Das revidierte generische Datenschutzkonzept der TMF: Neue Anforderungen an Tools zum ID-Management

TMF-Workshop „ID-Tools“

Mainz, 10. Mai 2012

Prof. Dr. Klaus Pommerening

Universitätsmedizin Mainz, IMBEI

TMF-AG Datenschutz

Warum Pseudonyme?

Verbergen von Identitäten

- Hier: von Patienten/ Probanden in medizinischer Forschung
- ⇒ Identitätsmanagement

Gute Pseudonyme: Zufallswerte, evtl. laufende Nummer

- Keine Information über Herkunft oder Zugehörigkeit preisgeben.
- Auch Information über Daten-/ Probenquelle kann kritisch sein.
- Zuordnung bekannt nur *entweder* beim DTH *oder* im lokalen Behandlungskontext [oder beim „Objekt"].
- Für andere Empfänger Daten (zumindest formal) anonym.

Ungeeignet als Pseudonyme:

- Initialen + Teile des Geburtsdatums,
- Nummer, die einem größeren Personenkreis bekannt ist (Fall-Nr. im KIS, Versicherungsnummer, …).

Warum mehrere verschiedene Pseudonyme?

Verschiedene Datensätze sollen nicht unbefugt verknüpft werden können (Untraceability).

Ein Pseudonym ist umso schwächer, je mehr Daten mit ihm verbunden sind.
(Hochdimensionale Datensätze faktisch nicht anonym.)

In verschiedenen Typen medizinischer Forschungsprojekte gibt es verschiedene rechtliche Rahmenbedingungen,
auch zum Umgang mit Pseudonymen.

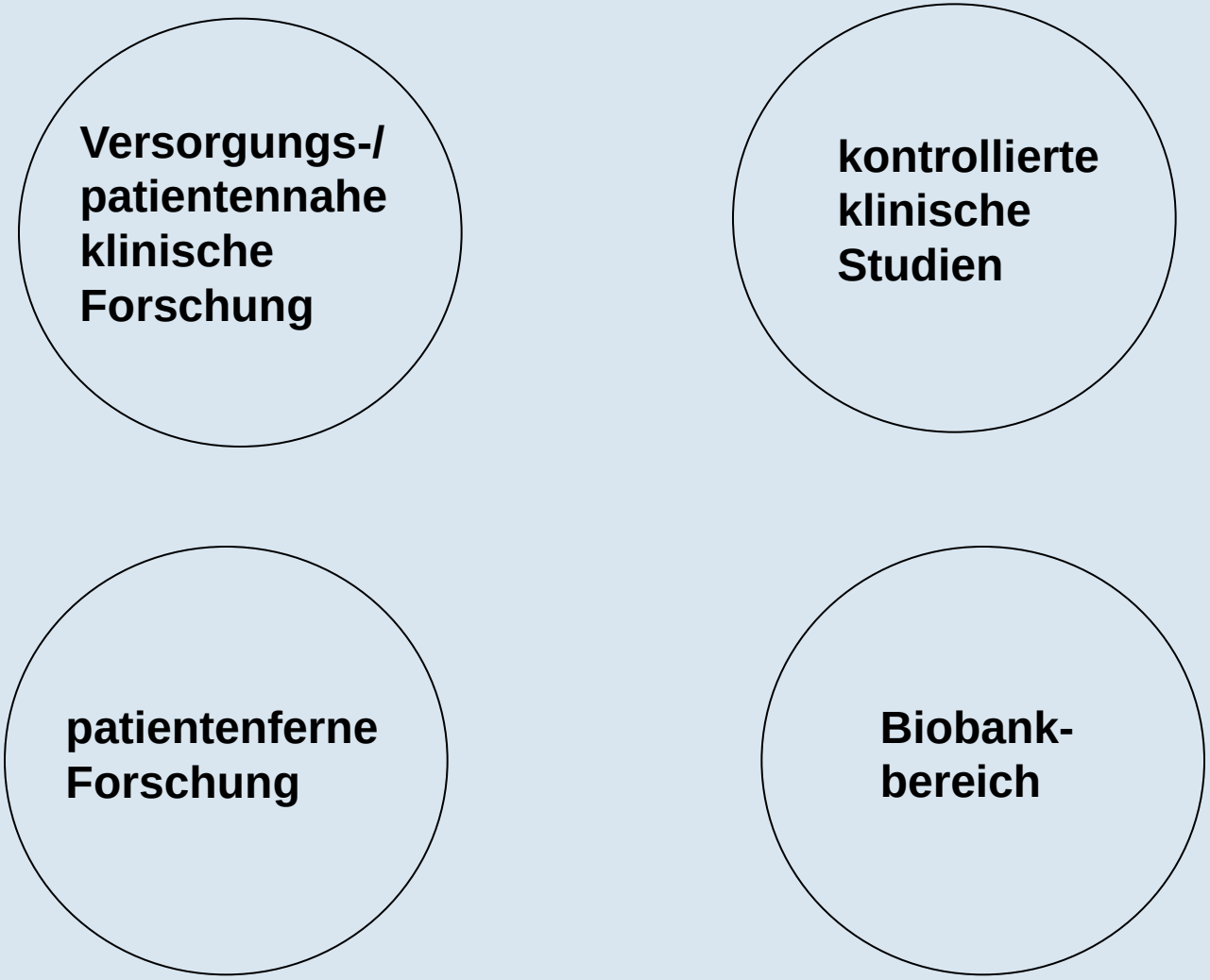
Reidentifizierungsrisiko durch Proben

Genetische Informationen in Materialien stets personenbeziehbar.

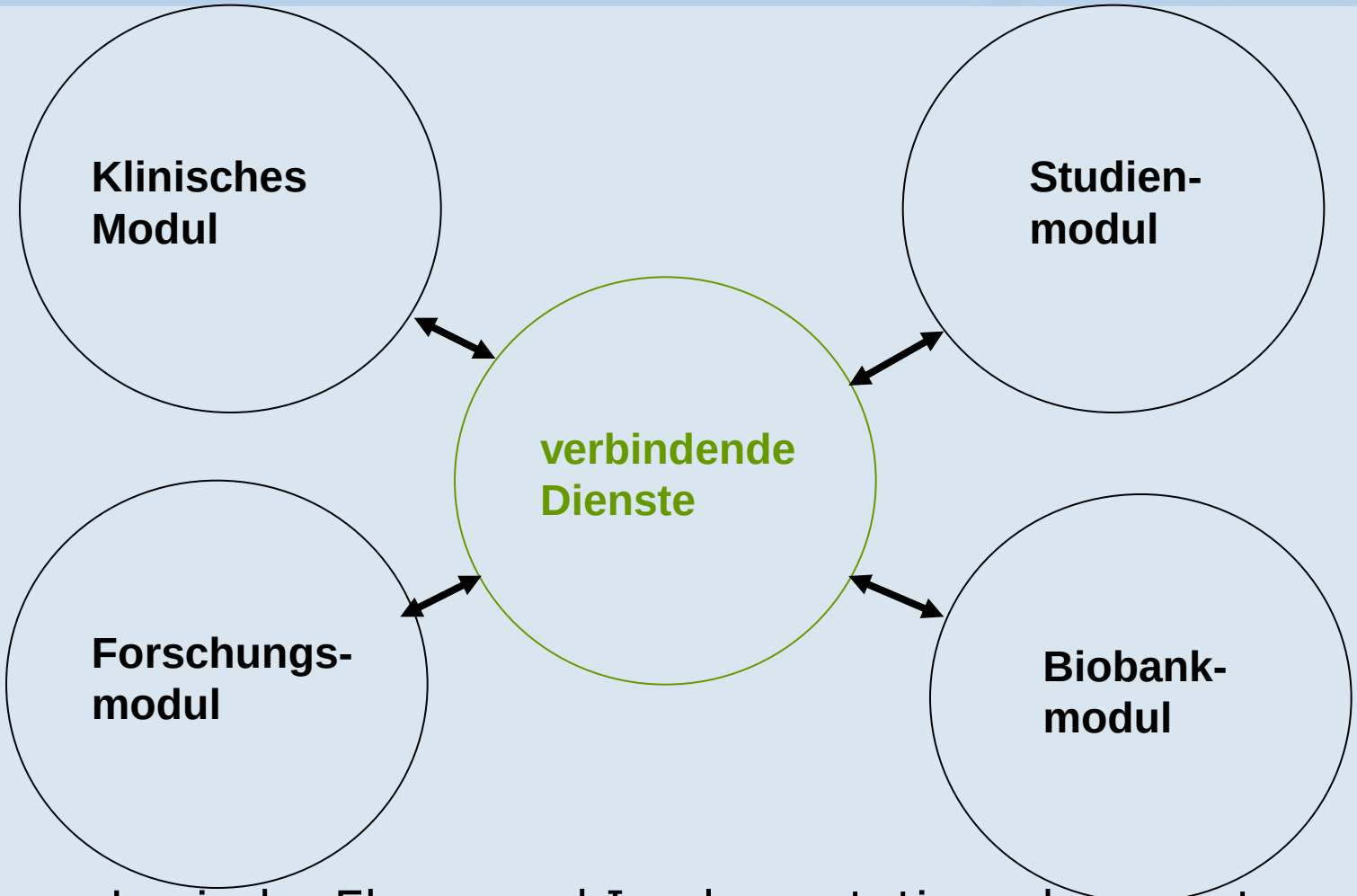
- Verknüpfung der Proben mit krankheitsbezogenen, soziodemographischen und anderen Daten –
  - über Referenzproben unbefugte Reidentifizierung möglich.
- Biomaterialien in der Regel nur mit klinischer Annotation nützlich.
  - Diese Daten werden einem Reidentifizierungsrisiko ausgesetzt.

Daher Proben und Daten durch Pseudonyme trennen.

4 Bereiche der medizinischen Forschung mit unterschiedlichem rechtlichem Rahmen



Mögliche Module eines Forschungsverbunds



Achtung: Logische Ebene und Implementationsebene unterscheiden.

Für die verbindenden Dienste gibt es mehrere Optionen (zentral, verteilt, dezentral/lokal).

Das revidierte (künftige) Modell

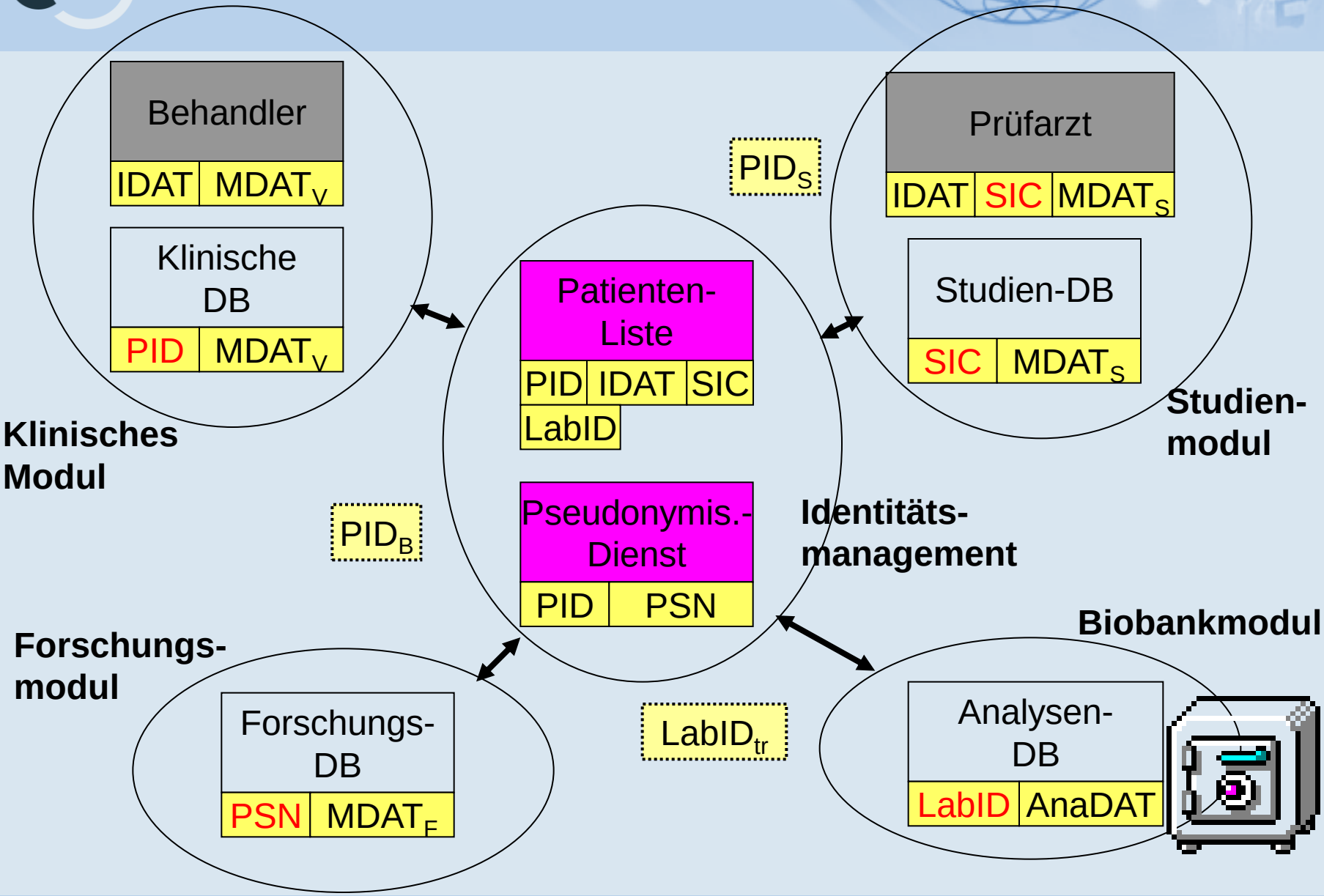
Im Maximalmodell vier Module mit unterschiedlichen rechtlichen Rahmenbedingungen:

- Klinisches Modul (↔ Modell A),
- Studienmodul (↔ klin. Studien),
- Forschungsmodul (↔ Modell B),
- Biobankmodul (↔ Modell BMB).

In jedem Modul eigenes Pseudonymisierungsschema
→ informationelle Gewaltenteilung.

Jedes Modul kann mehrere Datenbanken enthalten
(z. B. verschiedene SDBn im Studienmodul).

Module und Pseudonyme



Was soll das Identitätsmanagement alles können?

Widerprüchliche Anforderungen:

- Daten zum selben Individuum korrekt zuordnen,
- Identität vor Unberechtigten verbergen.

Identitätsmanagement

- verwaltet Zuordnung zwischen Pseudonymen und Identitäten auch fehlertolerant (→ Daten-QM)
- und zwischen verschiedenen Pseudonymen,
- wirkt (erforderlichenfalls) bei der Kontaktierung mit,
- hilft bei der Berücksichtigung von Einwilligungsoptionen und Anonymisierung nach Ablauf der eingewilligten Frist (→ Einwilligungsmnagement).

Z. B. durch Führung einer Patientenliste/ CRM*-Software.

Die Aufteilung der Pseudonyme auf die zentralen Dienste in der Grafik ist exemplarisch.
Exportpseudonyme ad-hoc, nicht im IDmgt.

* CRM = Customer Relationship Management

Identitätsmanagement: Anwendungsfälle

a) Anmeldung eines Patienten/ Studienteilnehmers
b) Datenübertragung klinisches Modul → Forschungsmodul
c) Datenübertragung Studienmodul → Forschungsmodul
d) Rückmeldung von Ergebnissen (Findings)
e) Anfragen eines Patienten/ Studienteilnehmers
f) Depseudonymisierung zur DQ-Sicherung
g) Rekrutierung
h) Widerruf mit Löschung
i) Widerruf mit Anonymisierung
j) Todesfall
k) Umpseudonymisierung

Identitätsmanagement: Überschneidungen

Datenqualitätsmanagement:

- fehlertolerantes Matchen
- Despeudonymisierung für Rückfragen
- Datenabgleich über verschiedene Module

Rechtemanagement:

- Nutzerzugriff auf Patientenliste/ PID-Generator
- Zugehörigkeit Patient zu Arzt/ Einrichtung
- Ticket-Handling

PID-Generator: Verwaltung von Pseudonymen

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

- Für Studiendatenbanken (SIC oder PID_S), genetische Analysen (LabID, $LabID_{tr}$), evtl. Bilddatenbank (PID_B).
- PID_V *oder* PID_S als Input an Pseudonymisierungsdienst senden.
- Umwandlung verschiedener pseudonymer Kennungen ineinander.
- Entgegennahme und Verwaltung auch extern erzeugter Kennungen (z. B. SIC i. d. R. in Studiensoftware erzeugt).

PID-Generator: Kommunikation und Schnittstellen

Einbindung in den Datenfluss von klinischem Modul („Modell A“) und Studienmodul
und Kommunikation mit Pseudonymisierungsdienst:

- Ausgabe geeigneter Zugriffstickets,
- Ticket-Vergabe oder PID-Rückgabe als Alternativen.
- Kommunikation mit KDB und SDB bzw. den dort angesiedelten Systemkomponenten des Pseudonymisierungsdienstes.

Dazu $OrgDat_{PL}$ benötigt: Kontext der meldenden Stelle + Datum (für Kontakt-Mgt., Rechte-Mgt.)

Erleichterte Einbindung in EDC-Software (auch DS-Bibliothek „Modell A“ und CDW-Software:

- Überarbeitung und Erweiterung der SOAP-Schnittstelle,
- verschiedene Konfigurationsoptionen.
- „PID-Dispatcher“ verallgemeinern.

PID-Generator: Administration

Web-Oberfläche (statt Kommandozeilen-Bedienung) zur

- Konfiguration,
- Auswertung.

(für revidiertes DS-Konzept nicht entscheidend)

Log-Funktionen verbessern:

- Alarme.
- Automatische Auswertung.

(für revidiertes DS-Konzept relevant)

PID-Generator: Interne Funktionen

Internationalisierung

- Zeichensätze: Transkription, Unicode,
- Phonetik.

Alternative Match-Verfahren zur Auswahl anbieten –
auch in Kombination –,

- z. B. stochastisches Matchen,
- verbesserte Algorithmen von Jörg Michael.

(Beides für revidiertes DS-Konzept nicht entscheidend.)

Pseudonymisierungsdienst

Ansprache und Nutzung des Pseudonymisierungsdienstes aus unterschiedlichen Ausgangskomponenten.

- Anstoß aus KDB wie aus SDB zum Datenexport an FDB.
  - (Aktuelle Implementierung passt nur zu SDB.)
- Folgende Komponenten müssten angepasst werden:
  - PSD-Service: Ansprache nicht nur vom SDB-Service und FDB-Service,
    sondern auch vom (neu zu konzipierenden) VDB-Service.
  - VDB-Service: muss neu implementiert werden (analog zum bestehenden SDB-Service).
    - Wegen der unterschiedlichen Handhabung des PID_V (nicht in der VDB bekannt) muss in diese Komponente auch eine Kommunikation mit der Patientenliste, insbesondere die Handhabung von Zugriffstickets (TKT), eingebaut werden.
- Umgang mit Tickets („Vorbeileitung“ von Daten zur „Bandbreitenschonung“)
- Umschlüsselung multipler IDs (z. B. Dicom-Header)

Prioritäten (aus Sicht des revidierten DS-Konzepts)

1. Umgang mit mehreren Pseudonymen (auch extern erzeugten)
   - Erweiterung des DB-Schemas im PID-Generator
   - Umwandlungsfunktionen
   - Schnittstellen im PSD einschl. Ticket-Handling
2. Verbesserung der Schnittstellen zu EDC-Software
   - kommerziell, DS-Bibl. „A“, CDW
   - einschl. Verwaltung der nötigen Zugriffstickets
3. CRM-Funktionalität: erst mal Kompatibilität/ Interoperabilität mit CRM-Software prüfen (Marktanalyse)
4. Einwilligungsmanagement
   - einschl. Anonymisierungs-/ Sperr-/ Löschauftrag.
5. Admin-Oberfläche des PID-Generators
6. Internationalisierung des PID-Generators
   - einschl. alternativen Match-Verfahren
7. verbesserte Handhabung des PSD
8. Pseudonymisierung von ADAT

Separate Schiene: Anonymisierungswerkzeuge