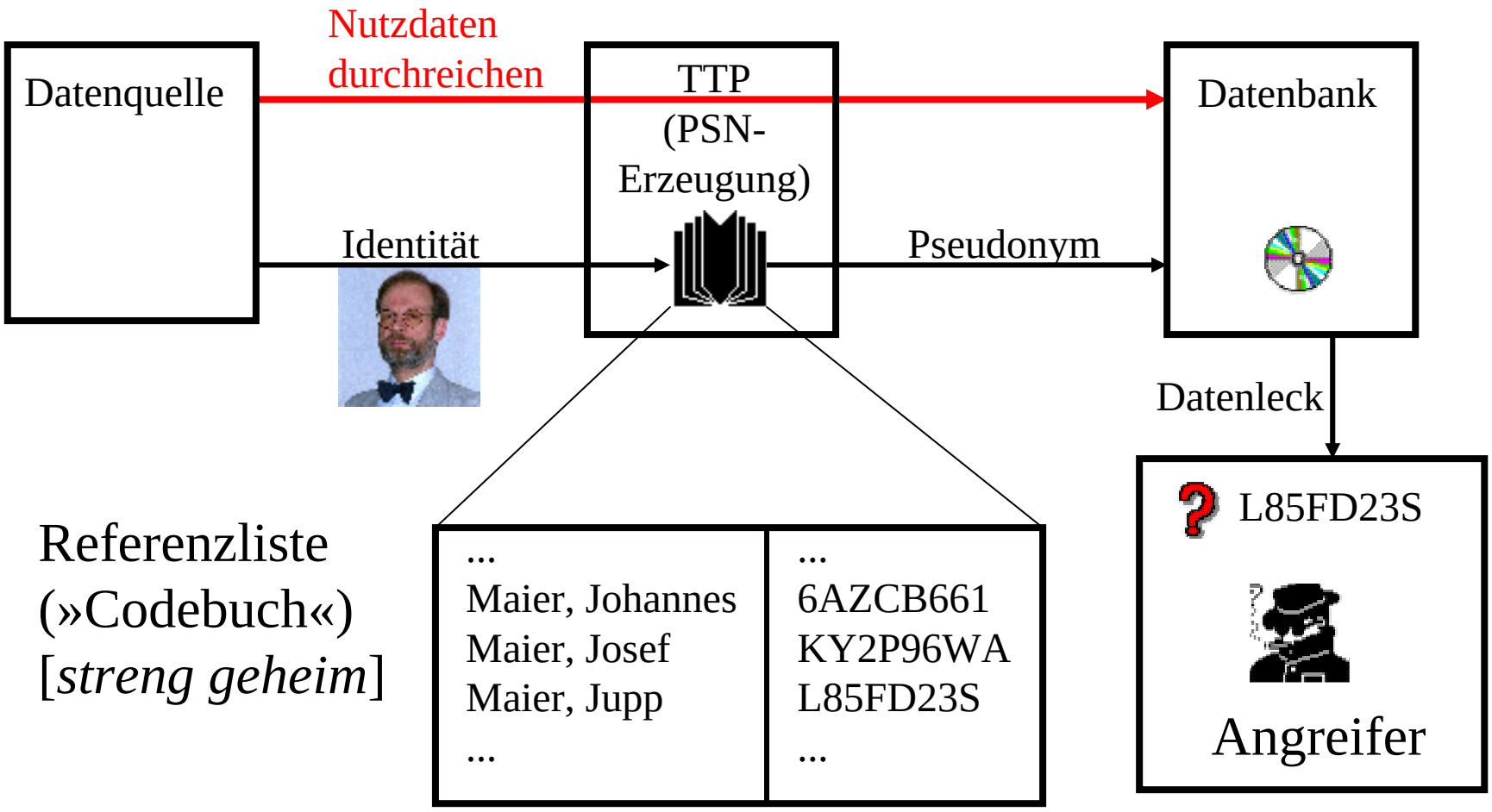
  → Anonyme Berechtigungsnachweise.
- Problem: Politisch nicht gewollt.

Grundtyp 2 von Pseudonymen

TTP-erzeugte Pseudonyme

- Trusted Third Party = »Vertrauensstelle« oder »Datentreuhänder« (z. B. ein Notar).
- Beispiele:
  - Krebsregister (Michaelis/Pomm. 1993),
  - QuaSiNiere,
  - KN Parkinson.
- Für Sekundärnutzung von Gesundheitsdaten besser geeignet:
  - z. B. Rückmeldung über behandelnden Arzt,
  - z. B. Rekrutierung für Studien.

Grundtyp 2: Das Basismodell



Alternative: Schlüssel statt Referenzliste

- Pseudonym-Erzeugung durch kryptographische Verschlüsselung;
  - garantierte Eindeutigkeit: Pseudonym = verschlüsselter Personenidentifikator (PID).
  - D. h., es muss ein wirksames Identitätsmanagement vorhanden sein, das für eindeutige PIDs sorgt.
  - ID-Management und PSN-Erzeugung trennbar.
- Die Zentralstelle speichert nichts außer ihrem geheimen Schlüssel (z. B. auf SmartCard).
  - »Schlanker« TTP-Service.
- Auch irreversible Pseudonyme möglich (durch Einweg-Verschlüsselung).
- Bei Wechsel des kryptographischen Verfahrens:
  - Um- oder Überverschlüsselung.

Auflösung von Pseudonymen

- Depseudonymisierung:
  - befugte Zuordnung zur Identität.
  - Typische Datentreuhänder-Funktion.
- Rückidentifizierung:
  - unbefugte Zuordnung zur Identität.
  - Methoden: Ausspähen von Geheimnissen, statistische Inferenz aus Datenbanken und Zusatzwissen.
- Ziel aller Pseudonymisierungslösungen:

Kontrolle des Rückidentifizierungsrisikos

Achtung: Initialen + Geburtsdatum als PSD nicht geeignet.

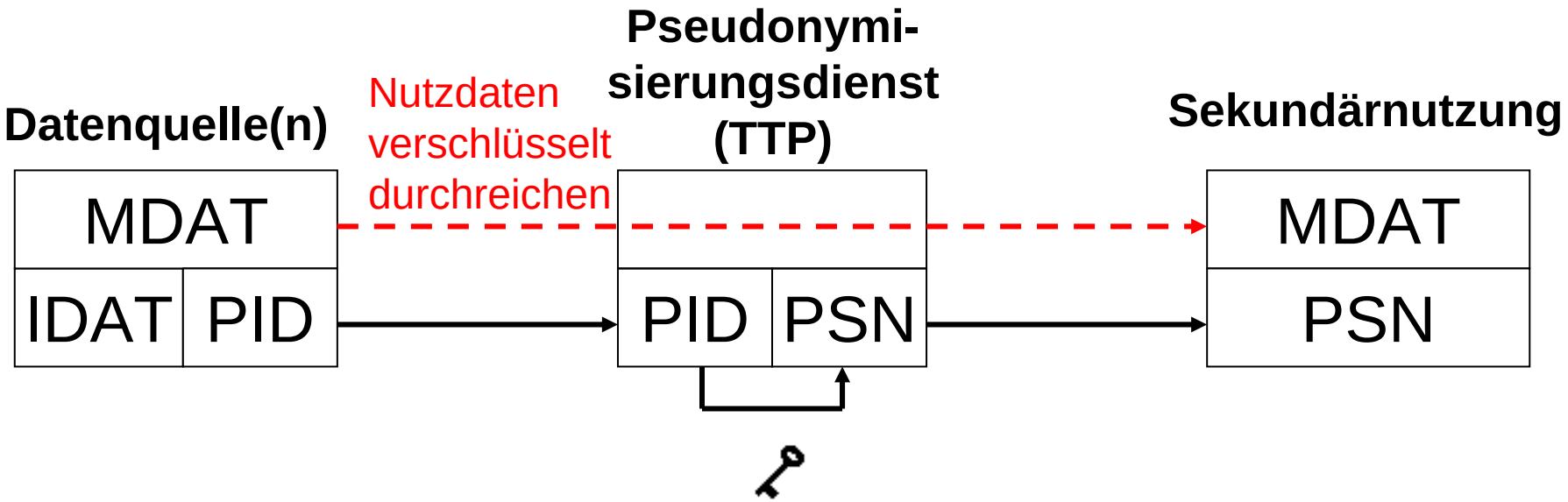
Szenario 1: Einzelne Datenquelle, Einmal-Sekundärnutzung

- (Eigentlich) typischer Anwendungsfall für Anonymisierung.
  - Beispiel: Einfache statistische Auswertung exportierter Datensätze.
- AMG §40 (2a): ... *die erhobenen Daten soweit erforderlich ...* ***pseudonymisiert*** *an den Sponsor oder eine von diesem beauftragte Stelle zum Zwecke der wissenschaftlichen Auswertung weitergegeben werden ...*
  - TMF-Rechtsgutachten steht noch aus.
  - Probleme bei wissenschaftsgetriebenen Studien.

Szenario 2: Mehrere Datenquellen mit Überschneidungen, Einmal-Sekundärnutzung

- Daten aus verschiedenen Quellen müssen zusammengeführt werden.
- Beispiele:
  - multizentrische Studien,
  - Follow-up-Daten in Registern,
  - Qualitätsstudien in der Versorgungsforschung.
- Typischer Anwendungsfall für Einweg-Pseudonyme.
  - Depseudonymisierung (fast) ausgeschlossen,
  - Verknüpfbarkeit der pseudonymen Daten untereinander bleibt.

Pseudonymisierung für Einmal-Sekundärnutzung



MDAT = Medizinische Daten
IDAT = Identitätsdaten

PID = Eindeutiger Patientenidentifikator
PSN = Pseudonym

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Besonderheiten von Szenario 2

- Medizinische Daten (MDAT) mit öffentlichem Schlüssel des Sekundärnutzers verschlüsselt –
  - Die TTP kann die MDAT nicht lesen.
  - Nur der Sekundärnutzer kann sie entschlüsseln.
- Das Pseudonym (PSN) ist der verschlüsselte PID
  - mit einem geheimen Schlüssel, den nur die TTP hat,
  - durch eine Einweg-Funktion.
- Die TTP speichert nichts (außer dem Schlüssel).
- Szenario 2 in Routinebetrieb seit 2002 in einem Projekt der Versorgungsforschung der TMF.

Szenario 3: Einmalige Sekundär-Nutzung mit Depseudonymisierungsmöglichkeit

- Verwendet wird das Modell von Szenario 2,
  - aber PSN-Dienst verschlüsselt *umkehrbar*,
  - Depseudonymisierbarkeit bleibt gewahrt.
- Identitätsmanagement (evtl. extern):
  - Eine „Patientenliste“ speichert die Zuordnung zwischen IDAT und PID.
- Die Depseudonymisierung läuft über PSN-Dienst und Patientenliste.
- Zwei TTPs involviert (je nach Verhältnismäßigkeit).

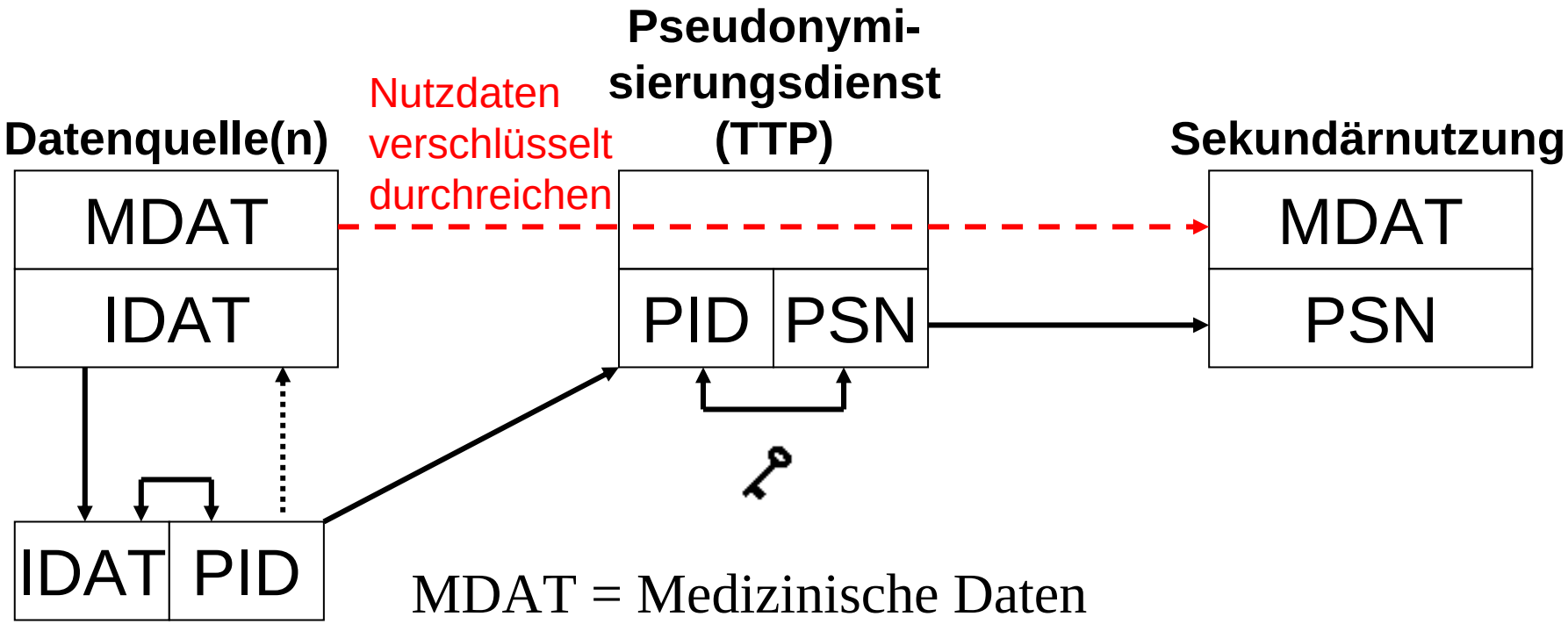
Nutzung der KV-Nummer als PID?

- Eindeutige Kennzeichnung auf KV-Karte bzw. eGK.
- Ohne gesetzliche Regelung nutzbar für
  - Projekte der Versorgungsforschung?
  - Forschungsnetze?
  - Vielleicht zur Einweg-Verschlüsselung?
- Probleme:
  - Kein Geheimnis (im Gegensatz zum PID, der manchmal als „Pseudonym erster Stufe“ genutzt wird, siehe Modell A).
  - Nicht langjährig invariant.
  - Nicht bei allen Datenquellen verfügbar.

Modell B der TMF: Pseudonymer Forschungs-Datenpool

- Datenfluss wie in Szenario 3,
  - aber der Sekundärnutzer baut einen (krankheitsspezifischen) Datenpool (z. B. Kohorte, Register, BMB) auf.
- Die Langzeit-Datensammlung erfordert
  - klar definierten organisatorischen Rahmen,
  - besondere technische Sicherheitsvorkehrungen,
  - entsprechende Patientenaufklärung und -einwilligung.
  - Ferner: ausgefeiltes Qualitätssicherungsmanagement.
- Identitätsmanagement (PID-Verwaltung) und Qualitätssicherung der Daten (mit Rückfragen) müssen vor Pseudonymisierung erfolgen.

Pseudonymisierung mit möglicher Depseudonymisierung



MDAT = Medizinische Daten
IDAT = Identitätsdaten

PID = Patientenidentifikator
PSN = Pseudonym

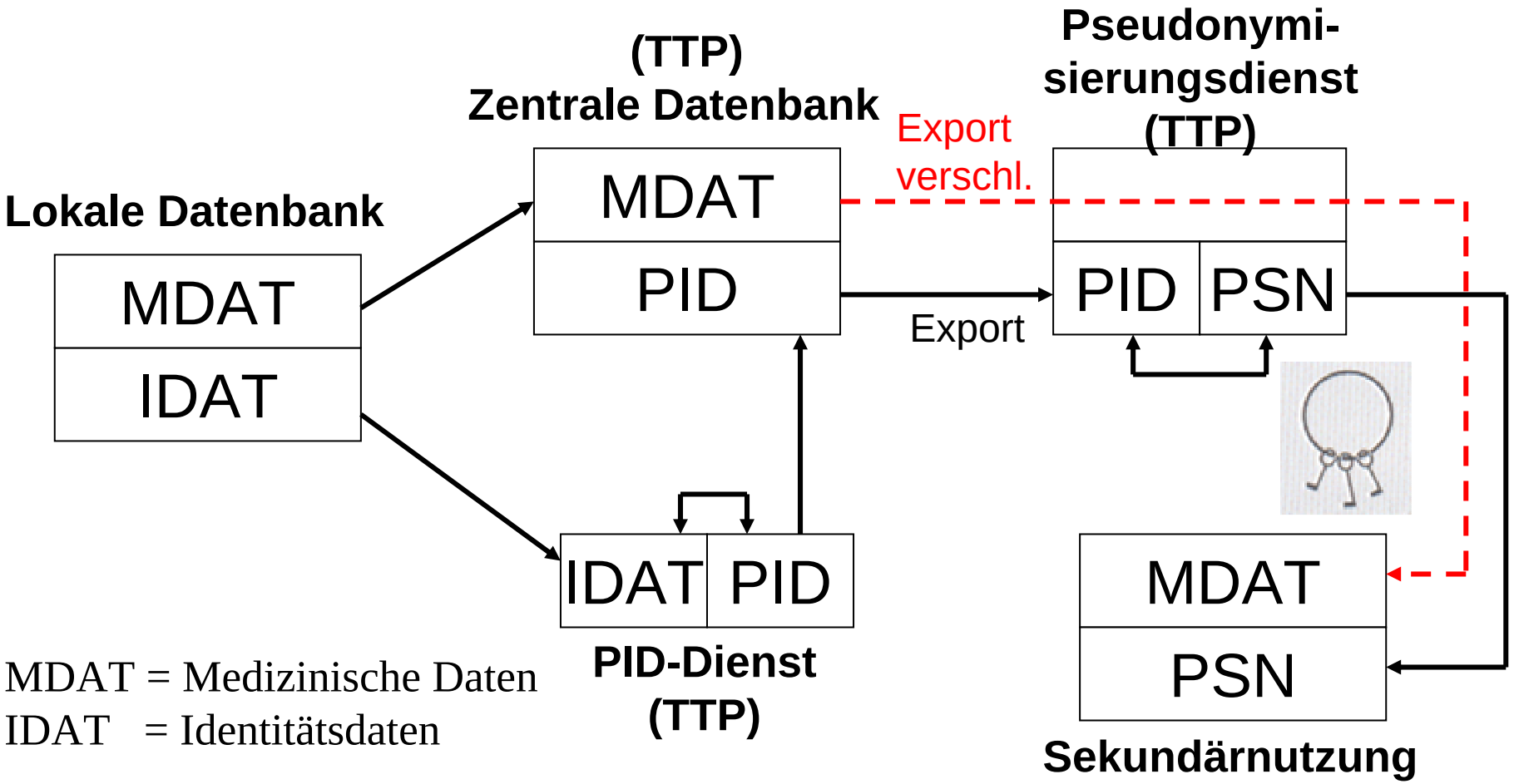
JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Modell A der TMF: Zentrale klinische Datenbank, mehrfache Sekundärnutzung

- Datenpool = zentrale »klinische« Datenbank.
  - Zentral für Forschungsverbund.
  - Zugriff für behandelnden Arzt (dezentral).
  - Keine Identitätsdaten, nur PIDs in DB.
  - PID wird als Pseudonym behandelt.
  - Zugriffsregelung über temporäre Token (tempID).
- Kein Online-Zugriff für Sekundärnutzer.
  - Für Sekundärnutzung wird jeweils ein Auszug der Datenbank exportiert (anonymisiert oder *ad hoc* pseudonymisiert).

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Die zentrale klinische Datenbank



MDAT = Medizinische Daten
IDAT = Identitätsdaten

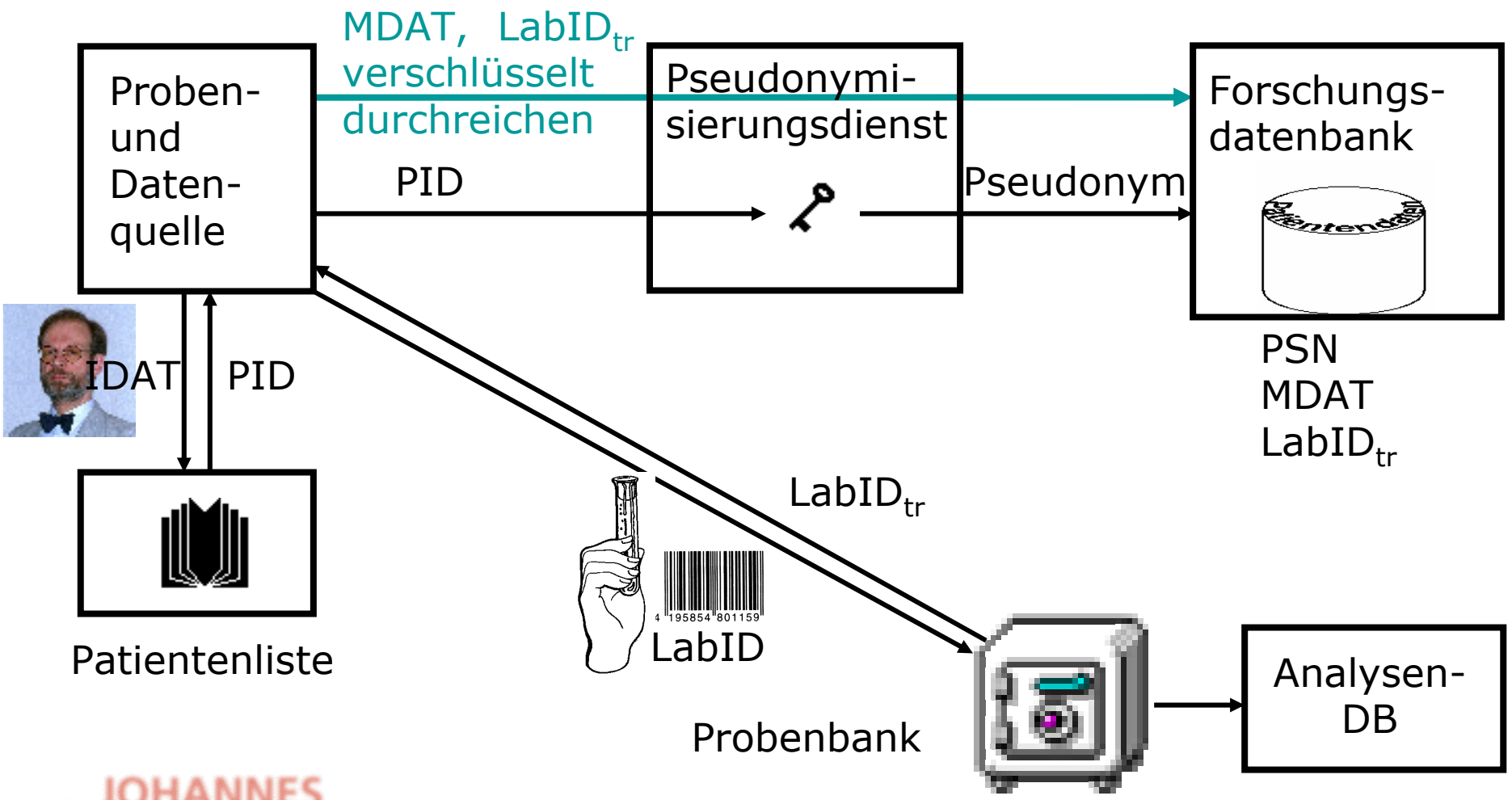
PID = Patientenidentifikator
PSN = Pseudonym

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Besonderheiten von Modell A

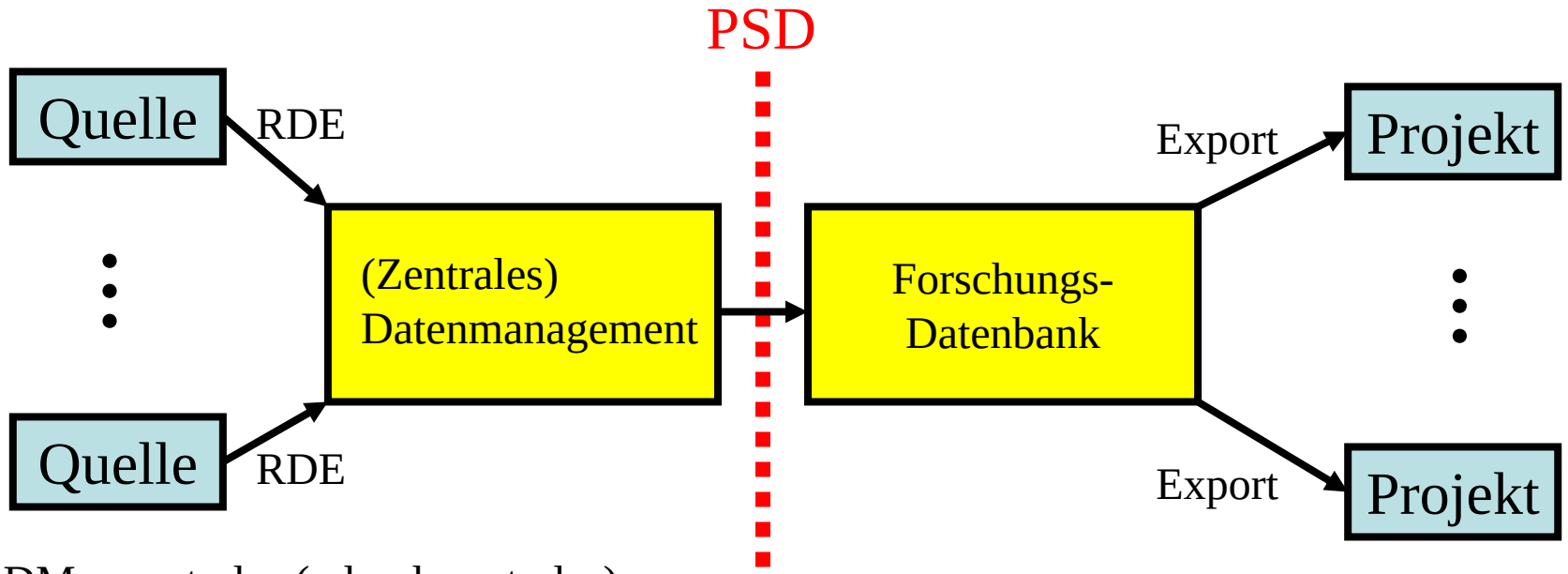
- Vorteile:
  - Gut geeignet für multizentrische Studien.
  - Bessere Unterstützung für Langzeitbeobachtung von Patienten mit chronischer Erkrankung.
  - Nützlich für den datenproduzierenden Arzt.
  - Gut an Patientenakten-Architektur mit *zentraler* DB anpassbar.
- Nachteile:
  - Komplizierte Kommunikationsprozeduren.
  - Viele TTP-Dienste und geheime Schlüssel benötigt.

TMF- Modell für Forschungsnetze mit Biomaterialbank



LabID$_{tr}$ zusätzliches Pseudonym zur Zuordnung von Proben

Künftiges modulares Modell (Vorschlag)



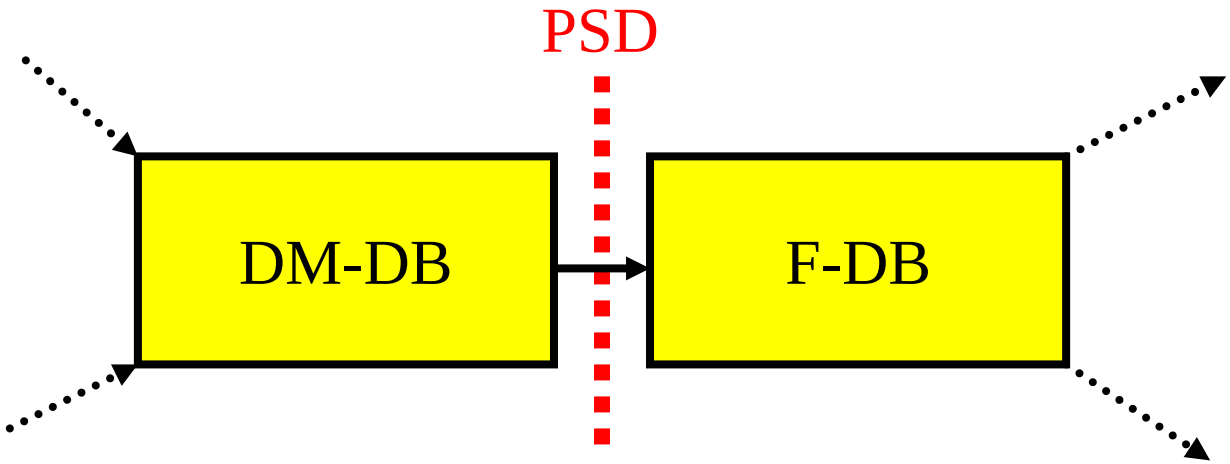
DM = zentrales (oder dezentrales)
Datenmanagement für
verteilte Versorgung
oder für (evtl. mehrere)
(i. d. R. multizentrische) Studien
z. B. bei einem KKS
Auch für Qualitätssicherung

Export an oder Zugriff für
Forschungsprojekte

DM liefert qualitätsgesicherte
Daten an Forschungsdatenbank.
F-DB dient auch zur (pseudonymen)
Archivierung von Studienrohdaten
nach GCP.

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Künftiges modulares Modell (Vorschlag)



DM-DB wird **nach Modell A** geführt, pseudonym, Patientenliste separat, RDE-Zugriff über ad-hoc-Token,

F-DB wird **nach Modell B** geführt, Rückweg zur DM-DB (oder zum Patienten) nur über PSD.

Datenüberführung von DM-DB in F-DB so früh wie möglich.

Künftiges modulares Modell (Vorschlag)

Die bisherigen Modelle sind Spezialfälle.

- „Reines Modell A“: wenn Daten auch langfristig patientennah verfügbar bleiben müssen.
- „Reines Modell B“: wenn keine direkte Rückkopplung an die Behandlung vorgesehen ist.

Weiterentwicklung

- „A-Teil“ Richtung elektronische Patientenakte weiterentwickeln.
  - Integration von Versorgung und Forschung besser abbilden,
  - insbesondere die Konsiliar- und Referenztätigkeiten im Rahmen multizentrischer Therapie-Optimierungsstudien.
- Identitätsmanagement für „A-Teil“ über PID-Generator/Patientenliste mit abgestuften Anforderungen nach Verhältnismäßigkeit durch
  - getrennte Datenbanktabelle,
  - getrennte Datenhoheit (u. U. reichen getrennte Fachabteilungen im Krankenhaus),
  - Datentreuhänder-/ Notardienst.

Wie wird das RI-Risiko kontrolliert?

- Sichere Verschlüsselung bei PSN-Erzeugung.
- Informationelle Gewaltenteilung:
  - Aufbewahrung des PSN-Schlüssels bei TTP,
  - ID-Management bei TTP,
  - Datensatz auf verschiedene Datenbanken verteilt,
  - nur kontrollierter Export statt online-Zugriff.
- Verhältnismäßigkeit: Kriterien z. B.
  - Größe des Netzes (z. B. europaweite vs. lokale BMB),
  - Brisanz des Netzes (z. B. HIV vs. Rheuma),
  - Existenz externer Vergleichsdaten,
  - ...
- SOPs, Verpflichtungen, vertragliche Regelungen.