ID-Management mit dem PID-Generator der TMF für das KPOH

TMF-Workshop ID-Management

Berlin, 15. Dezember 2008

Prof. Dr. Klaus Pommerening
Universität Mainz, IMBEI
KN Pädiatrische Onkologie und Hämatologie
TMF-AG Datenschutz

Gefördert vom
Bundesministerium für Bildung und Forschung

Übersicht

1. **Identität und Pseudonym**
2. Der PID-Generator im KPOH
3. Der PID-Generator im TMF-Datenschutzkonzept

Identität eines Patienten?

Behandlungskontext

[Primärnutzung]

Barriere: Ärztliche Schweigepflicht

[Sekundärnutzung/Forschungskontext]

klinische Forschung
Versorgungsforschung

direkte
Erfassung

Export erlaubt, wenn
- anonyme, (evtl. pseudonyme) Daten,
- Einwilligung,
- Gesetzesvorschrift

Register/
epidemiologische Forschung

Patienten-Identifikation

Behandlungszusammenhang:

- Identitätsdaten/ persönliche Ansprache
- künftig: elektronische Gesundheitskarte (eGK)

Sekundärnutzung von Patientendaten (Forschung, Qualitätssicherung, …):

- Anonymisierung oder
- Identitätsmanagement über Pseudonyme durch vertrauenswürdige Instanzen („Datentreuhänder“, „Trusted Third Parties“ (TTPs))

Pseudonyme

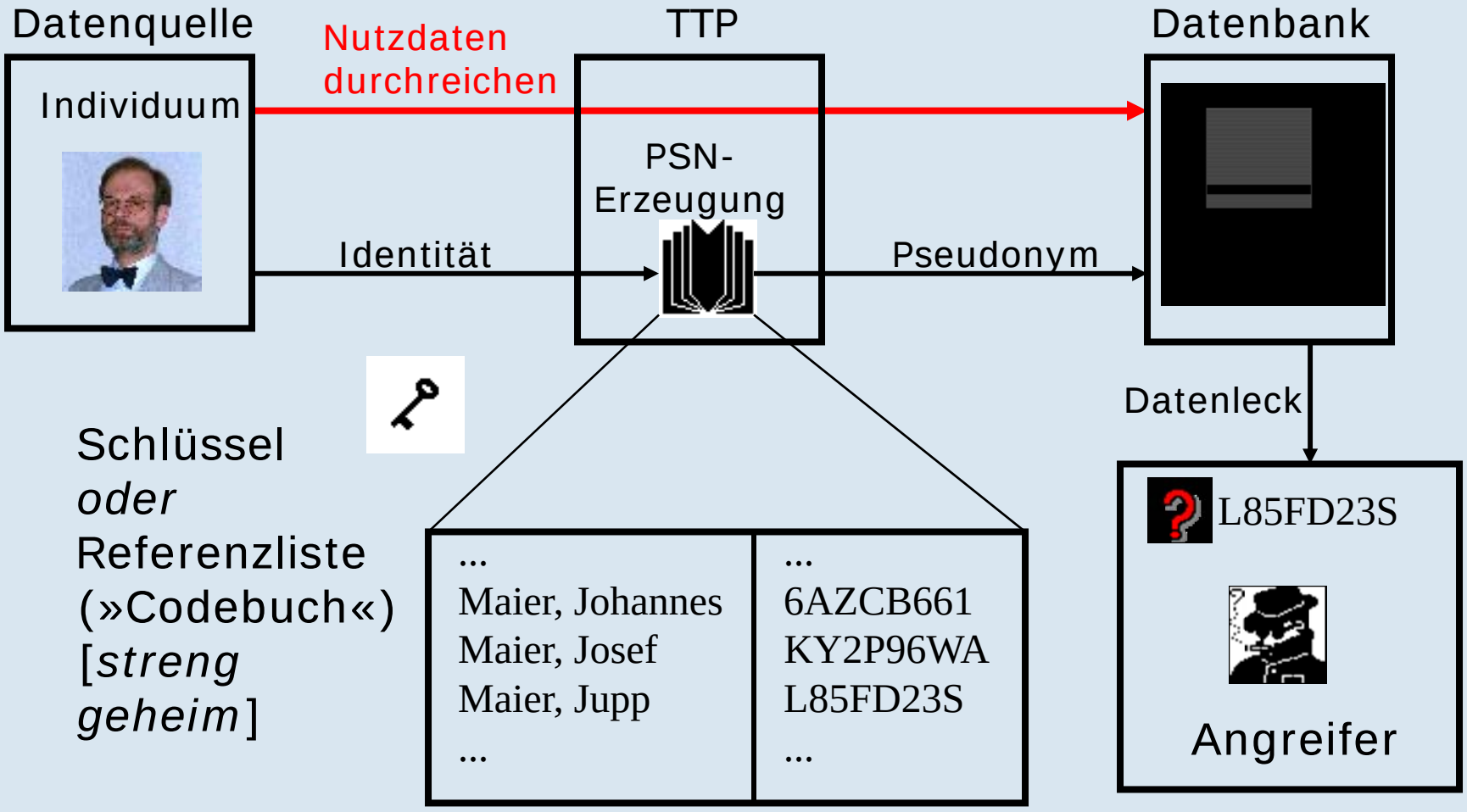
Grundtyp 1 (**aktive Pseudonyme**, Chaum ca 1980):

- vom Inhaber selbst verwaltet,
- Inhaber bei Nutzung präsent.
- Gültigkeit und Rechtssicherheit:
  Zertifikat durch blinde digitale Signatur.
- Geeignet für Rechtsbeziehungen mit Anonymitätsanspruch.

Grundtyp 2 (**passive Pseudonyme**,
Michaelis/Pomm ca 1990 für Krebsregister):

- von Datentreuhänder verwaltet,
- Inhaber bei Nutzung nicht präsent.
- Geeignet für Datensammlungen,
  z. B. in medizinischen Forschungsprojekten.
- Achtung: Rechtlich *nicht* zur Anonymität äquivalent!

Grundtyp 2: Das Basismodell



Übersicht

1. Identität und Pseudonym
2. **Der PID-Generator im KPOH**
3. Der PID-Generator im TMF-Datenschutzkonzept

Motivation der Entwicklung:

Struktur der Pädiatrischen Onkologie und Hämatologie (POH) weit verteilt – z. T. durch KPOH abgedeckt.

Wunsch nach eindeutigem Patienten-Identifikator für Patienten.

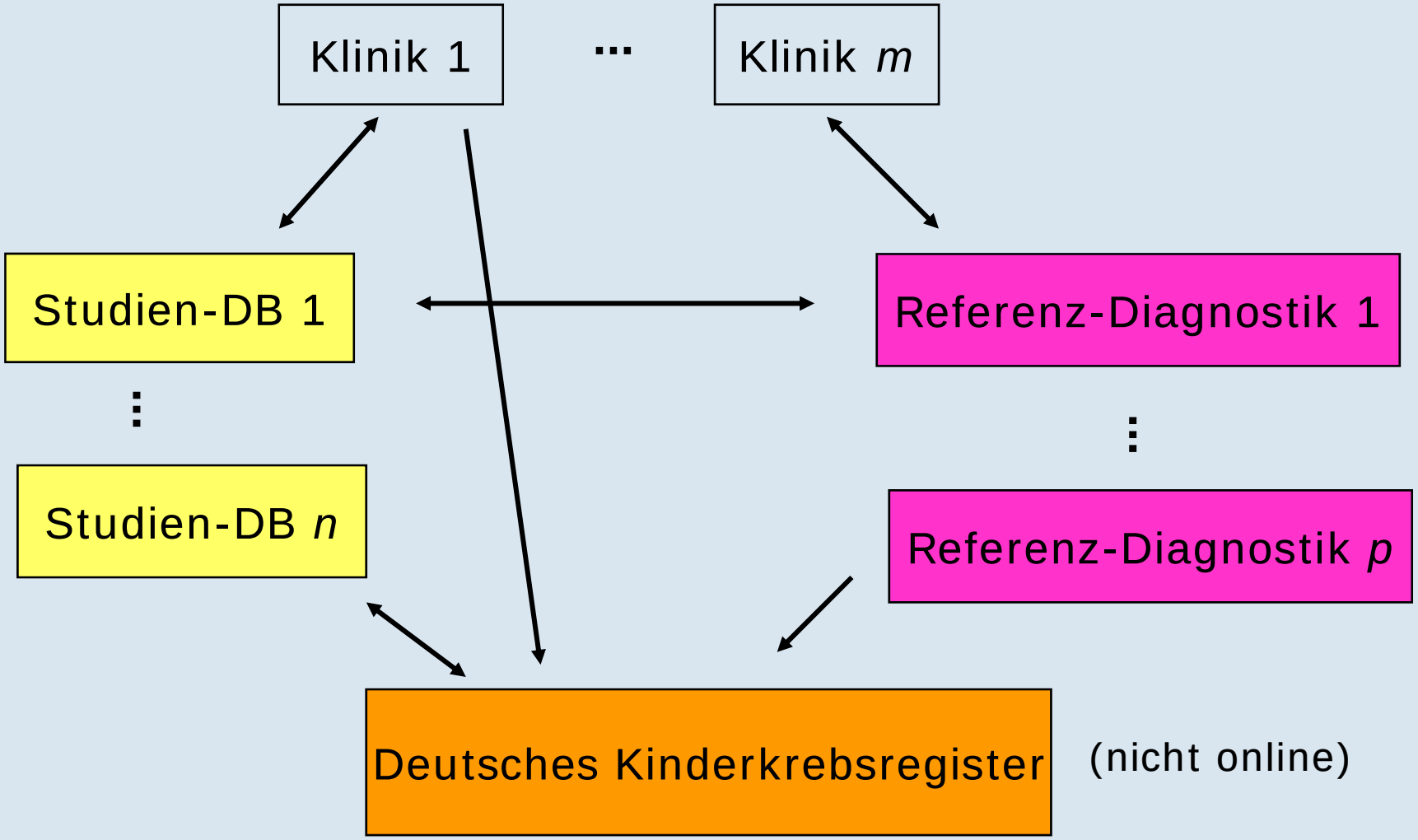
- Patient nicht überall physisch anwesend.

Einführung 2002 nach GPOH-Vorstandsbeschluss.

PID soll auch als Pseudonym dienen

- als SIC (Subject Identification Code) im Sinne des AMG
- [d. h., abweichend vom Basismodell – aber AMG-konform – ist das Pseudonym auch der Datenquelle bekannt]
- und nach Anonymisierung den Fall für die Krebsforschung verfügbar halten.
- Zwischenschritt „Kontrollnummern“ soll zum pseudonymen Abgleich mit Landeskrebsregistern dienen.

Verteilte Struktur der POH



Realisierung

- Software am IMBEI entwickelt – Entwickler und Betreuer:
  - Markus Wagner (bis 2003)
  - Jutta Glock (Moormann) (2003-2005)
  - Murat Sariyar (seit 2005)
- Zentraler Web-Service am IMBEI,
  - dedizierter Server.
- Web-Formular für Anforderung durch Studienärzte;
  - Übertragung mit „Copy & Paste“.
- Batch-Verarbeitung auch möglich
  - durch Server-Administrator,
  - z. B. für „Altfälle“ des DKKR (= deutsches Kinderkrebsregister).
- Einbindung in vorhandene Erfassungsprogramme (DKKR, Studien).
  - SOAP-Schnittstelle.

Anforderung eines Patienten-Identifikators (GPOH-PID)

Kompetenznetz Pädiatrische Onkologie und Hämatologie

Erklärung/Hilfe [Vor der ersten Verwendung unbedingt lesen!]

Identifizierende Angaben		Wie sicher ist der Name?	○ sicher ◉ unsicher
Nachname:		Vorname:	
früherer Nachname:		Geburtsdatum	TT: MM: JJJJ:
Ergänzende Angaben			
Geschlecht:	○ weiblich ○ männlich ◉ unbekannt		
Postleitzahl:		Wohnort:	
		Staat:	

Bevor Sie das Formular abschicken, vergewissern Sie sich bitte noch einmal, ob alle Einträge korrekt sind.

GPOH-PID anfordern

Formular zurücksetzen

Falls Sie als Reaktion nicht einen PID oder eine verständliche Fehlermeldung zurück erhalten, wenden Sie sich per E-Mail an `PIDservice@gpoh.de`.

Das Werkzeug PID-Generator

Match-Verfahren austauschbar, optimierbar

PIDs maschinenlesbar (AES von lfd. Nummer)
oder menschenlesbar (8-Zeichen, Faldum-Code)

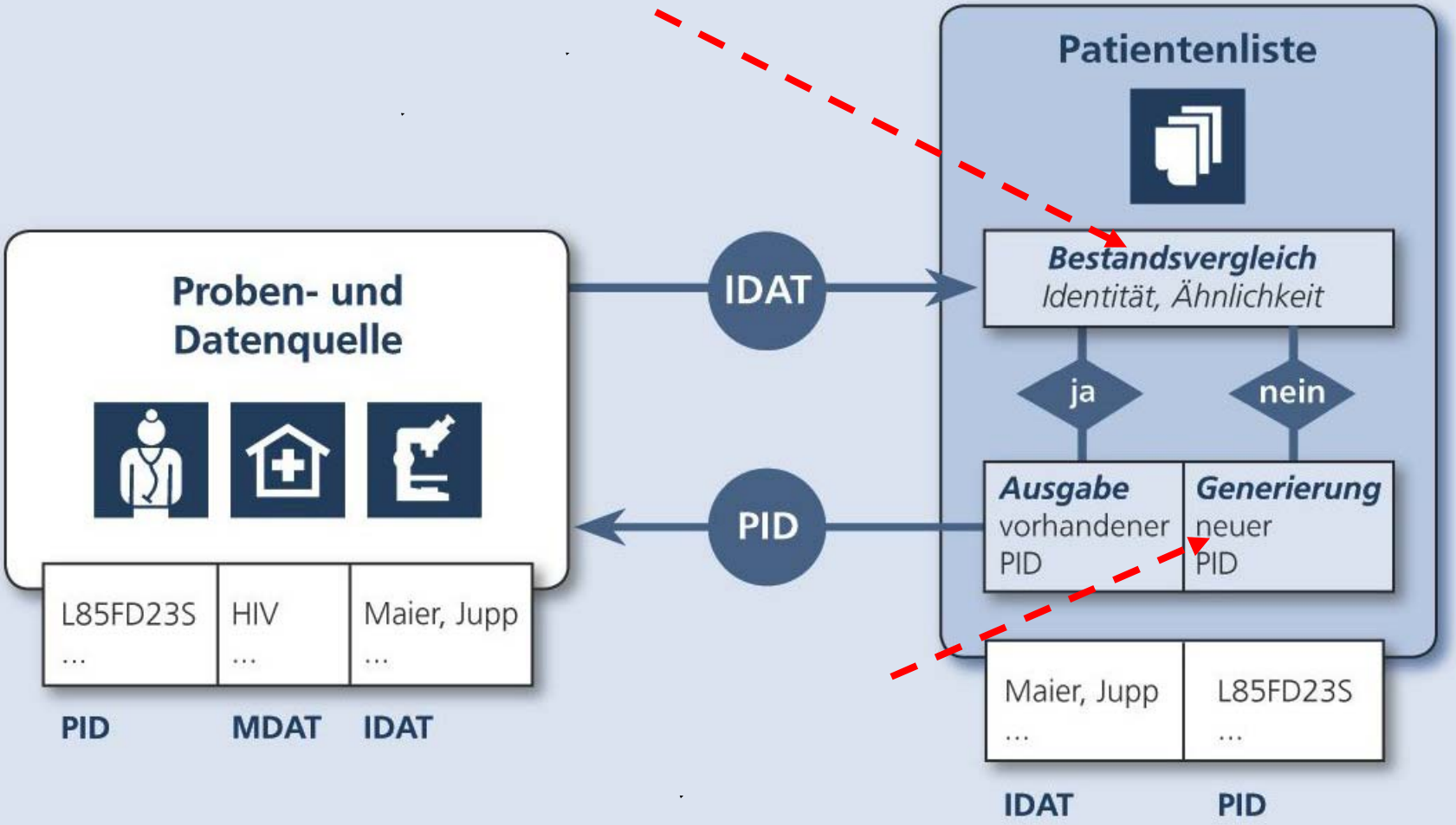
Konfigurierbar für verschiedene Anwendungsszenarien.

PID-Generator in Kompetenznetz POH seit 2002
(ca 57 000 PIDs).

- Die 44 248 „Altfälle“ des DKKR im Batch-Lauf eingespeist.
- Neu beginnende Studien wollen/sollen den PID verwenden.

Umfangreiche Evaluation, insb. des Match-Verfahrens.

PID-Erzeugung



Das Match-Problem (Record Linkage)

*Richtige Zuordnung (fast) nur **vor** Pseudonymisierung möglich*.

- Daher integraler Bestandteil des pseudonymen ID-Mgt.

Logisches Matchen

- Erkennen von (z. B.) Namensänderung, Namenszusatz
- Wenn möglich, KV-Nummer verwenden.

Matchen mit *»unsicherem«* Namen:

- Zusatzdaten und phonetische Codes werden mitverwendet,
- evtl. Warnhinweis.

Homonym- vs. Synonymfehler

- Stochastische/ KI-Matchverfahren getestet,
- „klassisches“ Record Linkage überlegen.
- Fehler nie ganz auszuschließen,
- Phonetik reduziert Synonymfehler.

Übersicht

1. Identität und Pseudonym
2. Der PID-Generator im KPOH
3. **Der PID-Generator im TMF-Datenschutzkonzept**

Einsatzszenarien

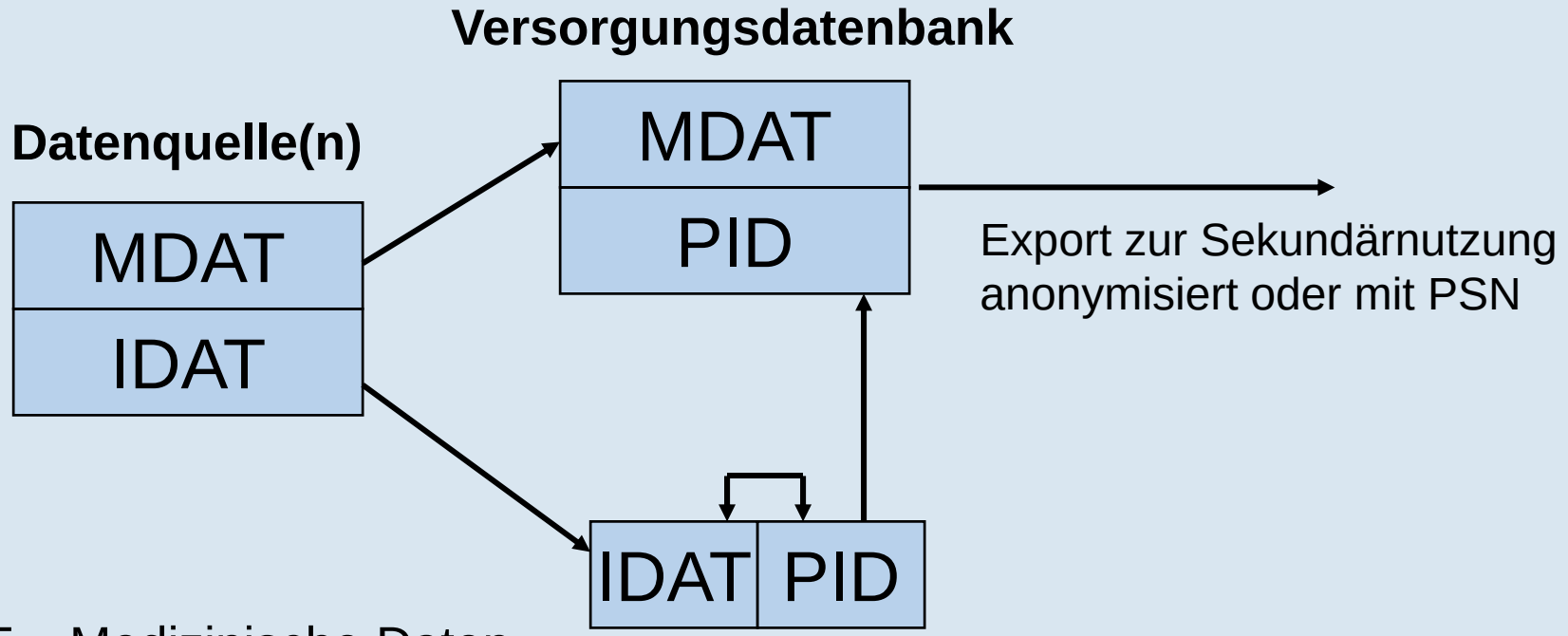
(A) PID als pseudonymes Kennzeichen in einrichtungsübergreifender **Versorgungsdatenbank** mit separater Patientenliste:

- Speicherung pseudonym,
- Online-Zugriff personenbezogen (für Berechtigte).

(B) PID als zusätzliches Kennzeichen zu IDAT:

- Pseudonym für „**Forschungsdatenbank**“ als verschlüsselter PID (durch Pseudonymisierungsdienst)
- Speicherung und Zugriff nur pseudonym.
- Fehlertoleranz bei Erzeugung zur Erhöhung der Datenqualität nötig (→ Record Linkage).

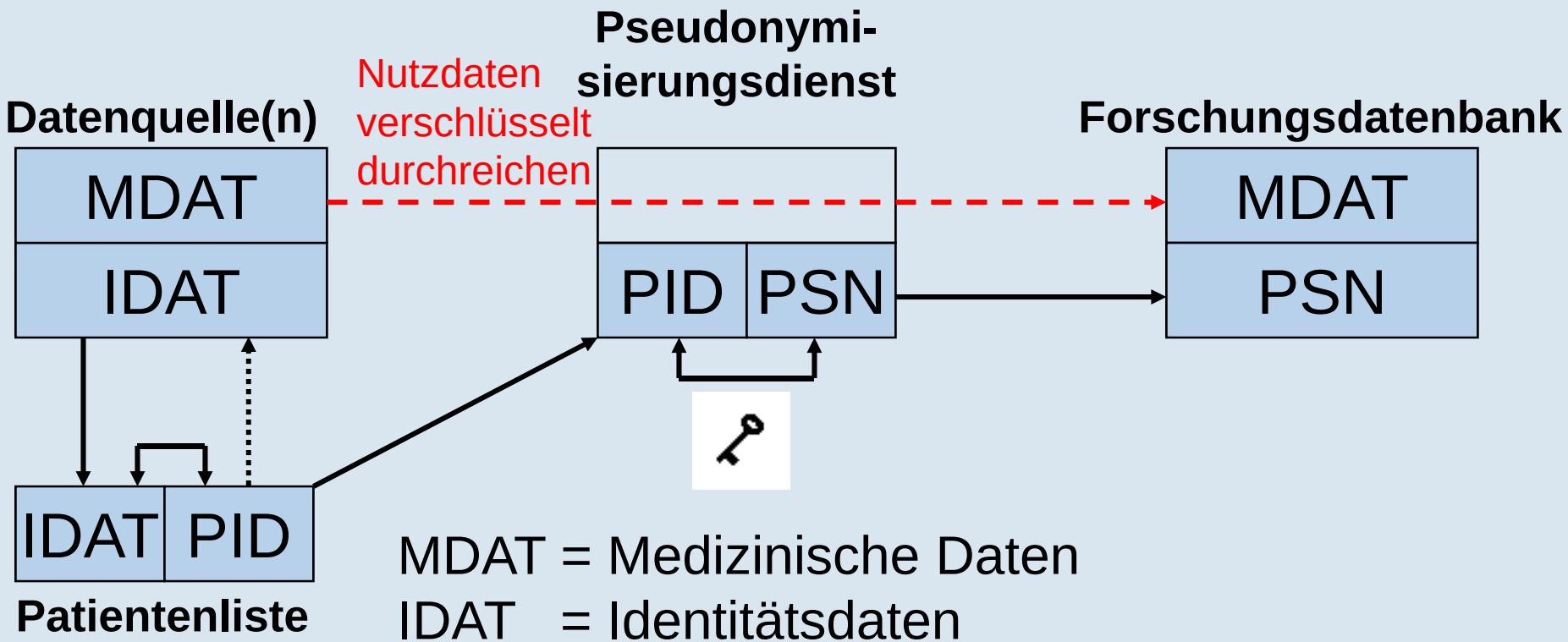
Identitätsmanagement im TMF-Modell A (patientennahe „Versorgungsdatenbank“)



MDAT = Medizinische Daten
IDAT = Identitätsdaten

PID = Patientenidentifikator
PSN = Pseudonym

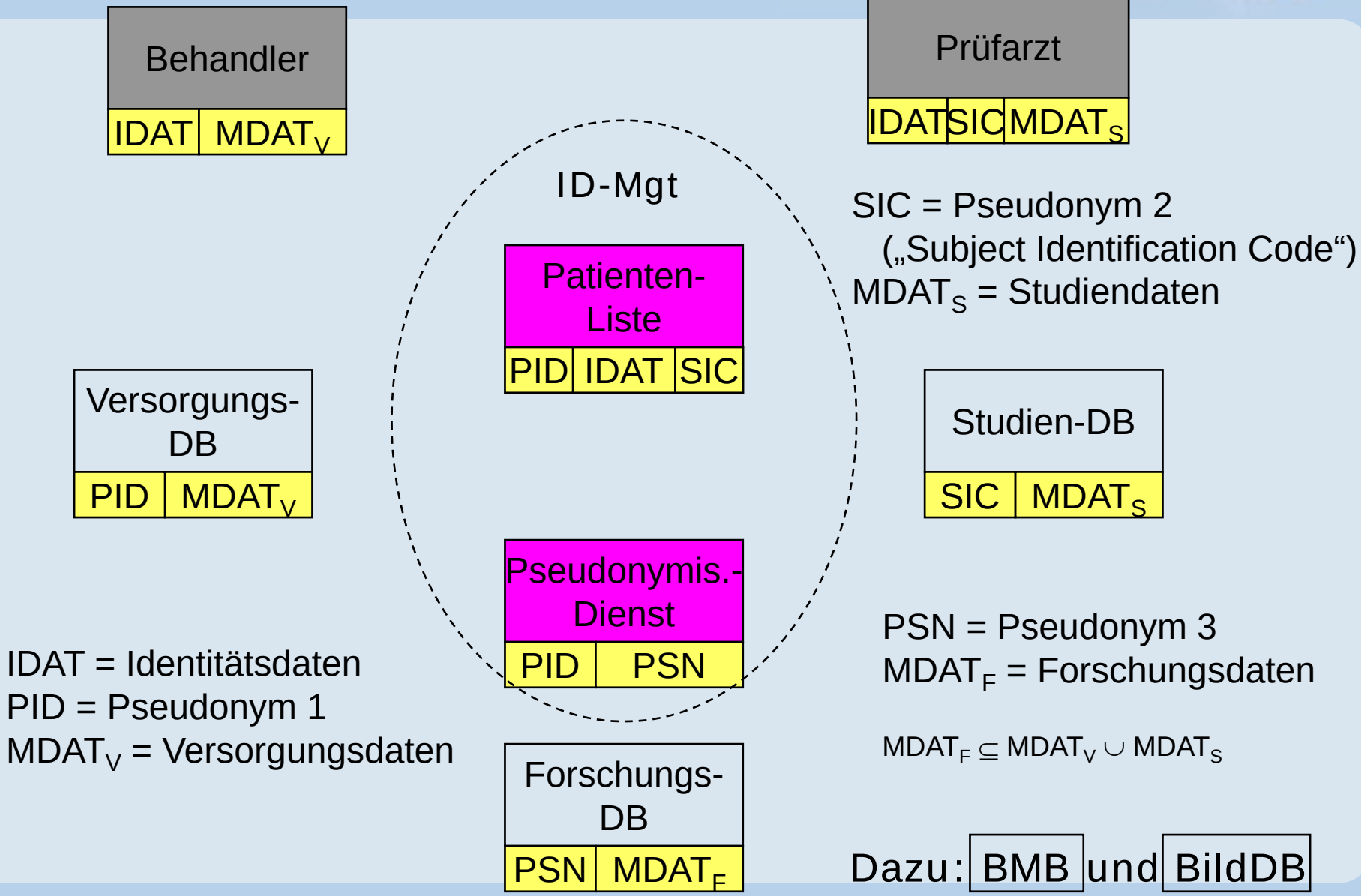
Identitätsmanagement im TMF-Modell B (patientenferne „Forschungsdatenbank“)



MDAT = Medizinische Daten
IDAT = Identitätsdaten

PID = Patientenidentifikator
PSN = Pseudonym

Referenzmodell im künftigen revidierten Konzept: Datenbanken und Daten



SIC = Pseudonym 2
(„Subject Identification Code“)
$MDAT_S$ = Studiendaten

IDAT = Identitätsdaten
PID = Pseudonym 1
$MDAT_V$ = Versorgungsdaten

PSN = Pseudonym 3
$MDAT_F$ = Forschungsdaten

$MDAT_F \subseteq MDAT_V \cup MDAT_S$

Dazu: BMB und BildDB

Die Versorgungs-Datenbank

… enthält Daten, die für die Versorgung des Patienten relevant sind,

… steht im unmittelbaren Behandlungskontext,

… ist aber einrichtungsübergreifend

… und wird daher pseudonym (PID) geführt.

Behandler haben einen personenbezogenen Zugriff auf die Daten ihrer Patienten (lesend und schreibend).

Der Zugriff geschieht mit Hilfe der Patientenliste (ID-Management für Patienten) und mit Hilfe eines Verzeichnisdienstes (ID-Management für Benutzer).

(Weiterentwicklung aus TMF-Modell A.)

Die Studien-Datenbank

... dient zur Durchführung klinischer Studien nach den Regularien des AMG und der guten klinischen Praxis (GCP).

... enthält Daten zum Patienten, die für die Studie relevant sind; die Überschneidung mit den Daten der reinen Versorgungsdokumentation ist groß.

... steht im unmittelbaren Behandlungskontext, soweit es um Zugriffe durch den Prüfarzt geht; sie steht im Forschungskontext, wenn es um Zugriffe durch den „Sponsor“ oder Studienleiter geht.

... ist einrichtungsübergreifend

... und wird daher pseudonym (SIC) geführt.

Prüfärzte haben einen personenbezogenen Zugriff auf die Daten ihrer Patienten (lesend und schreibend) und kennen SIC.

(Entspricht POH-Modell mit GPOH-PID als SIC.)

Die Forschungs-Datenbank

... dient zur Langzeitspeicherung pseudonymisierter medizinischer Daten für spätere Forschungsprojekte

- direkt zur epidemiologischen Forschung,
- zur Rekrutierung geeigneter Fälle für neue klinische oder epidemiologische Forschung.

... bietet den nochmals pseudonymisierten Export geeigneter Daten
(evtl. je nach Beurteilung des RI-Risikos einen Direktzugriff für Forscher).

(Weiterentwicklung aus TMF-Modell B.)

Erwünschte Weiterentwicklungen des PID-Generators

Für revidiertes DS-Konzept: Identitätsmanagement um weitere pseudonyme Kennzeichen erweitern (durch Verschlüsselung des PID)

- Für Studiendatenbanken (SIC), Bilddatenbanken (BildID), genetische Analysen (LabID).

Internationalisierung

- Zeichensätze: Transkription, Unicode,
- Phonetik.