GMDS Jahrestagung –
Leipzig, 12. September 2006

Pseudonymisierungsdienst auf Basis von Web Services – erfolgreiche Praxis der Generischen Datenschutzkonzepte für medizinische Forschungsnetze

Semler SC[1], Drepper J[1], Pommerening K[2],
Schlösser-Faßbender M[3], Schröder M[3], Rienhoff O[4]

1 Telematikplattform für Medizinische Forschungsnetze e.V. (TMF), Berlin, Deutschland

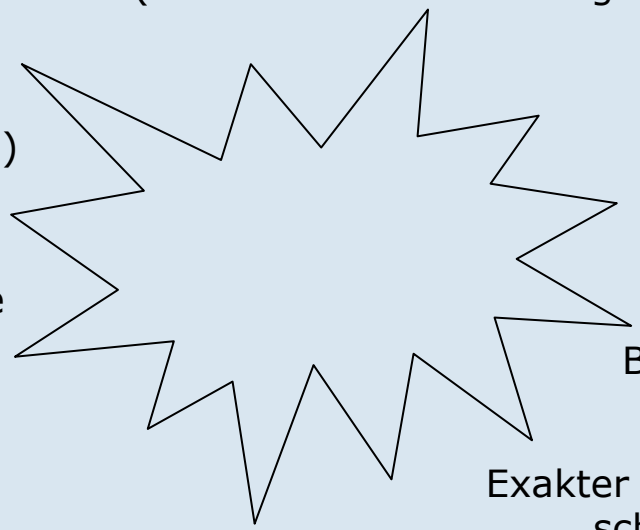
2 IMBEI, Univ. Mainz, Deutschland

3 Tembit Software GmbH, Deutschland

4 Inst. f. Medizininformatik, Univ. Göttingen, Deutschland

Einführung: Datenschutz
Forschung - trotz oder mit Datenschutz ?

- Das deutsche Datenschutzrecht stellt hohe Hürden für die vernetzte medizinische Forschung auf.



Aufgabe:

- Austarieren von Persönlichkeitsrechten von Patienten/Probanden (informationelle Selbstbestimmung) und Durchführbarkeit medizinischer Forschung (mit vertretbarem Aufwand)

Begriffsbestimmung

- Personenbeziehbarkeit
- Personenbezug (Personenbezogenheit)
- Anonymisierung (Anonymisiertheit)
- Pseudonymisierung (Pseudonymisiertheit)
  - einstufig vs. mehrstufig
  - zentral vs. dezentral
  - Einweg- vs. bidirektionel
- Depseudonymisierung
- Re-Identifikation

Definition der Grundbegriffe in §3 & §3a Bundesdatenschutzgesetz (BDSG)

§ 3 Weitere Begriffsbestimmungen

(1) **Personenbezogene Daten** sind Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person (Betroffener).

(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.

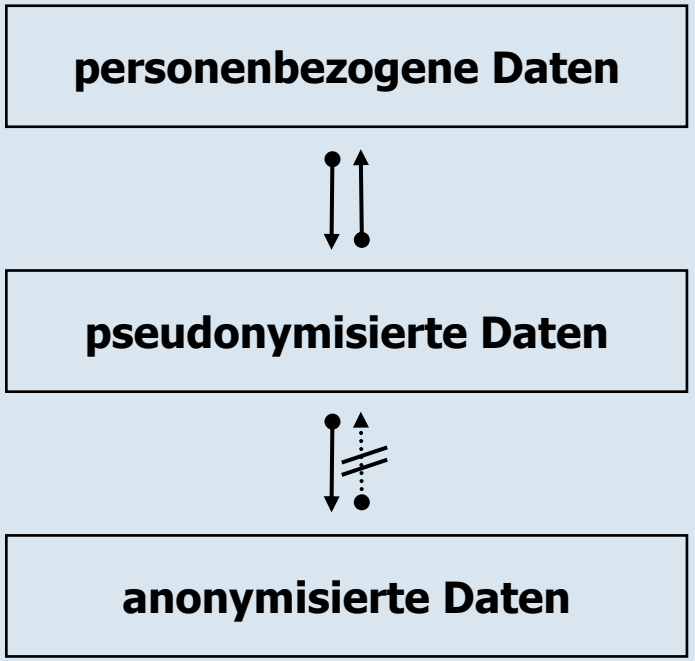
(6a) **Pseudonymisieren** ist das Ersetzen des Namens und anderer Identifikationsmerkmale durch ein Kennzeichen zu dem Zweck, die Bestimmung des Betroffenen auszuschließen oder wesentlich zu erschweren.

(9) **Besondere Arten personenbezogener Daten** sind Angaben über die rassische und ethnische Herkunft, politische Meinungen, religiöse oder philosophische Überzeugungen, Gewerkschaftszugehörigkeit, **Gesundheit** oder Sexualleben.

§ 3a Datenvermeidung und Datensparsamkeit

Gestaltung und Auswahl von Datenverarbeitungssystemen haben sich an dem Ziel auszurichten, keine oder so wenig personenbezogene Daten wie möglich zu erheben, zu verarbeiten oder zu nutzen. **Insbesondere ist von den Möglichkeiten der Anonymisierung und Pseudonymisierung Gebrauch zu machen**, soweit dies möglich ist und der Aufwand in einem angemessenen Verhältnis zu dem angestrebten Schutzzweck steht.

Personenbeziehbarkeit, Personenbezogenheit, Personenbezug



Unterscheide:

- ***Personen identifizierende Daten (IDAT)***
- ***medizinische Daten (MDAT)***

faktisch anonymisiert

vollkommen anonymisiert (κ-Anonymisierung)

formale Anonymisierung = bezieht sich auf technischen Anonymisierungsvorgang

Depseudonymisierung ≠ Re-Identifikation

Datenschutz – Grundsätzliche Aspekte der Umsetzung

- Recht auf informationelle Selbstbestimmung des Probanden / Patienten
  - grundsätzliche Patienteneinwilligung, Auskunftspflicht
- Prinzip der Datensparsamkeit
  - need-to-know-Prinzip
  - Anonymisierung, wo immer möglich
  - immanente Gefahr der Personenidentifikation (Kumulation eindeutiger Merkmale)
- Pseudonymisierungslösungen
  - nötig bei Verlaufsstudien / Follow-ups
  - hoher Geheimnisschutz des Pseudonyms erforderlich
  - kurzfristig: einstufig – langfristig: zweistufig (IDAT -> PID -> PSN)
- „informationelle Gewaltenteilung“
  - verteilte Datenhaltung zur Minimierung der Re-Identifizierbarkeit
  - Trennung von identifizierenden und medizinischen Daten
- komponentenbasierte Software-Architektur (SOA)
  - passend zu verteilt organisierten Forschungsnetzen
  - gewährleistet Unabhängigkeit des administrativen Zugriffs

Realisierung in der TMF
Konzepte, Spezifikationen, Softwarekomponenten

- 2001 – 2003 Generische Datenschutzkonzepte (I)
  Konsentierung mit Landesdatenschutzbeauftragten
- 2001 – 2003 Softwarespezifikationen (PIDGen, PSD)
- 2002/2003 Softwarerealisierung PIDGen
  erster Betrieb der PIDGen-Komponente (KN POH)
- 2003/2004 Softwarerealisierung PSD
- 2004 – 2006 Erweiterung der Datenschutzkonzepte: Biobanken
  Konsentierung mit Landesdatenschutzbeauftragten
- 2005 Re-Design der PIDGen-Schnittstellen (SOAP)
- 2005 Revision und Erweiterung der PSD-Spezifikation
- 2006 komplette Reimplementierung PSD
  erster Betrieb der PSD-Komponente (KN AHF)
- ab 2006 Fortschreibung gener. Datenschutzkonzepte (II)
- ab 2006 Spezifikation für Datentreuhänderdienst
- PKI mit Anbindung an die Telematikinfrastrukturen im Patientenversorgungsbereich (Gematik) ungelöst (2009 ?).

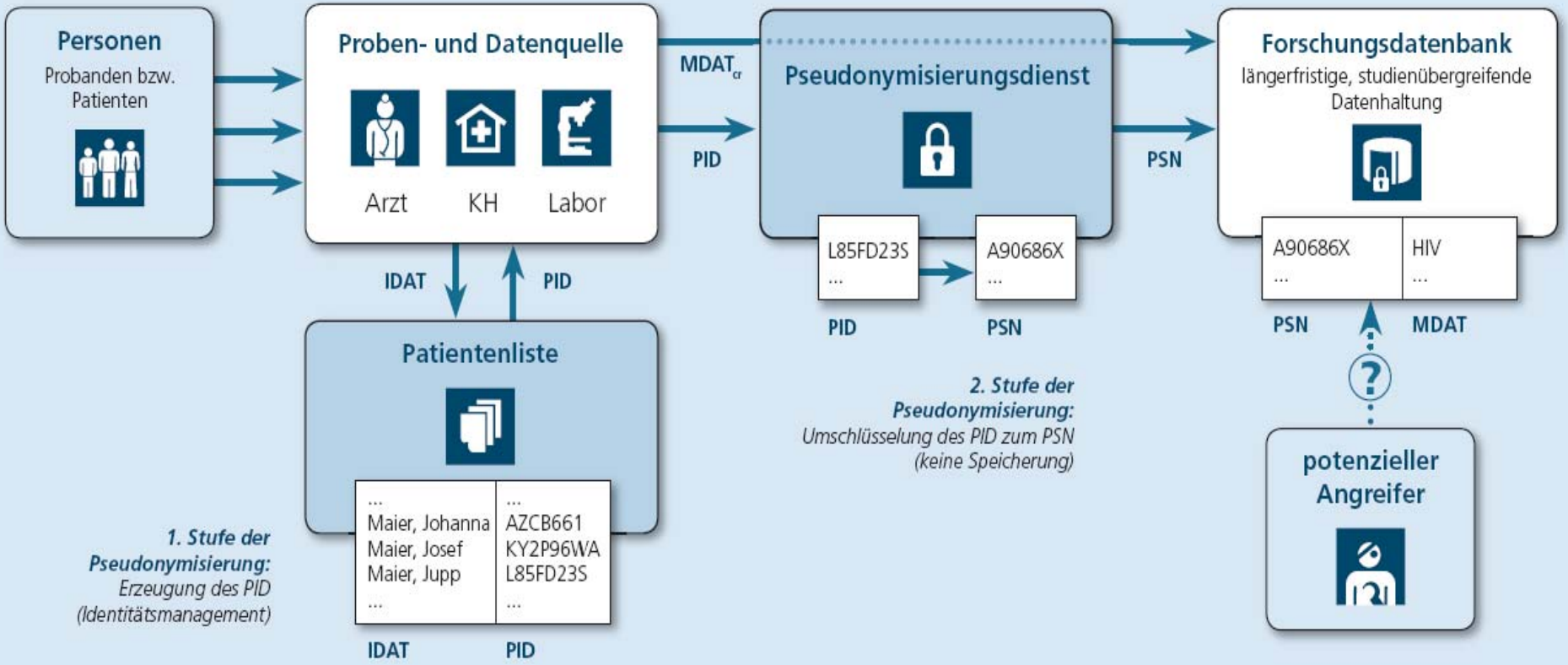
Partner und Dienstleister

- Fraunhofer ISST, Berlin (bis 2003)
- Fraunhofer SIT, Darmstadt
- Schlumberger (heute zu ATOS gehörig)
- Debold & Lux
- interactive systems GmbH (ias), Berlin
- Tembit GmbH, Berlin
- IMBEI, Universität Mainz
- CIOffice / Abt. f. Medizininformatik, Universität Göttingen
- Kompetenznetz POH
- Kompetenznetz AHF
- Kompetenznetz Rheuma
- Kompetenznetz CED
- Kompetenznetz HIV / KKS Köln
  … und weitere Kompetenznetze und KKS

Derzeit Kontakte zu industr. Partnern (Hosting, SW-Development)

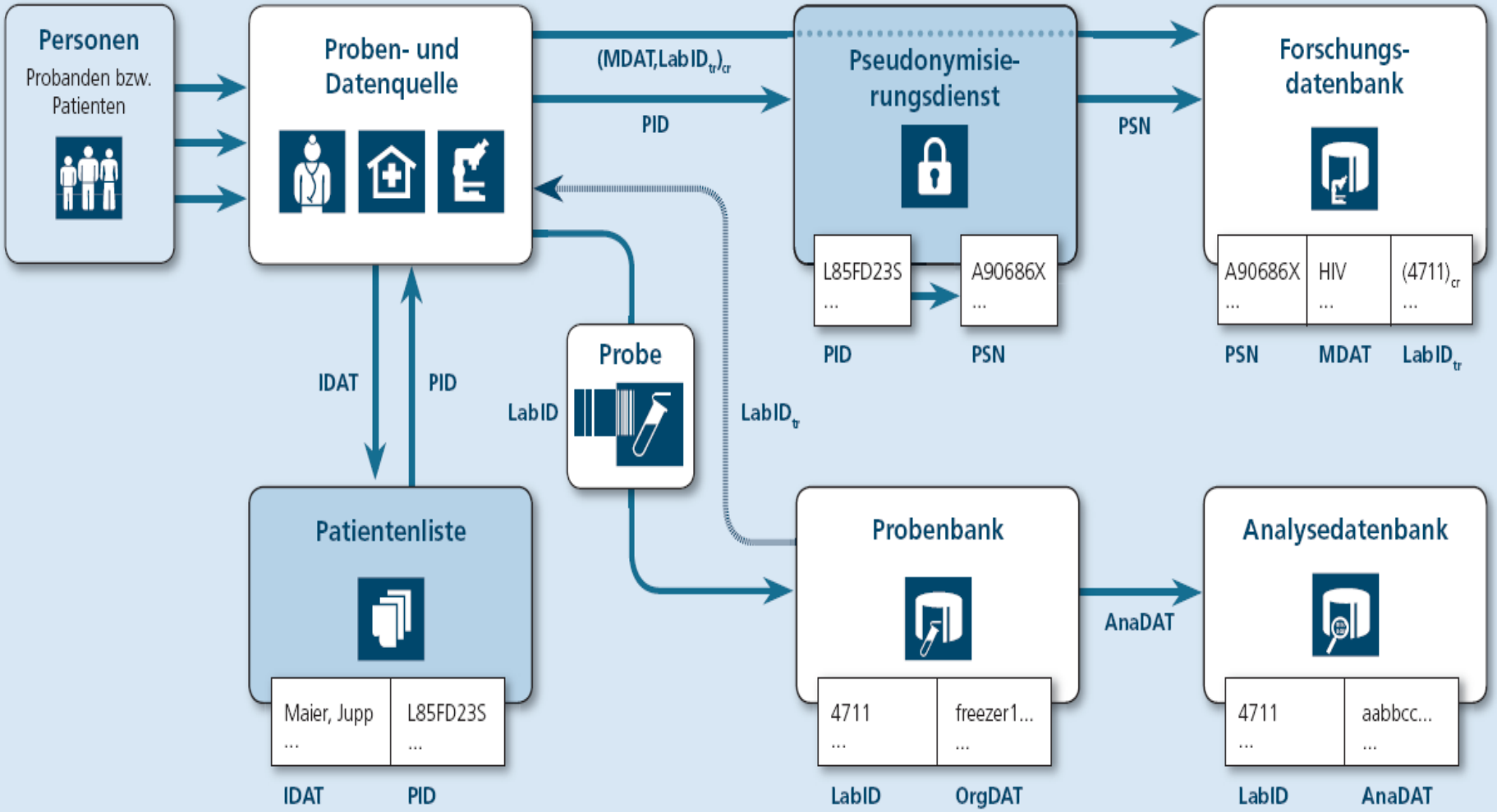
Pseudonymisierung, zweistufig

... gleichartige Szenarien in EPA-Projekten !



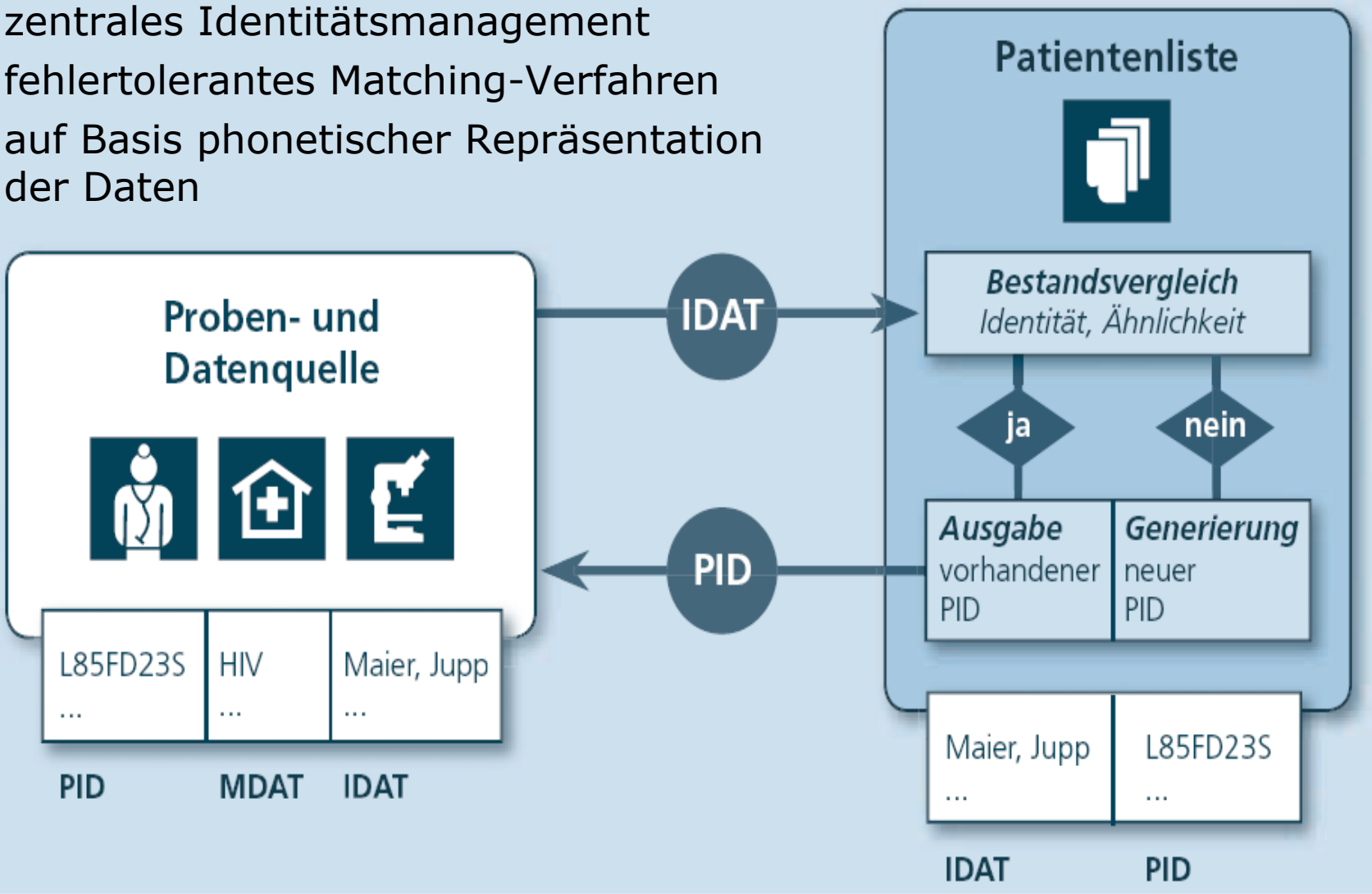
Generisches Datenschutzkonzept für Biobanken

„Maximalmodell" – konsentiert mit dem AK Wissenschaft der Landesdatenschützer (2006)



Erste Stufe der Pseudonymisierung: PID-Generator

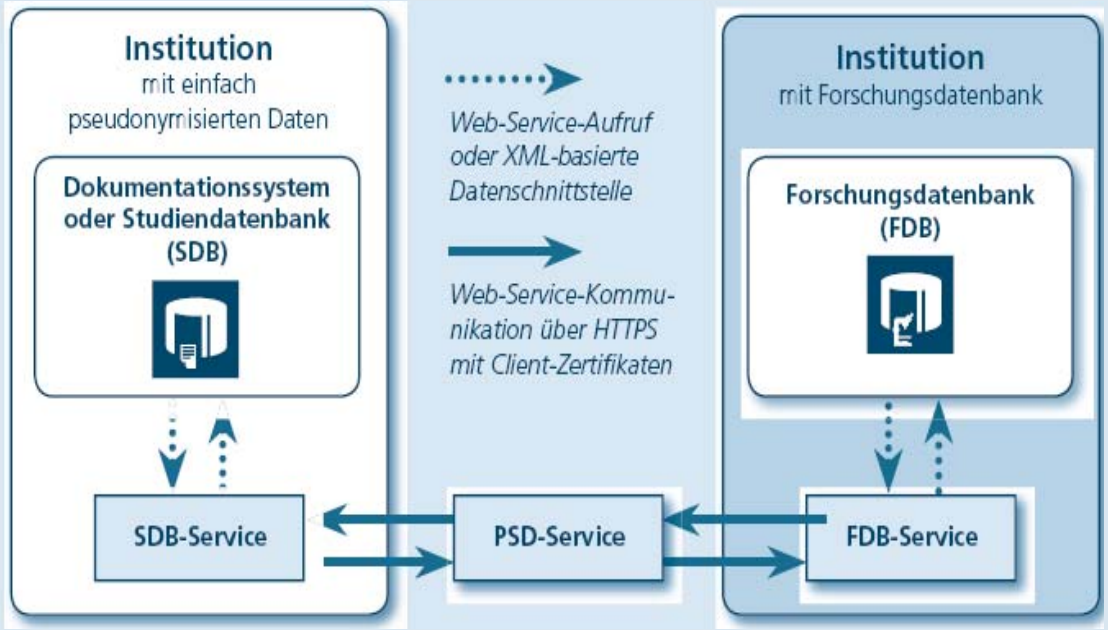
- zentrales Identitätsmanagement
- fehlertolerantes Matching-Verfahren
- auf Basis phonetischer Repräsentation der Daten



Zweite Stufe der Pseudonymisierung: PSD

3 Komponenten:

1. SDB-Service:
   - nimmt die unverschlüsselten medizinischen Daten und das einstufige Pseudonym entgegen
   - verschlüsselt asymmetrisch die medizinschen Daten (mit öffentlichem Schlüssel des FDB-Service) über separierte Crypter-Komponente
   - signiert vollständigen Datensatz
   - bei De-Pseudonymisierung: Anforderung an FDB



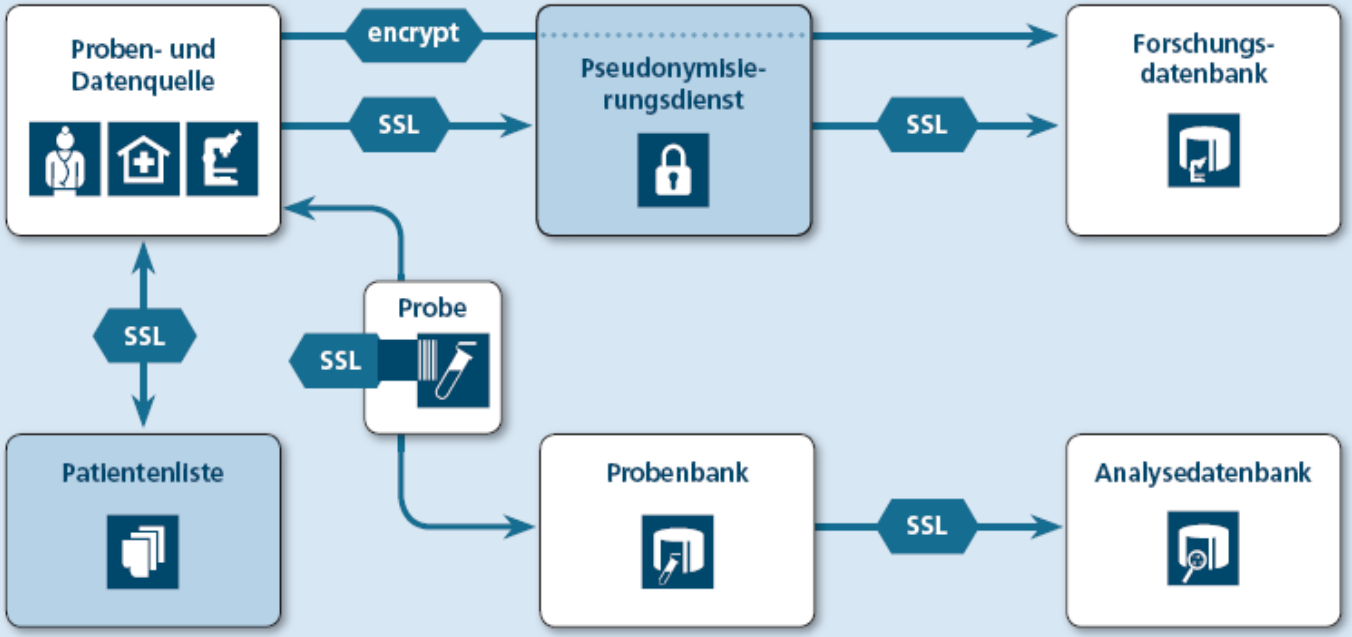
2. PSD-Service:
   - tauscht das einstufige Pseudonym (PID) gegen dessen symmetrisch verschlüsselte Entsprechung (PSN)

3. FDB-Service:
   - entschlüsselt die medizinischen Daten (mit seinem privaten Schlüssel)
   - übergibt Daten an FDB (SQL oder Dateischnittstelle: XML, ASCII, CDISC ODM/SDTM, HL7)
   - bei De-Pseudonymisierung: Request, Verschlüsselung (analog)

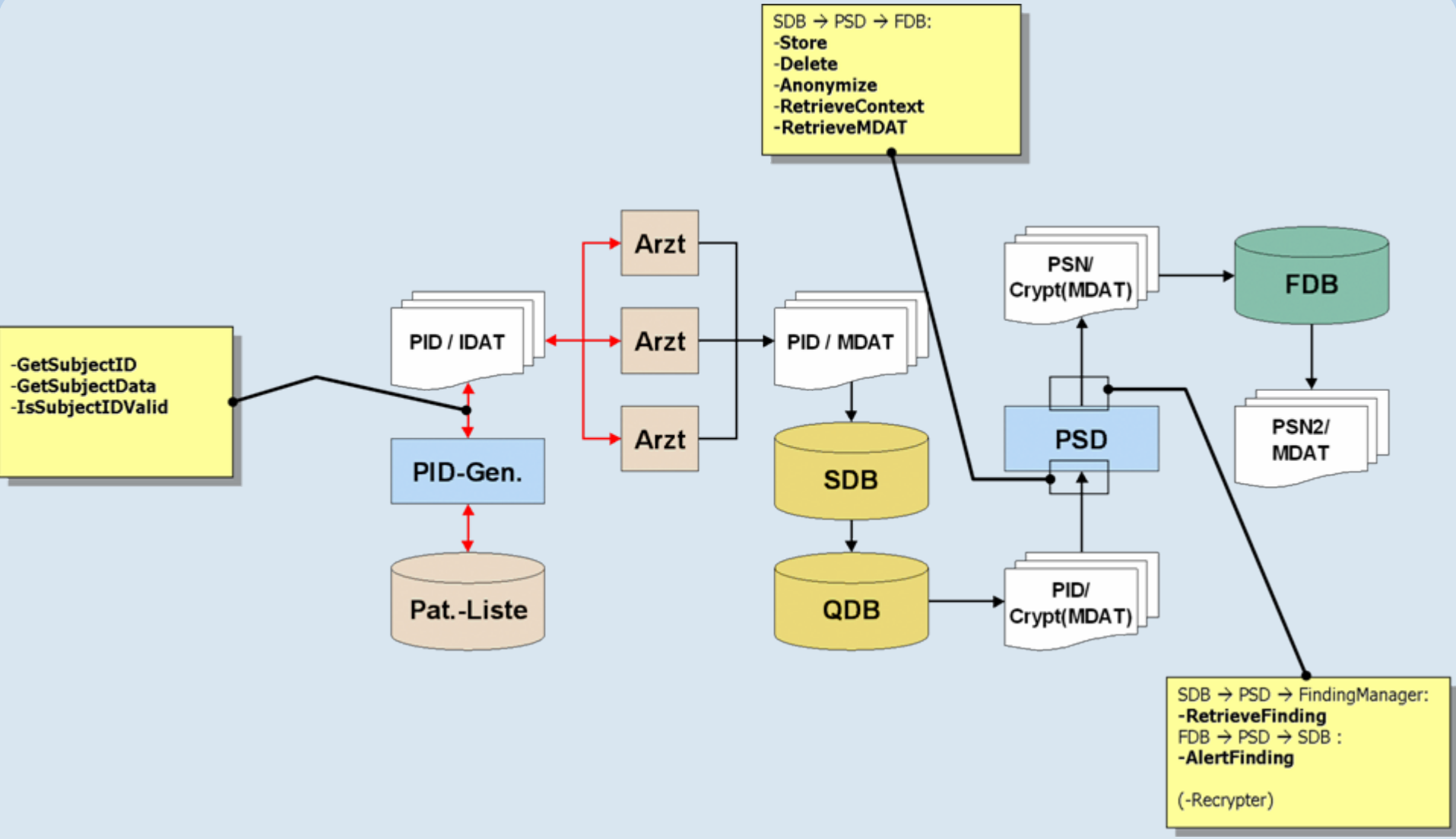
weitere Komponenten: Finding Manager, Recrypter

Security

- HTTPS-Kommunikation mit SSL-Verschlüsselung und Client-Zertifikaten (zur beidseitigen Authentifizierung)
- asymmetrische Verschlüsselung der MDAT zwischen SDB und FDB (Software-Zertifikate oder komponentengebunden Chipkarten) und Signatur des Datensatzes
- Smartcard-basierter Pseudonymisierungsalgorithmus (AES)
- separate Crypter- und Recrypter-Komponenten



Webservice-Funktionen der Komponenten im Pseudonymisierungsdienst der TMF



Anwendungsfälle

- Register
- Kohorten, epidemiologische Studien
- klinische Studien nach AMG / MPG
  - Kopplung an Studienmanagement-Systeme (RDE/EDC)
- Bio(material)banken

aber auch ähnliche Anwendungsfälle in der Versorgung:

- sektorenübergreifende Elektronische Patientenakten
- Versorgungsforschung

Nicht immer werden alle Funktionen gebraucht !

Heterogene Anforderungen und Systemarchitekturen !

Zusammenfassung

- Erarbeitung „Generischer Datenschutzkonzepte“ (2003) incl. Biomaterialbanken (2006) und „Leitfaden & Checkliste zur Patienteneinwilligung“
- in Zusammenarbeit und Abstimmung mit den 17 Datenschutzbeauftragten der Länder und des Bundes
- dient erfolgreich als „Vorlage“ für die spezifischen Datenschutzkonzepte der vernetzten medizinischen Forschung
- Verkürzung und Verschlankung der Genehmigungsverfahren (normalerweise: ca. 1,5 Jahre – Reduktion auf ca. ½ Jahr)
- IT-Lösungen zur Anonymisierung und Pseudonymisierung von Patientendaten (Pseudonymisierungsdienst) über TMF verfügbar.
- Komponentenorientierung und die technische Umsetzung als Web Services, entsprechend einer Service Oriented Architecture, haben die Integrationsaufwände dieses komplexen Lösungsansatzes in sehr heterogene Anwendungsumgebungen erheblich gemindert.

Wer ist die TMF ?

Telematikplattform für Medizinische Forschungsnetze

- Dachorganisation der medizinischen Forschungsverbünde
- gefördert vom BMBF
- 1999 parallel gegründet zur Etablierung der
  - Kompetenznetze in der Medizin (KN)
  - Koordinierungszentren für klinische Studien (KKS)
- Parallelinitiative zur Gesundheitstelematik im Versorgungsbereich (ATG, bit4health-Projekte etc.)
- nationale Zentralinstanz für die vernetzte medizinische Forschung – **eine der wenigen zentralen Telematik-Institutionen im deutschen Gesundheitswesen**
- „Jahresumsatz" 2,5 Mio €

gefördert vom Bundesministerium für Bildung und Forschung

Vielen Dank für Ihre Aufmerksamkeit!

Weitere Informationen:

http://www.tmf-ev.de/