Datenschutz in der medizinischen Forschung

GMDS-Jahrestagung Mainz, 28. September 2011

Univ.-Prof. Dr. Klaus Pommerening
Universitätsmedizin Mainz, IMBEI

Dr. Johannes Drepper
TMF e. V. Berlin

Gefördert vom
Bundesministerium für Bildung und Forschung

Inhalt

1. Szenario und Anwendungsfälle
2. Rechtliche Grundlagen der medizinischen Forschung
3. Anonymisierung und Pseudonymisierung
4. Patienteninformation und Einwilligungserklärung
5. **Die Datenschutzkonzepte der TMF**
6. Praktisches Vorgehen

Was ist ein Datenschutzkonzept?

Beschreibung aller Maßnahmen, die zum wirksamen Schutz von personenbezogenen Daten beitragen, insbesondere wie das RI-Risiko minimiert und kontrolliert werden soll.

Ausgangspunkte:

- Projektziele
- organisatorische Strukturen, Prozesse und Kommunikationswege
- IT-Konzept: Daten, Datenspeicher, Datenflüsse, Datenverwendung

Begründung von „Erforderlichkeit“ und „Angemessenheit“

Allgemeine Maßnahmenbereiche

- Definition verbindlicher organisatorischer Strukturen
  (z. B. als Hochschuleinrichtung* oder e. V.)
  mit rechtlich verbindlichen Regeln
  (z. B. Verträge, Policies, SOPs)
- Patienten-/ Probanden-/ Spender-Aufklärung und -einwilligung
  - Einwilligungsmanagement
- Rechtemanagent,
  insb. Kontrolle der Zugriffe,
  der Exporte und der Datennutzung.
- IT-Sicherheit

* Hier ist die Universität oder Universitätsklinik die rechtsfähige juristische Person.

Was ist ein generisches Datenschutzkonzept?

Vorlage zur Erstellung eines konkreten DS-Konzepts, die möglichst viele Anwendungsbereiche abdeckt.

- Allgemein gültige Regeln
- Daraus abgeleitete zwingende Maßnahmen oder Optionen
- Kriterien zur Verhältnismäßigkeit

Anwendung durch Adaption an die konkrete Situation („Mapping“ generischer Objekte auf konkrete), Identifikation von notwendigen Abweichungen oder Varianten

Wichtig also:

- Vollständigkeit
- Flexibilität
- Leichte Adaptierbarkeit

Beispiel: Ausschuss Datenschutz

Generisch definiertes Gremium;

- entscheidet über Verwendung und Weitergabe von Daten unter Berücksichtigung des Datenschutzkonzepts und der Policies,
- besteht aus z. B. Vertretern von Lenkungsausschuss, Datenschutzbeauftragtem, wissenschaftlichem Beirat.

Konkretisierung:

Organisatorische Einbindung, Kommunikationswege und Befugnisse verbindlich in der Satzung festlegen.
Personen oder Rolleninhaber benennen.

Versionen des generischen TMF-Datenschutzkonzepts

Erste Version (2003) mit zwei alternativen Modellen für typische Netze:

- Modell A (versorgungsnahe zentrale Datenbank) (Beobachtungsstudie bei chronischen Erkrankungen)
- Modell B (versorgungsferne zentrale Datenbank) (Kohorte oder epidemiologische Langzeitstudie)

Ergänzt (2006) durch Modell BMB

- = Datenschutzkonzept für Bio(material)banken

Ergänzt durch Leitfaden, Checkliste und Online-Assistent zur Patientenaufklärung und -einwilligung.

Basis: Gutachten führender Medizinrechtler.

Konsens vom AK Wissenschaft der Datenschutzbeauftragten.

In vielen Netzen adaptiert und implementiert

- mit mehr oder minder großem Anpassungsaufwand.

Revision des TMF-Datenschutzkonzepts

In Arbeit: Revision,

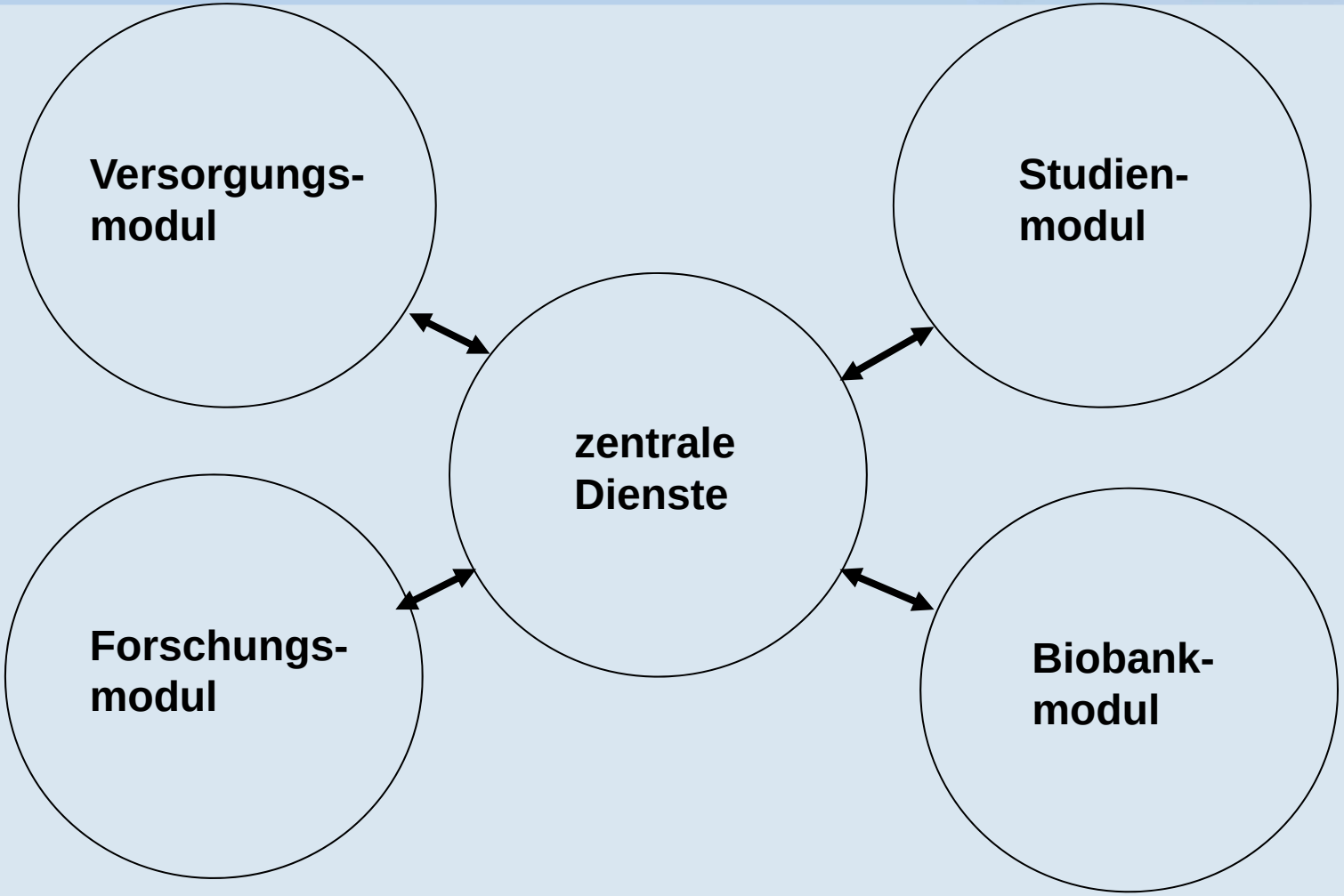
- einheitliches Modell mit modularer, skalierbarer Netzarchitektur.
- Ziele: breitere Anwendbarkeit, leichtere Umsetzbarkeit.

Die allgemeinen Bereiche mit unterschiedlichen rechtlichen Rahmenbedingungen werden in entsprechenden Modulen zusammengefasst.

Die Einteilung in Module dient als Ansatz zur informationellen Gewaltenteilung.

Die Module kooperieren miteinander über zentrale Dienste und Komponenten (z. B. ID-Management).

Mögliche Module eines Forschungsverbunds



Benötigt: DS-Konzepte für die einzelnen Bereiche („Module“)
DS-Konzept für das Gesamtszenario („Maximalmodell“)

TMF-Modell A
(↔ Versorgungsmodul* + Zubehör)

Versorgungsnahe Datenbank („Versorgungs-Datenbank“*)

- Zweck: Forschung im direkten Patientenbezug,
- Möglicherweise, aber nicht notwendig direkte Rückwirkung auf die Behandlung,
  insbesondere *kein Ersatz für Patientenakte*.
- Datenqualitätssicherung „an der Quelle“.
- Ansatz: pseudonyme Speicherung,
  personenbezogener Zugriff für behandelnde Ärzte.
- Kein Direktzugriff für Forschungsprojekte,
  - statt dessen Datenexport.

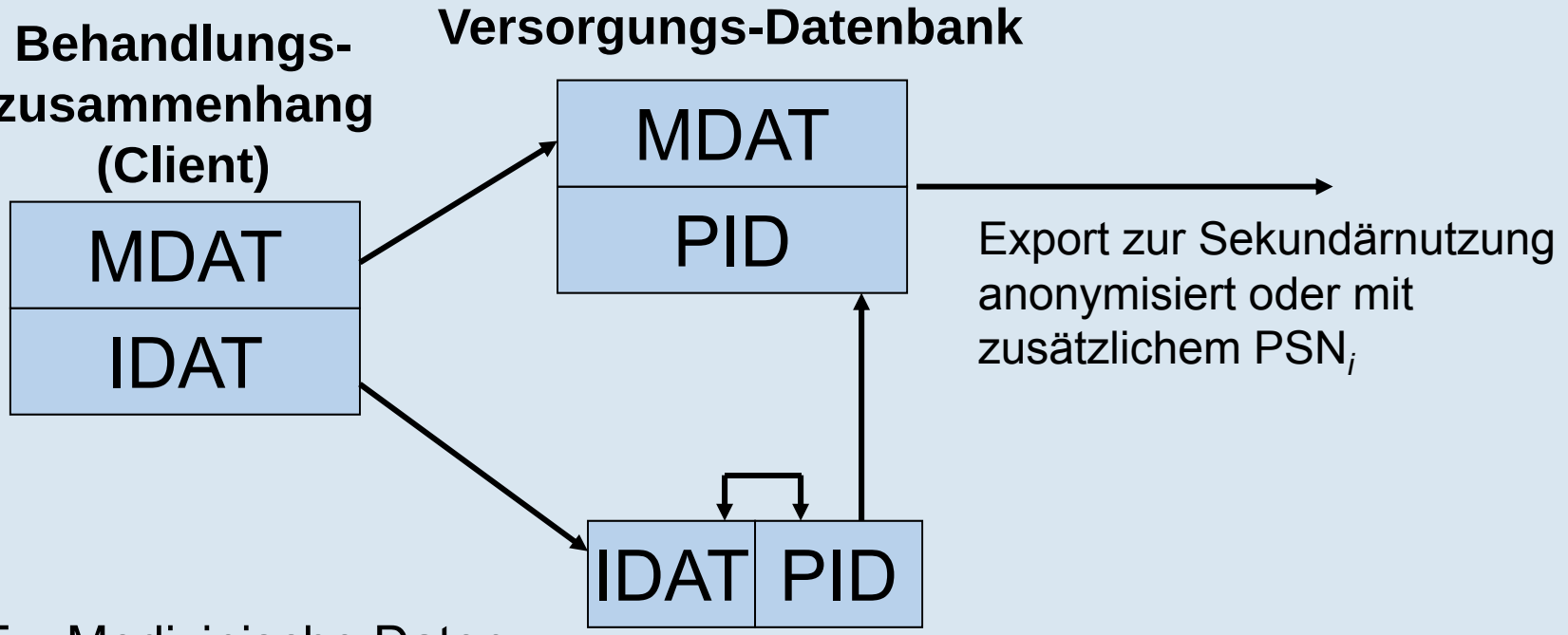
Pseudonym (hier PID genannt) ist nur in DB und TTP bekannt.

Zugriff über (temporäres) Zugriffsticket geregelt.

MDAT und IDAT kommen *nur beim behandelnden Arzt* zusammen.

* Achtung: Bezeichnung wird wg. Missverständnis-Potenzial möglicherweise noch geändert.

TMF-Modell A



MDAT = Medizinische Daten
IDAT = Identitätsdaten

PID = (pseudonymer) Patientenidentifikator
PSN$_i$ = Pseudonym für Export *i*.

TTP = Trusted Third Party (Treuhänderdienst)

Anforderungen an EDC-Systeme im Versorgungsmodul

IDAT und MDAT dürfen nur auf Bildschirm eines Behandelnden gemeinsam sichtbar sein.

- Technisch: Kommunikation mit zwei Datenbanken.
- Achtung: Cross-Site-Scripting oft gesperrt.

IDAT und PID nur in Patientenliste gemeinsam sichtbar.

PID nur in Patientenliste und Datenbank bekannt

- nicht für Export verwendet!

Notwendige IT-Sicherheitsmaßnahmen:

- Authentisierung,
- Zugriffsregelungen (Rollen-/ Rechtekonzept),
- verschlüsselte Übertragung (→SSL als Standard-Technik mit selbsterzeugtem Server-Zertifikat vorläufig ausreichend),
- *Achtung*: Einfacher Passwortschutz ist ungenügend.
- lokale Sicherheit auf Servern und Client-Rechnern.

*Unter diesen Voraussetzungen Internet datenschutzkonform für EDC nutzbar, auch bei Browser-basiertem Betrieb**.

* Software-Bibliothek bei TMF erhältlich

TMF-Modell B
(↔ Forschungsmodul + Zubehör)

Versorgungsferne Datenbank („Forschungs-Datenbank“),

- Zweck: Forschung ohne direkten Patientenbezug,
- Ansatz: Speicherung und Zugriff pseudonym.
- Vom Client nur Datenübermittlung, sonst kein Zugriff.

TTP „Patientenliste“ sorgt für eindeutige Zuordnung von Daten aus verschiedenen Quellen durch Vergabe eines PID,

- auch, wenn IDAT fehlerhaft (Teil des Datenqualitätsmanagements).

TTP „Pseudonymisierungsdienst“ verschlüsselt PID zu PSN.

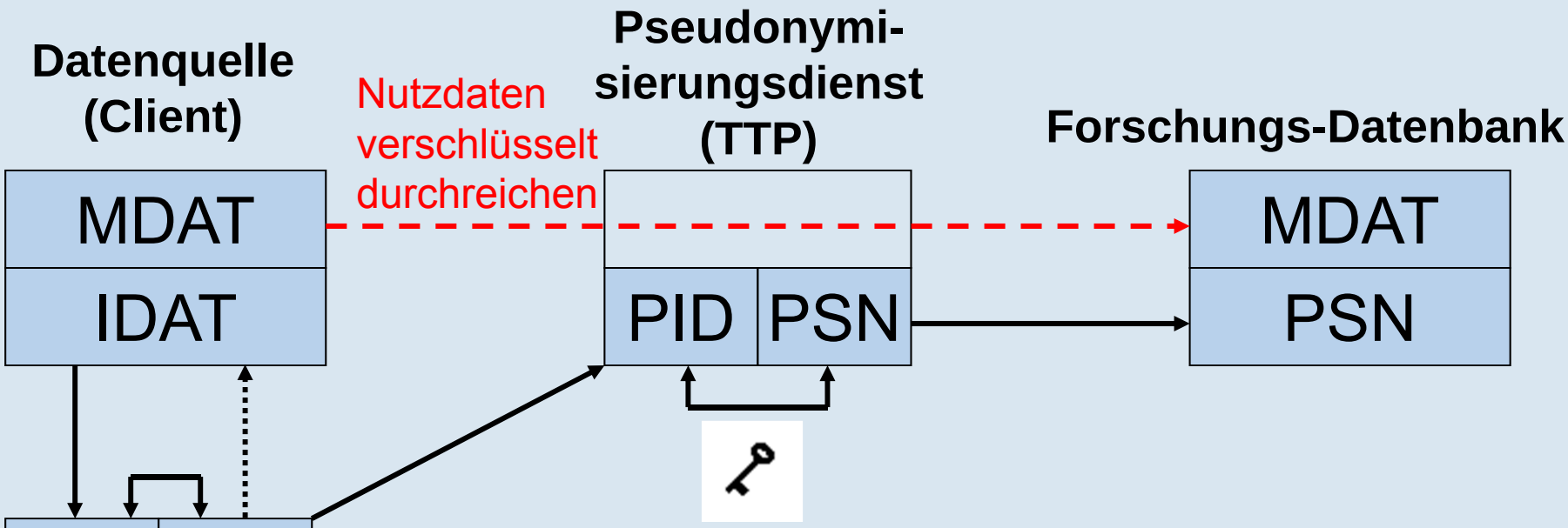
- Schlanke TTP, speichert nur Schlüssel.

MDAT werden (asymmetrisch) verschlüsselt durchgereicht

- oder über Einmal-Ticket zugeordnet.

Datenqualitätssicherung erfordert gesondert aufgesetzten Prozess (Rückfragen von Datenbank an Datenquelle), evtl. mit temporärer Depseudonymisierung.

TMF-Modell B



MDAT = Medizinische Daten
IDAT = Identitätsdaten

PID = Patientenidentifikator
PSN = Pseudonym

TMF-Modell BMB (↔ Biobankmodul + Zubehör)

Zweck: Sammlung von Proben in Biobank(en) für Forschungszwecke, insb. genetische Forschung.

Physische Trennung von Material und Daten.

Trennung von Analysedaten und medizinischen Daten („Annotation“).

Trennung von IDAT und MDAT wie Modell B (oder A).

Pseudonymisierungsdienst PID ↔ PSN wie Modell B.

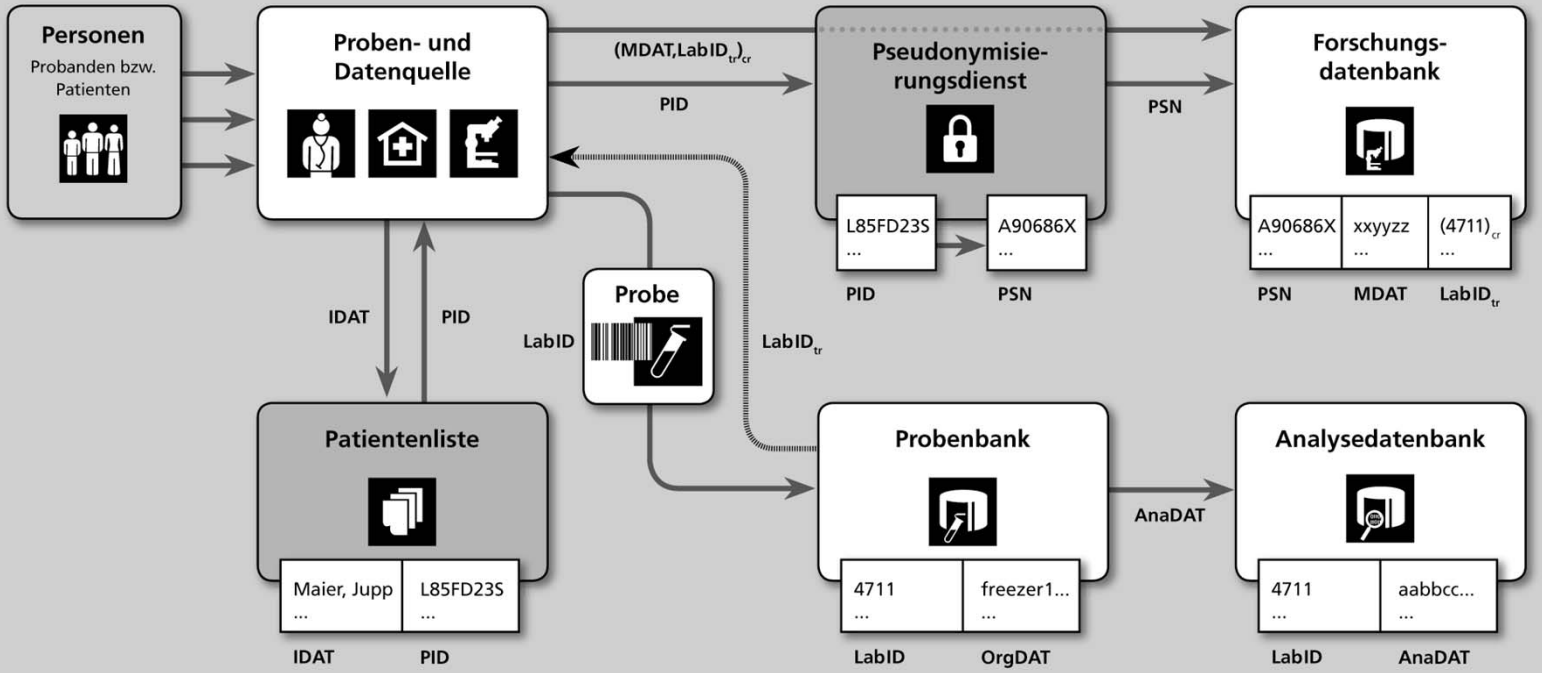
Pseudonyme Probenkennzeichnung LabID in Probenbank.

Verweis auf Probe in MDAT verschlüsselt: $LabID_{tr}$.

Ausführliche Diskussion der Verhältnismäßigkeit.

In der Annotations-DB („Forschungs-DB)
sind sowohl Spender (PSN)
als auch Proben ($LabID_{tr}$) pseudonymisiert.

TMF-Modell BMB



(Erweiterung von TMF-Modell B: MDAT in patientenferner Forschungsdatenbank - Adaption auf andere Szenarien möglich.)

Das revidierte (künftige) Modell

Vereinigung der Modelle A, B und BMB

- mit Erweiterung für klinische Studien.

Im „Maximalmodell“ vier Bereiche („Module“) mit unterschiedlichen rechtlichen Rahmenbedingungen und separater Verantwortlichkeit:

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

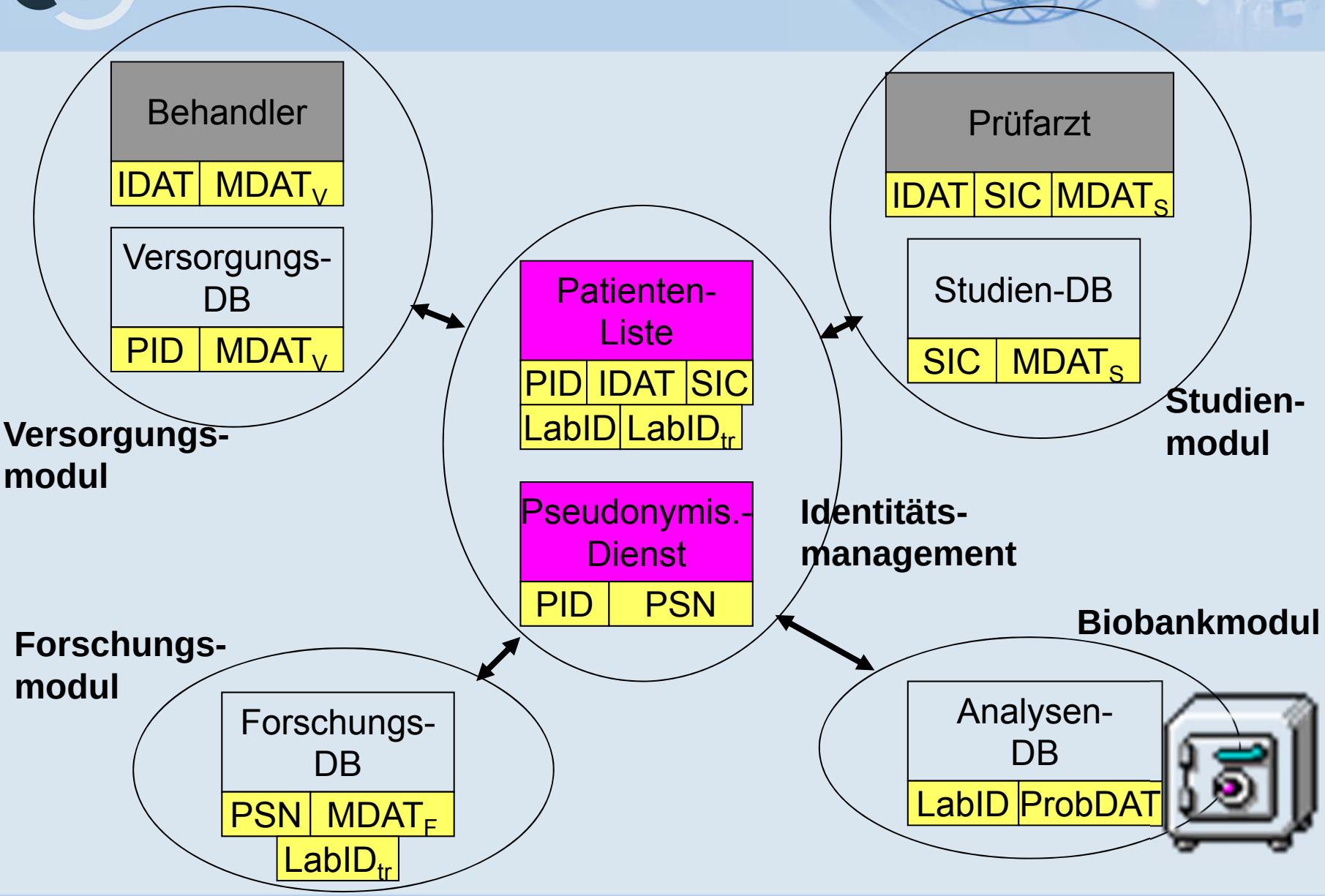
In jedem Modul eigenes Pseudonymisierungsschema
→informationelle Gewaltenteilung.

Jedes Modul kann mehrere Datenbanken enthalten
(z. B. verschiedene Studien-DBn im Studienmodul).

Pseudonym- (und ggf. IDAT-) Zuordnung im zentralen Dienst „Identitätsmanagement“ mit den Komponenten

- Patientenliste,
- Pseudonymisierungsdienst.

Module und Pseudonyme



SIC als Pseudonym in klinischen Studien

SIC = Subject Identification Code

als Pseudonym im AMG definiert.

Falls mehrere unabhängige Studien im Studien-Modul:

evtl. jeweils eigene SIC-Vergabe,
im ID-Mgt über PID_S zusammengeführt.

(Achtung: $PID_S \neq PID_V$ – die Pseudonyme in den Modulen sind verschieden und können nur vom ID-Mgt zusammengeführt werden.)

Zentrale Komponenten im Datenschutzkonzept

- Identitätsmanagement
- Rechtemanagement
- Einwilligungsmanagement
- Datenqualitätsmanagement

Zentral oder dezentral, aber kooperierend
verschiedene Realisierungsoptionen nach Verhältnismäßigkeit

Durch zentral vorgegebene Policies gesteuert

Identitätsmanagement von Patienten / Spendern

Das Identitätsmanagement sorgt für

- Eindeutige Identifizierung (richtige Zuordnung aus verschiedenen Quellen auch bei fehlerhaften IDAT),
  - repräsentiert durch die Patientenliste und PID,
  - gute Basis für dezentrale Strukturen.
- Wahrung der Vertraulichkeit der Identität
  - repräsentiert durch Patientenliste und Pseudonymisierungsdienst sowie verschiedene Pseudonyme.

Pseudonyme: PID_V, PID_S, SIC, PSN, LabID, $LabID_{tr}$.
Dazu noch „Exportpseudonyme“ PSN_i.

Die Aufteilung der Pseudonyme auf die zentralen Dienste in der Grafik ist exemplarisch.

Mindestens einer der Dienste sollte bei einem Datentreuhänder angesiedelt sein.

Was ist ein Datentreuhänder? (DTH)

Unabhängige Instanz, die Daten verwahrt
(hier: Zuordnungsregeln für Pseudonyme)
und das Vertrauen aller Beteiligten genießt.

Geeignet, aber selten (und teuer): Notar.
Achtung: Beschlagnahmeschutz *nicht* gesichert.

Pragmatische Lösung: Rechenzentrum oder MI-Institut.
Wichtig: DTH ist regelgebunden, darf aber nicht weisungsgebunden gegenüber anderen Modulen des Forschungsverbunds sein.

Vertragliche Regelungen und SOPs notwendig.

Insbesondere Depseudonymisierung nur nach verbindlichem Regelwerk.

Achtung: Unterauftrag reicht nicht, der DTH muss „gleichberechtigter“ Partner sein.
Rechtsfähig ist Universität bzw. Universitätsklinikum.

Beispiel: Für einen Treuhänderdienst am IMBEI schließt das KN XY einen Vertrag mit der Universitätsmedizin Mainz.

Werkzeuge zum ID-Management: Der PID-Generator der TMF

- Entwickelt am IMBEI Mainz für das Kompetenznetz POH.
- Führung der Patientenliste in einer Datenbank
  - mit IDAT und (einem) PID.
- Fehlertolerante Zuordnung von IDAT.
- PID als 8-Zeichen-Code mit Fehlerkorrektur.
- Betriebsarten: Web-Service, SOAP-Schnittstelle, Batch-Modus.
- Weitgehende Konfigurierbarkeit.
- Installation unter Linux und MS-Windows.
- Erhältlich bei der TMF.

Werkzeuge zum ID-Management: Der Pseudonymisierungsdienst der TMF

- Web-Dienst
- Schnittstellen für Datenquelle und Forschungsdatenbank
- kryptographische Verschlüsselung mit Smart-Card

PID ↔ PSN

Erhältlich über TMF.

Update von PID-Generator und Pseudonymisierungsdienst
als Projekt beantragt.
(Zur Anpassung an Revision des DS-Konzepts)

Einwilligungsmanagement

In Forschungsverbünden mit mehreren Projekten oder unbestimmten Auswertungsmöglichkeiten kann die Einwilligung abgestuft oder selektiv sein.

Bei Datenverwendung dürfen nur Fälle eingeschlossen werden, für die eine passende Einwilligung vorliegt.

Empfohlen dafür: Zentraler Online-Dienst mit zentraler oder verteilter Haltung von „Einwilligungsdatensätzen“.

Standardisierte Erfassung nach zentralen Vorgaben nötig.

Achtung: Die Speicherung desselben differenzierten Einwilligungsdatensatzes in verschiedenen Modulen kann die Pseudonymtrennung aushebeln.

Regelungen für Nutzung und Weitergabe

- Projektprüfung und ggf. -zulassung durch „Ausschuss Datenschutz“ des Forschungsverbunds.
- Einschränkungen des Empfängerkreises aus Einwilligung beachten.
- Datensparsamkeit beachten:
  - Nur die benötigten Daten herausgeben
  - Keine Herausgabe der „internen“ Pseudonyme PID, PSN, $LabID_{tr}$
  - Export mit PSN_i
- Verbindliche Vereinbarung mit Empfänger nötig.
  - Keine weitere Weitergabe.
  - Keine Dauerspeicherung, kein Restmaterial aufheben.
  - Keine Reidentifizierungsversuche.

Verhältnismäßigkeit

Wichtiges Prinzip des Datenschutzes.

Sicherheit benötigt Redundanz.
Vorkehrungen für den Fall, dass Sicherheitsmaßnahmen ausfallen oder durchbrochen werden.

Verhältnismäßigkeit bedeutet hier: Einige Redundanzen werden evtl. weggelassen oder durch schwächere Maßnahmen ersetzt.

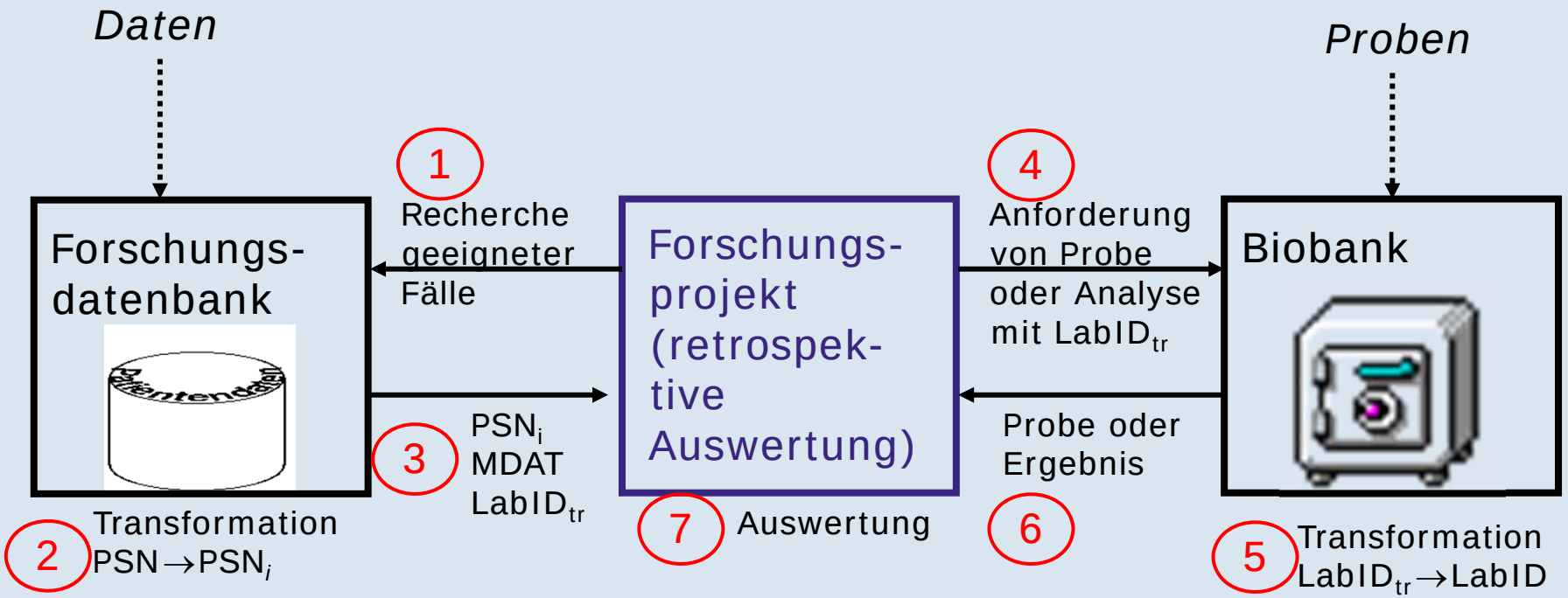
Beispiele:
- Pseudonymisierung an der Quelle (statt DTH) bei „kleinen“ Projekten,
- Teil des ID-Mgt (LabID $\leftrightarrow$ $LabIT_{tr}$) im Biobankmodul, wenn illegale Kooperation mit Forschungsmodul ausgeschlossen.

Kriterien zur Verhältnismäßigkeit

- Größe und Komplexität der Datenbank oder des Netzes,
- Dauer der Aufbewahrung,
- Brisanz der Daten- oder Biobank,
  - z. B. stigmatisierende Krankheit,
  - Attraktivität für Reidentifizierungsversuche.
- Anforderungen von Patientenverbänden/ Interessentengruppen
- Stringenz der Organisation,
  - z. B. Policies, SOPs, etabliertes Monitoring,

Was die IT nicht sichern kann, muss durch organisatorische Maßnahmen abgedeckt werden.

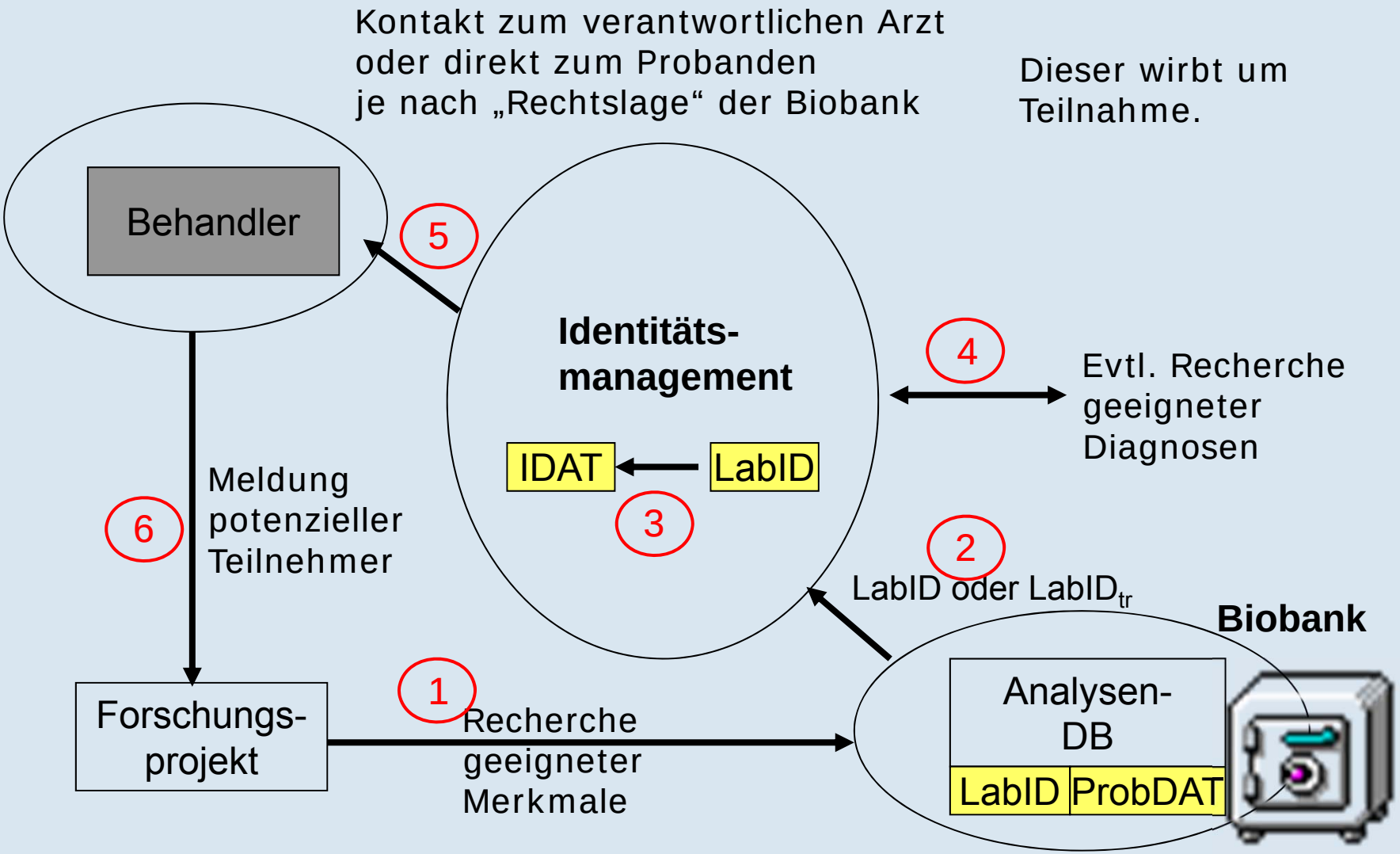
Anwendungsbeispiel: Fallsuche in Biobank für retrospektive Studie



PSN_i = ad-hoc-Pseudonym
(zusätzl. Stufe der Pseudonymisierung –
Sinn: *verschiedene Projekte können nicht unabhängig Daten zusammenführen, Rückfrage bleibt möglich*)

Analog bei Primär-Abfrage der Analysen-DB

Anwendungsbeispiel: Rekrutierung für neue Studien (über genetische Merkmale)



IT-Sicherheit

Ansätze:

- kryptographische Infrastruktur
  - Nutzung von SSL, evtl. auch von Nutzerzertifikaten, vorzugsweise mit Smartcards
  - kryptographische Kommunikation und Datenspeicherung
- Serverhärtung
- lokale Sicherheitsmaßnahmen bei Clienten
- zentrales Rollen- und Rechtemanagement
  - Regelung und Protokollierung von Zugriffen, Vier-Augen-Prinzip
  - Policy Enforcement

Das generische Datenschutzkonzept der TMF für Datensammlungen und Biobanken

- Das („alte“) Konzept hat den Konsens des AK Wissenschaft der Datenschutzbeauftragten.
- Publikation in der TMF-Buchreihe (bei MMV) (für Biobanken in Vorbereitung).
- Das Konzept ist flexibel genug, um angemessenen Spielraum bei der Implementierung zu lassen.
  Es ist kein Königsweg, aber eine entscheidende Hilfe.
- Abschluss der Revision 2011 (?).