

# Pseudonymisierung in der medizinischen Forschung und Sekundärnutzung von Patientendaten

*Klaus Pommerening, Mainz*

Tutorium AG DGI, 17. September 2007

GMDS 2007, Augsburg

JOHANNES  
GUTENBERG  
UNIVERSITÄT  
MAINZ

Kompetenznetz  
Pädiatrische Onkologie  
und Hämatologie



TMF = Telematikplattform für die  
medizinischen Forschungsnetze

Gefördert vom



Bundesministerium  
für Bildung  
und Forschung

# Inhalt

- 1. Grundlagen von Anonymisierung und Pseudonymisierung**
2. Methoden und Szenarien
3. Die Pseudonymisierungswerkzeuge der TMF
4. Diskussion und Ausblick



# Nutzung von Patientendaten

- Primärnutzung: Behandlungskontext.
- Sekundärnutzung:
  - Versorgungsforschung, Qualitätssicherung, Gesundheitsökonomie,
  - krankheitsspezifische klinische oder epidemiologische Studien,
  - Aufbau von zentralen Datenpools und Biomaterialbanken.

## Typische Aspekte der Sekundärnutzung:

- Außerhalb des Behandlungskontexts und der Schweigepflicht (des *behandelnden* Arztes);
- die Identität des Patienten ist ohne Belang.



Behandlungskontext



[Primärnutzung]

**Barriere: Ärztliche Schweigepflicht**

[Sekundärnutzung/Forschungskontext]

klinische Forschung  
Versorgungsforschung



direkte  
Erfassung



Export erlaubt, wenn  
- anonyme Daten,  
- Einwilligung,  
- Gesetzesvorschrift

Register/  
epidemiologische Forschung



BDSG §3 (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.

[„faktische Anonymisierung“]

# Für die Sekundärnutzung von Patientendaten (und Proben):

- Identität der Patienten schützen.
  - Informationelle Selbstbestimmung erfordert Einwilligung oder Anonymisierung
- Anonymisierung, wann immer möglich.  
Nachteile der Anonymisierung:
  - Keine Zusammenführung von Daten aus verschiedenen Quellen
  - ... oder von verschiedenen Zeitpunkten.
  - Kein Weg zurück zum Patienten für Rückmeldungen
  - ... oder zur Rekrutierung für neue Studien
  - ... oder zum Rückruf von Proben.

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.

# Pseudonyme

- Goldener Mittelweg zwischen Anonymität und offener Identität („indirekter Personenbezug“).
- Pseudonymisierung ist rechtlich *nicht* äquivalent zur Anonymisierung,
  - sondern erfordert Zusatzüberlegungen und -maßnahmen;
  - z. B. nur mit Einwilligung oder gesetzlicher Regelung erlaubt!
  - Denn Pseudonyme vom Typ 2 (s. u.) sind *personenbeziehbar*.

# Inhalt

1. Grundlagen von Anonymisierung und Pseudonymisierung
- 2. Methoden und Szenarien**
3. Die Pseudonymisierungswerkzeuge der TMF
4. Diskussion und Ausblick

# Grundtyp 1 von Pseudonymen

Inhaber-erzeugte Pseudonyme (Chaum ca. 1980)

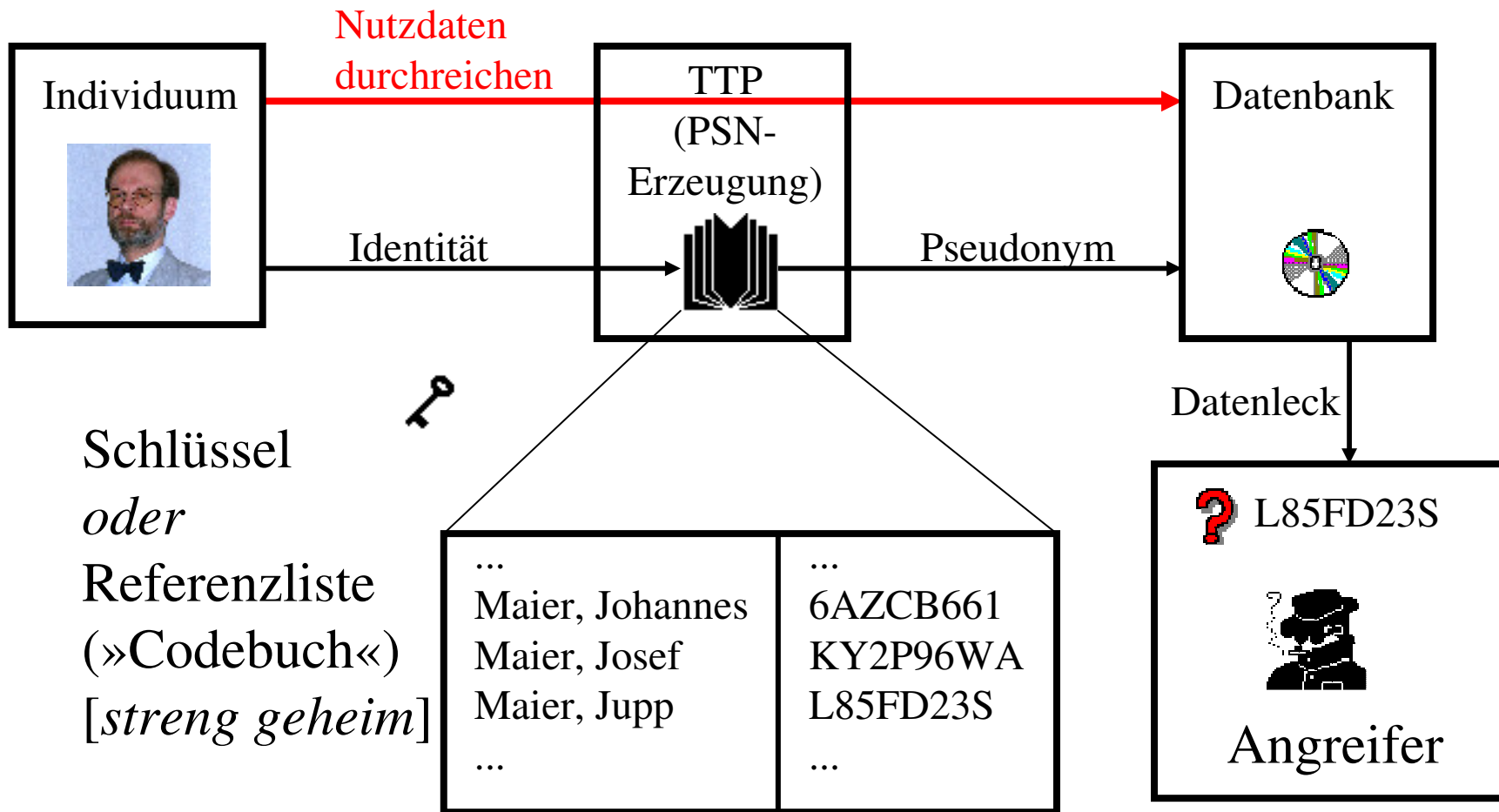
- Erzeugung durch blinde digitale Signatur.
- Kontrolle beim Besitzer.
- Für Sekundärnutzung von Gesundheitsdaten nicht geeignet.
- Geeignet für E-Commerce.
- Lüftbar im Betrugsfall.

# Grundtyp 2 von Pseudonymen

## TTP-erzeugte Pseudonyme

- Trusted Third Party = »Vertrauensstelle« oder »Datentreuhänder« (z. B. ein Notar).
- Beispiel: Krebsregister (Michaelis/Pomm. 1993).
- Für Sekundärnutzung von Patientendaten besser geeignet:
  - z. B. Rückmeldung über behandelnden Arzt,
  - z. B. Rekrutierung für neue Studien.
- Aber: *personenbeziehbar* (durch TTP).

# Grundtyp 2: Das Basismodell



# Besser: Schlüssel statt Referenzliste

- Pseudonym-Erzeugung durch kryptographische Verschlüsselung;
  - garantierte Eindeutigkeit: Pseudonym = verschlüsselter Personenidentifikator (PID).
  - Voraussetzung: Es gibt einen eindeutigen Identifikator
  - D. h., ein Identitätsmanagement ist nötig.
- Die Zentralstelle speichert nichts außer ihrem geheimen Schlüssel (z. B. auf SmartCard).
  - »Schlanker« TTP-Service.
  - Auch irreversible Pseudonyme möglich (durch Einweg-Verschlüsselung).

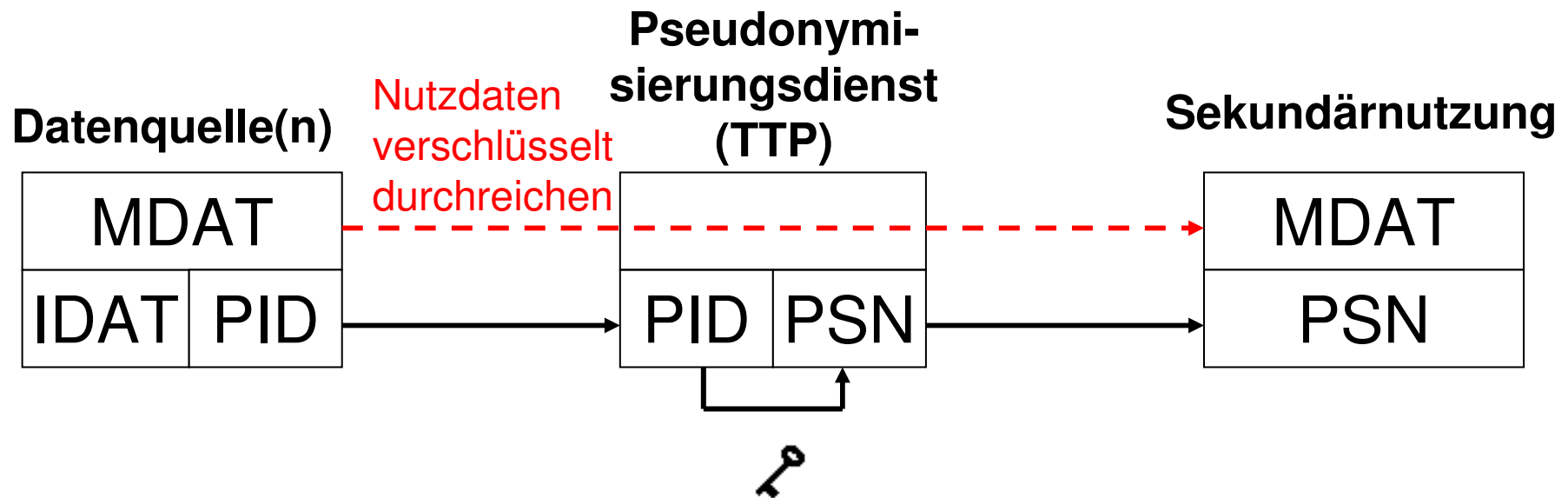
# Szenario 1: Einzelne Datenquelle, Einmal-Sekundärnutzung

- Typischer Anwendungsfall für Anonymisierung.
- Beispiel: Einfache statistische Auswertung exportierter Datensätze.

# Szenario 2: Mehrere Datenquellen mit Überschneidungen, Einmal-Sekundärnutzung

- Daten aus verschiedenen Quellen müssen zusammengeführt werden.
- Beispiele:
  - multizentrische Studie,
  - Follow-up-Daten.
- Typischer Anwendungsfall für Einweg-Pseudonyme.
  - Direkter Identitätsbezug wird aufgehoben, Verknüpfbarkeit bleibt.

# Pseudonymisierung für Einmal-Sekundärnutzung



MDAT = Medizinische Daten  
IDAT = Identitätsdaten

PID = Eindeutiger Patientenidentifikator  
PSN = Pseudonym

# 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.
- Szenario 2 in Routinebetrieb seit 2002 in einem Projekt der Versorgungsforschung der TMF.

# Szenario 3: Sekundär-Nutzung mit Rückidentifikationsmöglichkeit

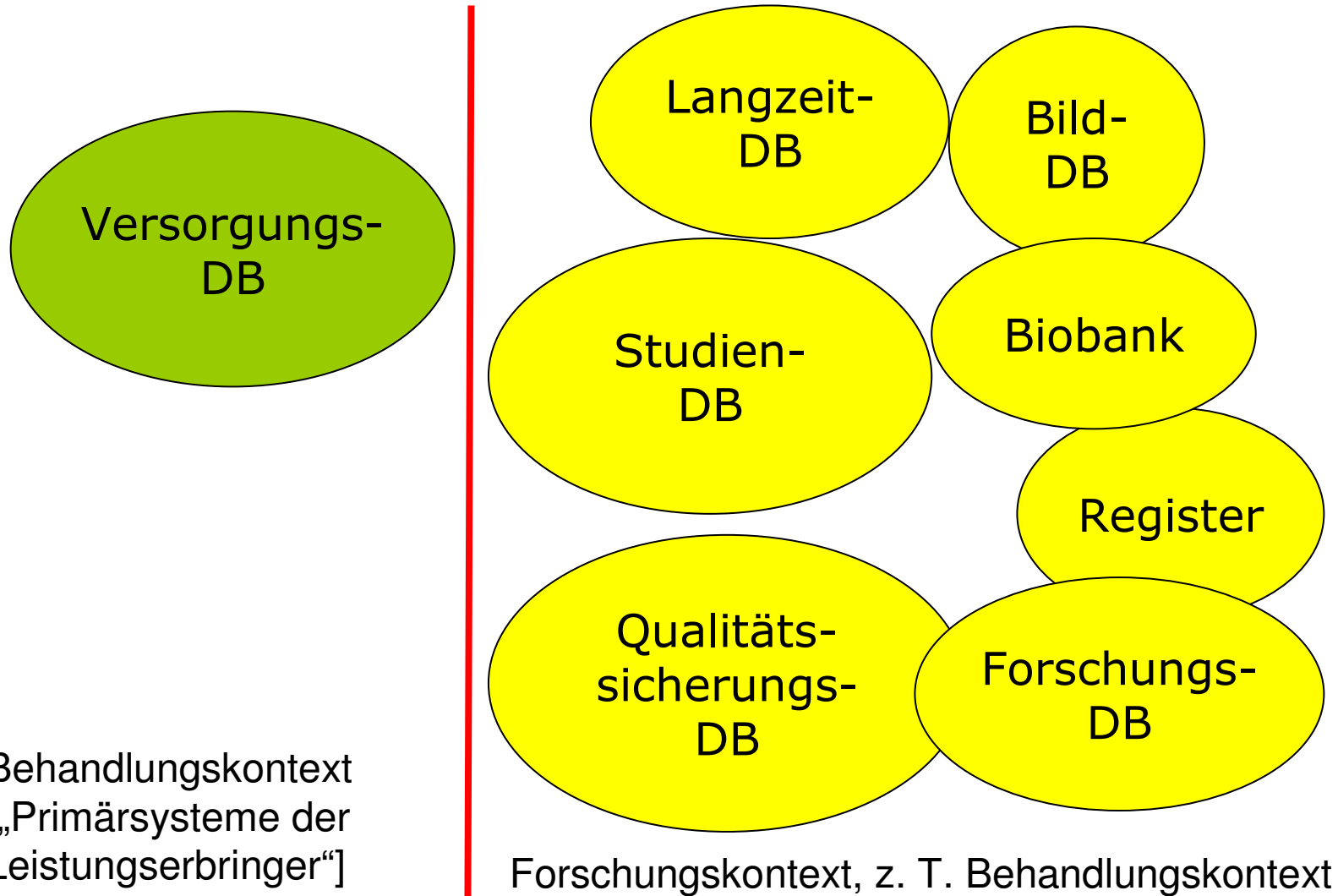
- Verwendet wird das Modell von Szenario 2,
  - aber PSN-Dienst verschlüsselt *umkehrbar*,
  - Rückverknüpfbarkeit bleibt erhalten.
- Evtl. Identitätsmanagement als TTP-Dienst (je nach Verhältnismäßigkeit):
  - Eine Patientenliste speichert die Zuordnung zwischen IDAT und PID.
- Die Rückidentifikation läuft über PSN-Dienst und Patientenliste.
  - Informationelle Gewaltenteilung

# Inhalt

1. Grundlagen von Anonymisierung und Pseudonymisierung
2. Methoden und Szenarien
- 3. Die Pseudonymisierungswerkzeuge der TMF**
4. Diskussion und Ausblick



# Typische Datenbanken



# Aufbau von Langzeit-Datenpools für die medizinische Forschung

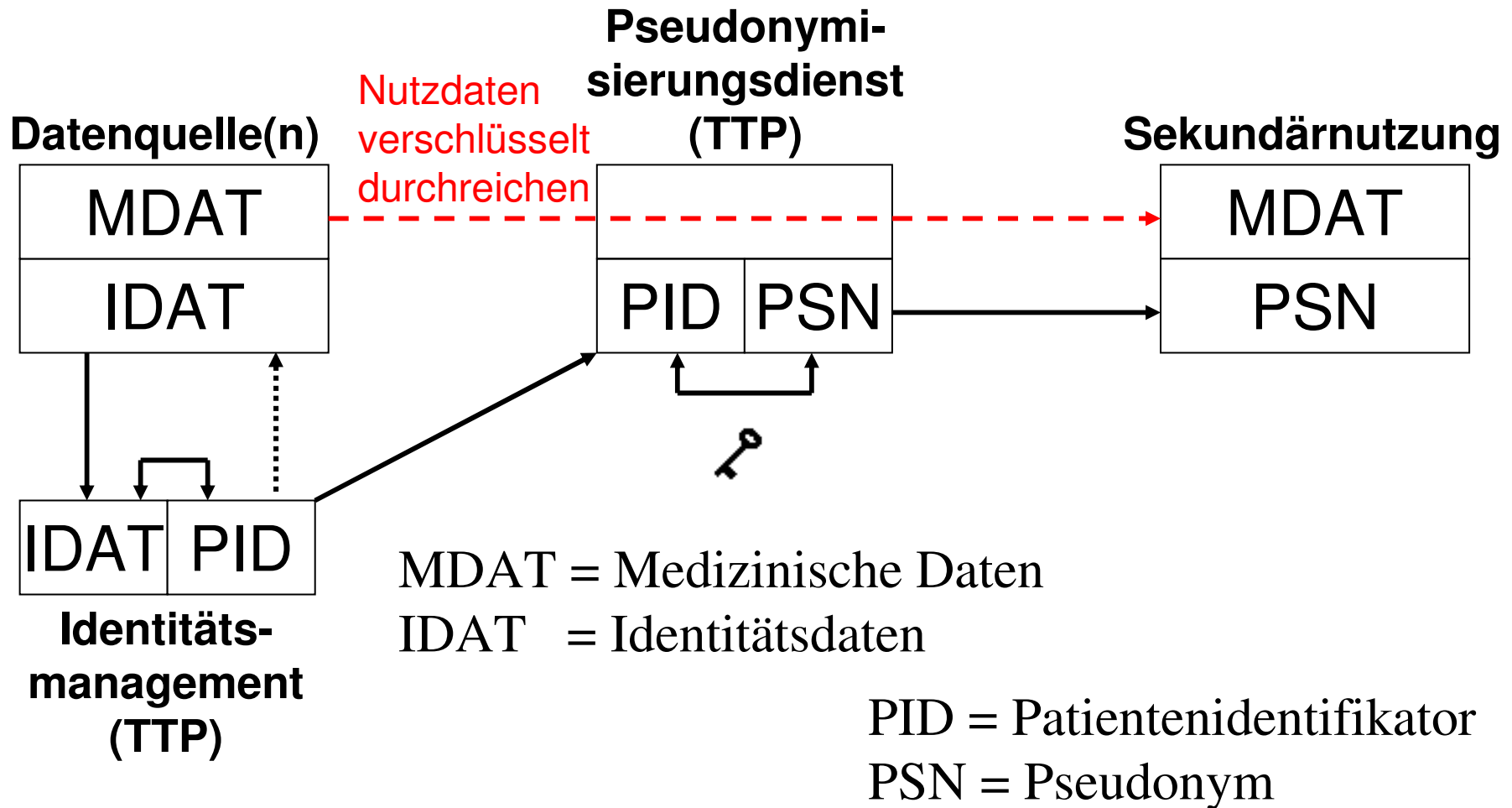
- *Patientennaher Bereich*, z. B. ePA, klinisches Register, Begleitung chronischer oder seltener Erkrankungen, klinische Studie, ...
  - TMF-Werkzeug A: „Klinische Datenbank“
  - Speicherung pseudonym, Zugriff für Berechtigte personenbezogen
- *Patientenferner Bereich*, z. B. epidemiologisches Register, Referenzbilder-Sammlung, ...
  - TMF-Werkzeug B: „Forschungsdatenbank“
  - Speicherung und Zugriff nur pseudonym

# Langzeit-Datenpools

Die Langzeit-Datensammlung erfordert

- Pseudonymisierung und TTP-Dienste,
- klar definierten organisatorischen Rahmen,
- Informationelle gewaltenteilung,
- besondere technische Sicherheitsvorkehrungen,
- entsprechende Patientenaufklärung und -einwilligung.
- Identitätsmanagement (PID-Verwaltung) und Qualitätssicherung der Daten (mit Rückfragen) müssen vor Pseudonymisierung erfolgen.

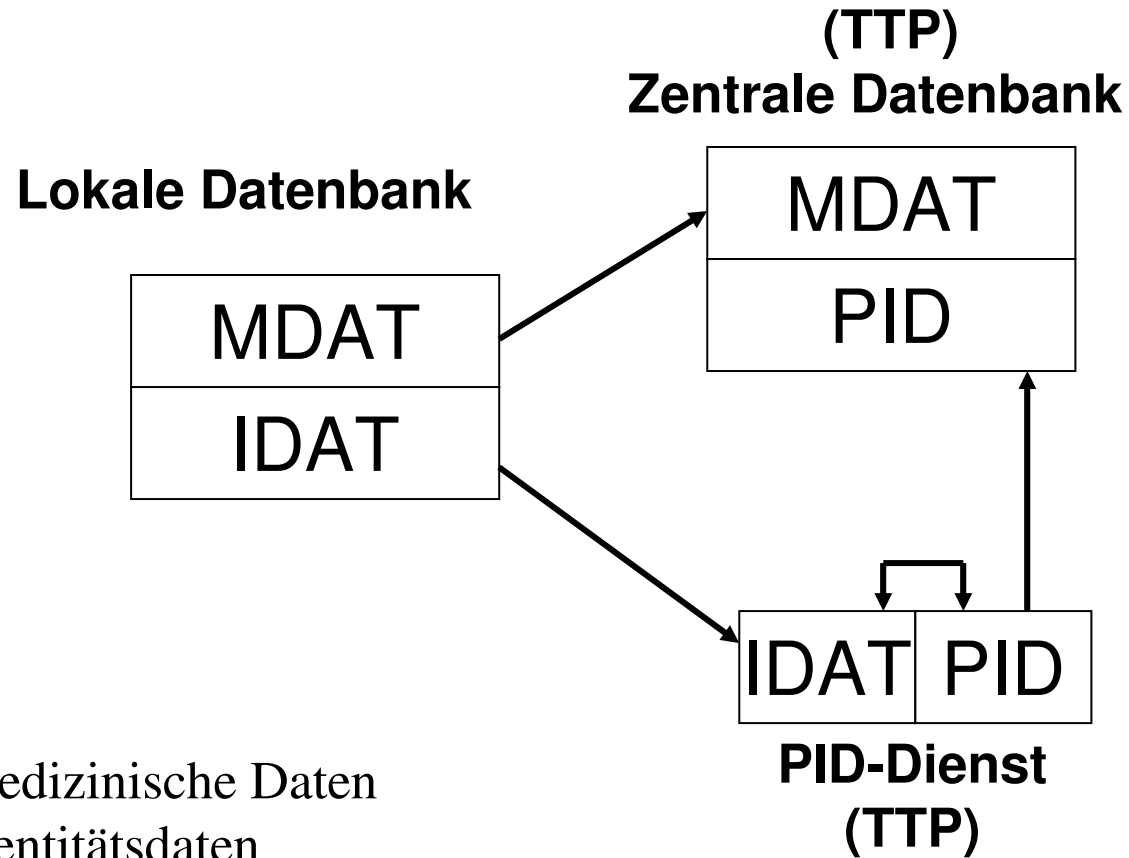
# Werkzeug B: Forschungsdatenbank



# Werkzeug A: Klinische Datenbank

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

# Die zentrale klinische Datenbank



MDAT = Medizinische Daten

IDAT = Identitätsdaten

PID = Patientenidentifikator (*hier als Pseudonym behandelt*)

PSN = Pseudonym



# Besonderheiten von Werkzeug A

- 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.
- Allgemein: für „patientennahe“ Forschung

# Ergebnisse I

- TMF-Datenschutzkonzept mit alternativer Verwendung von Werkzeug A oder B von den Datenschutzbeauftragten positiv bewertet
  - (Arbeitskreis Wissenschaft und AK Gesundheit der Datenschutzbeauftragten des Bundes und der Länder)
- Modell A in einem Forschungsnetz implementiert.
  - Weitere in Vorbereitung oder Einführung.
- Modell B von mehreren Netzen adaptiert;
  - Implementierungen in Arbeit.

# Ergebnisse II

- Die TMF bietet Software-Tools für die TTP-Dienste.
- Zugehörige Policies, Musterverträge, Mustereinwilligungserklärungen von der TMF erhältlich (frei für Mitglieder).
- Buchveröffentlichung in TMF-Schriftenreihe:
  - Reng/Debold/Specker/Pommerening:  
»*Generische Lösungen zum Datenschutz für die Forschungsnetze in der Medizin*«.


Schriftenreihe der Telematikplattform  
für Medizinische Forschungsnetze

C.-M. Reng | P. Debold  
Ch. Specker | K. Pommerening



# Generische Lösungen zum Datenschutz für die Forschungsnetze in der Medizin

im Auftrag des Koordinierungsrates  
der Telematikplattform für  
Medizinische Forschungsnetze

 Medizinisch Wissenschaftliche Verlagsgesellschaft



# Inhalt

1. Grundlagen von Anonymisierung und Pseudonymisierung
2. Methoden und Szenarien
3. Die Pseudonymisierungswerkzeuge der TMF
- 4. Diskussion und Ausblick**

# Anwendungserfahrungen

- Workshop Ende 2005 mit 14 Netzen.
- Ergebnisse:
  - Begutachtung des Datenschutzkonzepts (meist) zügig und positiv.
  - Anpassung an Modell A oder B oft schwierig, nicht alle Anforderungen abgedeckt.
  - Implementierung oft langwierig und aufwendig.
- Fazit: Überarbeitung des generischen Datenschutzkonzepts bis 2008.

# Revision des TMF-Datenschutzkonzepts

## Ziele:

- Dichotomie „A oder B“ beseitigen, statt dessen:
- Modularer Aufbau mit optionalen Komponenten;
  - für jede Komponente das passende DS-Werkzeug.
- Skalierbarkeit, Verhältnismäßigkeitskriterien.
- Verzahnung mit Versorgung und klinischen Studien besser berücksichtigen.
- Etablierung zentraler Dienstleistungen.

Internationalisierung als separates Projekt.

