6 Organisatorisches und technisches Konzept für Forschungsverbünde

Ein medizinischer Forschungsverbund umfasst in der Regel mehrere oder alle der in Kapitel 5 beschriebenen Module. Durch deren Zusammenwirken ergeben sich komplexe Datenflüsse und Kommunikationsbeziehungen, die erhebliche organisatorische und technische Maßnahmen erfordern, auch zur Wahrung eines angemessenen Datenschutzniveaus. Auf der technischen Seite sind zentrale Komponenten und Verfahren vorzusehen, die für

- das Identitätsmanagement von Patienten und Probanden (Kap. 6.1), welches auch ein Kontaktmanagement einschließt,
- das Rechte- und Rollenmanagement für Netzteilnehmer (Kap. 6.2) und
- die Qualitätssicherung von Daten (Kap. 6.8)

zuständig sind und mit sorgfältig geplanten Sicherheitsmaßnahmen und -richtlinien, oft sogar bei organisatorisch unabhängigen Stellen (Trusted Third Parties, Datentreuhänder) betrieben werden.

Das Zusammenspiel verschiedener Module wird exemplarisch für den kombinierten Einsatz von

- Klinischem Modul und Studienmodul (Kap. 6.3) sowie
- Studien- und Forschungsmodul (Kap. 6.4)

erläutert; der kombinierte Einsatz von Forschungs- und Biobankenmodul war schon Gegenstand des Datenschutzkonzepts für Biomaterialbanken. Das Zu-

sammenspiel aller Module wird im Kapitel 6.5 als Maximalmodell beschrieben. Hinzu kommen Überlegungen und Vorschläge zu

- organisatorischen Regelungen (Kap. 6.6) und
- der Verhältnismäßigkeit von Maßnahmen (Kap. 6.7).

6.1 ID-Management

Das Identitätsmanagement für Patienten (und andere Studienteilnehmer) in einem medizinischen Forschungsverbund dient dazu,

- Daten, die zum selben Individuum gehören, korrekt zuzuordnen
- und dabei die Identität dieses Individuums vor Unberechtigten zu verbergen.

Diese beiden Ziele stehen in einem gewissen Spannungsverhältnis, da die korrekte Zuordnung eine Erkennbarkeit voraussetzt. Um diesen Zielkonflikt bestmöglich aufzulösen, werden sie generisch im Modul Identitätsmanagement zusammengefasst und auf zwei funktionale Komponenten aufgeteilt, deren Zusammenspiel sorgfältig austariert werden muss. Diese beiden Komponenten werden hier als

- Patientenliste und
- Pseudonymisierungsdienst

bezeichnet. Sie werden in Kapitel 6.1.1 in ihren Funktionen und in Kapitel 6.1.5 in ihrer organisatorischen Ausgestaltung beschrieben. Wie weit diese konzeptionelle Aufteilung des Identitätsmanagements in zwei Komponenten bei einer Implementierung abgebildet werden muss oder kann, ist Gegenstand späterer Erläuterungen.

Die persönliche Identifikation eines Patienten soll nur den unmittelbar an seiner Behandlung Beteiligten möglich sein, nicht aber anderen, z.B. wissenschaftlich tätigen Mitarbeitern des Forschungsnetzes. Außerhalb des direkten Behandlungskontexts ist daher – da die Ziele eines medizinischen Forschungsverbundes in der Regel mit anonymisierten Daten nicht erreicht werden können – ein pseudonymes Identitätsmanagement aufzubauen. Sinngemäß gilt das gleiche für Studienteilnehmer (Probanden), die nicht Patienten sind (z.B. Kontrollpersonen, Teilnehmer an epidemiologischen Studien).

6.1.1 Zweck und Verwendungsbereich

Patienten oder Studienteilnehmer werden in verschiedenen Bereichen des Forschungsverbundes durch unterschiedliche pseudonyme Kennzeichen repräsentiert[26]:

26 Die unterschiedlich aufgebauten Bezeichnungen und Akronyme für die verschiedenen Pseudonyme resultieren aus dem historischen „Wildwuchs“ und den in anderen Kontexten bereits etablierten Nomenklaturen.

- PID_K: im Klinischen Modul
- PID_S: im Studienmodul
- SIC: im Studienmodul für einzelne Studien
- PSN: im Forschungsmodul
- LabID: zur Kennzeichnung von Proben im Laborbereich (Probenbank)
- $LabID_{tr}$: zur Kennzeichnung von Proben im Forschungsmodul

Siehe hierzu auch Abbildung 8. Eventuell kommen dazu weitere Pseudonyme wie PID_B in einer Bilddatenbank, sofern eine solche unter getrennter Datenhoheit geführt wird, temporäre Pseudonyme für die Verarbeitung im Grid oder in der Cloud (vgl. Kap. 6.1.3.7) oder PSN_I für Datenexporte an Forschungsprojekte.

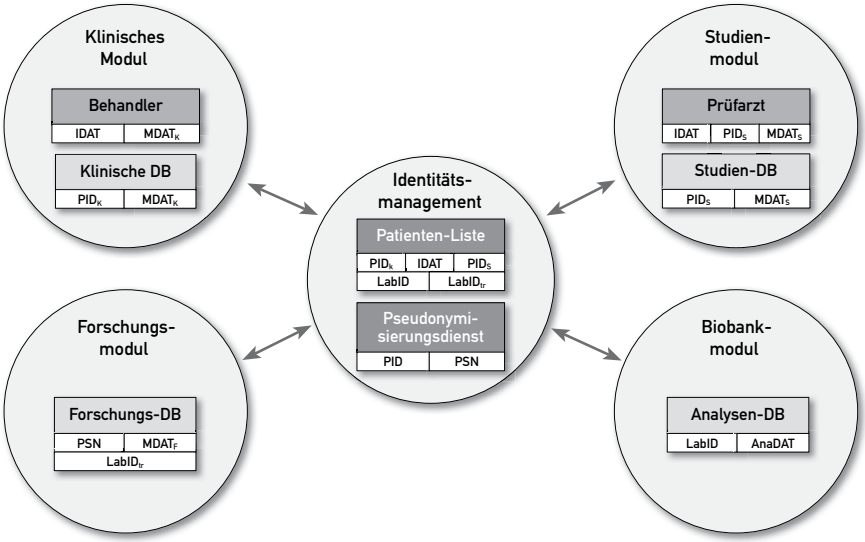

Abb. 8 Die zentrale Stellung des Identitätsmanagements; die Akronyme sind in Tabelle 2 zusammengestellt.

Grundsätzliche Aufgabe des Identitätsmanagements ist, die Zuordnung dieser Pseudonyme zueinander und zu den Identitätsdaten in den Anwendungsfällen, die dieses erfordern, herzustellen. Für solche Zuordnungs- oder Depseudonymisierungsprozesse sind nach den Regeln des Forschungsverbundes Entscheidungsprozesse und Kontrollen notwendig; das Identitätsmanagement soll organisatorisch und technisch so gestaltet werden, dass diese Regeln unterstützt bzw. erzwungen werden.

6.1.1.1 Patientenliste (mit PID-Dienst und IDAT-Datenbank)

Zweck der Patientenliste ist die Anmeldung und Registrierung eines Patienten oder Studienteilnehmers im Forschungsverbund sowie die Zuordnung eines eindeutigen nichtsprechenden Identifikators zu den Identitätsdaten IDAT.

Tab. 2 Akronyme

Abkürzung	Bedeutung	Verwendung
IDAT	Identitätsdaten	Direkter Behandlungszusammenhang
PID	(nichtsprechender) Patientenidentifikator	im Pseudonymisierungsdienst, im generischen Fall PID = PID_S, aber auch PID = PID_K möglich
PID_K	(nichtsprechender) Patientenidentifikator	Klinisches Modul
PID_S	(nichtsprechender) Patientenidentifikator	Studienmodul
SIC	(nichtsprechender) Subject Identification Code	einzelne Studie oder Studien-DB
PSN	Pseudonym	Forschungsmodul
LabID	Probenkennzeichnung	Biobank
$LabID_{tr}$	Verschlüsselte Probenkennzeichnung	Forschungsmodul
$MDAT_K$	Medizinische Daten	Klinisches Modul
$MDAT_S$	Medizinische Daten	Studienmodul
$MDAT_F$	Medizinische Daten	Forschungsmodul
AnaDAT	Analysedaten aus Proben, insbesondere genetische Daten	Biobank
OrgDAT	organisatorische Daten	siehe Kapitel 6.5.2.4

Dieser Identifikator wird hier zunächst als PID bezeichnet und kann als Pseudonym oder als Teil der IDAT behandelt werden. Es kann sich dabei um einen PID_K (aus dem Klinischen Modul) oder PID_S (aus dem Studienmodul) handeln, je nachdem, welche dieser beiden Kennungen in diesem Forschungsverbund benötigt wird oder aus welchem Bereich die Anmeldung erfolgt. Je nach Meldeweg wird diese Kennung

- an der Datenquelle erzeugt und an die Patientenliste übergeben
- oder erst dort erzeugt bzw. aus dem schon vorhandenen Bestand entnommen;

die andere der beiden (sowie weitere benötigte) Kennungen wird daraus in der Software der Patientenliste durch eine kryptographische Transformation abgeleitet. Werden im Verbund mehrere klinische Studien mit verschiedenen SICs durchgeführt, wird die Zuordnung zwischen diesen und dem PID_S ebenfalls in der Patientenliste zusammen mit dem Hinweis auf den Kontext der jeweiligen Kennung ($OrgDAT_{PL}$) aufbewahrt. Dieser Kontext enthält Angaben zur meldenden Stelle und das Meldedatum, um einen für den Patienten verantwortlichen Arzt als Ansprechpartner identifizieren (s.a. Kap. 6.5.2.4) und im Bedarfsfall einen Kontakt herstellen zu können.

Die eindeutige Identifikation des Patienten durch die Patientenliste wird als ein Mittel der Qualitätssicherung verstanden. Zugrunde liegt ein Szenario, in

dem ein Patient mit einer chronischen oder langwierigen Erkrankung über einen längeren Zeitraum von unterschiedlichen Einrichtungen behandelt oder beobachtet wird. Ein Patient kann also von verschiedenen Stellen zu unterschiedlichen Zeitpunkten zur Teilnahme am Forschungsverbund angemeldet oder seine dort bereits vorhandenen Daten ergänzt werden. Durch die Arbeitsweise der Patientenliste soll sichergestellt werden, dass ein einmal angemeldeter Patient bei einer späteren Meldung wieder erkannt wird. Die Identifikation eines Patienten geschieht über die Stammdaten (IDAT), welche den Patienten im Klartext identifizieren und in der Patientenliste gespeichert werden. Über die IDAT ist im Bestand dieser Liste zu prüfen, ob der Patient bereits erfasst und ein PID vergeben ist. Im negativen Fall ist ein neuer PID zu erzeugen und mit den IDAT in den Bestand der Patientenliste zu übernehmen. Eine mögliche technische Komponente zur Umsetzung einer solchen Patientenliste ist der PID-Generator der TMF.

Ein wichtiges Problem der Identifikation besteht darin, sicherzustellen, dass bei der Vergabe des PID Synonymfehler (ein Patient hat mehrere PIDs) und Homonymfehler (zwei oder mehr Patienten haben einen identischen PID) mit möglichst hoher Sicherheit vermieden werden, und zwar auch dann, wenn die IDAT durch Änderung (z.B. des Namens) oder durch unterschiedliche Schreibweise oder Eingabefehler voneinander abweichen. Dazu dienen folgende Maßnahmen:

Die Erhebung der IDAT wird möglichst einheitlich gestaltet. Als Basis der IDAT wird der Datensatz der Versichertenkarte (VK oder eGK) empfohlen, da hiermit das größtmögliche Maß an Normierung erreicht wird und durch elektronische Übernahme der Daten fehlerhafte Eingaben vermieden werden können. Nach dem Rechtsgutachten [11] sind nicht direkt versorgungsbezogene Daten, wie z.B. Angaben zur Versicherung und insbesondere der lebenslang konstante Teil der Versichertennummer, hierfür allerdings nicht nutzbar (s. Kap. 4.3.2).

Zusätzlich soll der Geburtsname oder ein anderer früherer Name erfasst werden, wenn ein Patient während seiner Verweilzeit im Forschungsnetz den Namen gewechselt hat.

Für den Bestandsabgleich wird ein fehlertoleranter Algorithmus mit einstellbarer Empfindlichkeit verwendet.

Unklare Zuordnungen können durch manuellen Eingriff eines Administrators und ein Rückfragemanagement aufgelöst werden.

Mit der Anmeldung eines Patienten oder Studienteilnehmers bei der Patientenliste werden ein Kennzeichen der meldenden Stelle und das Datum der Meldung übertragen und in der Liste gespeichert. Dies gilt auch dann, wenn einem Patienten bereits ein PID zugewiesen wurde und dieser einer neu meldenden Klinik übermittelt wird. Kennzeichen und Datum werden nicht als Historie geführt, sondern durch die jeweils aktuelle Meldung überschrieben.

Die Daten ($OrgDAT_{PL}$ mit Kontextinformation) werden benötigt, damit die Stelle, welche die Patientenliste führt, erkennen kann, über welche Klinik oder welchen verantwortlichen Arzt in einem entsprechenden Anwendungsfall ein Patient kontaktiert werden kann. Sie können auch als Entscheidungshilfe bei der Prüfung von Zugriffsberechtigungen herangezogen werden (s.a. Kap. 6.5.2.4).

Die Funktion der Patientenliste ist weitgehend automatisiert. Bei der Anmeldung eines Patienten können Fälle auftreten, in denen die Zuordnung der Meldung zum Bestand der Liste zwar möglich, aber wenig gesichert ist. Ein solcher Fall führt zu einer Fehlermeldung. Abhängig davon, wie wichtig es für die Forschungsziele ist, Synonyme und Homonyme zu vermeiden, kann für solche Fälle ein Verfahren zum manuellen Abgleich von Daten vereinbart werden; bei der Führung der Patientenliste ist dann der Eingriff durch einen Operator bzw. eine Dokumentationsfachkraft erforderlich.

Hinweis: Das Identitätsmanagement bei einzelnen Projekten, insbesondere klinischen Studien nach AMG, ist evtl. unabhängig zu betreiben, da eine Systemvalidierung bei zentral genutzten Diensten wesentlich erschwert ist; dies wird relevant, sobald das Identitätsmanagement mit Fehlerkorrektur- und Record-Linkage-Mechanismen versehen ist. Daher ist es hier in der Regel zu empfehlen, dass der Prüfarzt einen SIC vergibt oder einmalig aus einer externen Quelle übernimmt, der nur ihm – und darüber hinaus bei Notwendigkeit der Patientenliste im Verbund – bekannt ist. Ein solcher Mechanismus zur SIC-Erzeugung ist in der Regel in der Studiensoftware implementiert.

6.1.1.2 Pseudonymisierungsdienst

Zweck des Pseudonymisierungsdienstes ist der besondere Schutz der Daten in dem auf Langzeitspeicherung angelegten Forschungsmodul. Mittel dazu ist die Transformation des PID aus der Patientenliste in ein Pseudonym PSN, das in der Forschungsdatenbank als Kennung genutzt wird; hierfür wird ein kryptographisches Verfahren angewendet. Der Datenfluss in die Forschungsdatenbank wird über den Pseudonymisierungsdienst geleitet, wobei der PID durch das PSN ersetzt wird. Da der Pseudonymisierungsdienst die medizinischen Daten (MDAT) weder benötigt noch überhaupt sehen soll, werden diese in asymmetrisch verschlüsselter Form durchgereicht oder ganz an ihm vorbei geleitet, siehe Abbildung 9. Die zweite Option hat den Vorteil geringerer Anforderungen an die Bandbreite des Datendurchsatzes im Pseudonymisierungsdienst und den Nachteil erhöhter Komplexität der Kommunikation.

Die Pseudonymisierung ist eine reine Maschinenfunktion, die keines Eingriffs durch das Personal bedarf. Um eine unberechtigte Nutzung dieses Dienstes, der mit der Weiterleitung der Daten an die Forschungsdatenbank verbunden ist, auszuschließen, werden die Daten nur von zugelassenen Absendern übernommen (s. Kap. 6.1.4.2).

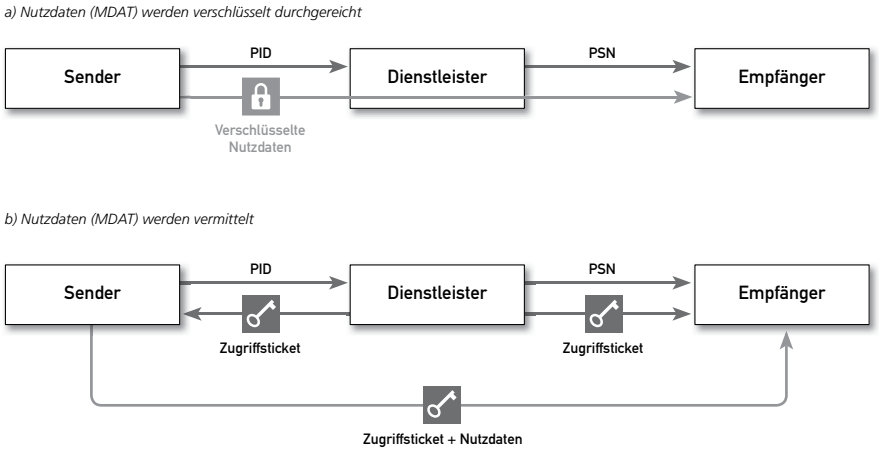

Abb. 9 Verschlüsselte Durchleitung oder Vorbeileitung von Daten; Aufgabe des Dienstleisters ist nur, pseudonyme Kennungen (z.B. PID und PSN oder LabID und $LabID_{tr}$) ineinander umzuwandeln.

Hinweis: Die Pseudonymisierung wird hier für die Patientendaten beschrieben. Sie kann in gleicher Weise eingesetzt werden, um die Kennung medizinischer Einrichtungen oder individueller Ärzte (ADAT) in den Forschungsdaten unkenntlich zu machen. Selbstverständlich ist dies auch in einem nachfolgenden Schritt beim Export von Daten aus der Forschungsdatenbank möglich. Die Lösung muss in der Vertragsgestaltung zwischen den Ärzten und dem Forschungsverbund und bei der Regelung des Zugangs zu Forschungsdaten definiert werden. Ein entsprechender Depseudonymisierungsvorgang muss eingerichtet werden.

6.1.2 Anwendungsfälle

a) **Anmeldung** eines Patienten oder Studienteilnehmers beim Forschungsverbund: Die direkte Anmeldung erfolgt bei der Patientenliste. Das Identitätsmanagement sorgt dafür, dass die nötigen pseudonymen Kennzeichen für die verschiedenen Bereiche des Forschungsverbundes erzeugt werden, siehe auch Abbildung 10.

b, c) **Übermittlung** von Daten an die Forschungsdatenbank aus dem Versorgungskontext (Fall b) oder aus dem Studienkontext (Fall c): Die für die Forschungsdatenbank asymmetrisch verschlüsselten Daten (MDAT) werden über das Identitätsmanagement geleitet, das die verwendete Kennzeichnung in das im Forschungskontext verwendete PSN umwandelt. Als Alternative können die MDAT auch mit Hilfe eines vom Pseudonymisierungsdienst vergebenen Zugriffstickets direkt an die Datenbank übergeben werden, siehe Abbildung 9.

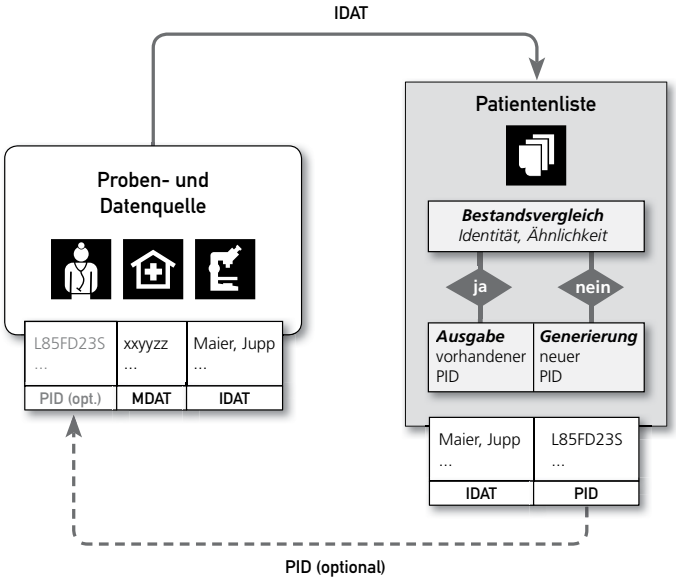

Abb. 10 Anmeldung eines Patienten oder Studienteilnehmers in der Patientenliste. Der pseudonyme Patientenidentifikator PID wird nur im Studienmodul (als PID_S), nicht aber im Klinischen Modul (als PID_K) zurückgemeldet.

d, e) **Mitteilung** von Ergebnissen (Findings) aus einem Forschungsprojekt (Fall d) an einen Patienten oder Studienteilnehmer: Hier wird über das Identitätsmanagement das Pseudonym in eine Kennung (in der Regel IDAT) umgewandelt, die dem für den Betroffenen verantwortlichen Arzt bekannt ist, und diesem das jeweilige Finding mitgeteilt; auch hier ist die asymmetrisch verschlüsselte Übertragung zu nutzen, siehe Abbildung 22 in Kapitel 6.8.3.2. Auf gleiche Weise kann auch auf **Anfragen** eines Patienten (Fall e) reagiert werden, sofern diese Art des Auskunftsrechts mit ihm vereinbart wurde. Hier wendet sich der Patient an seinen zuständigen Arzt oder einen sonstigen Auskunftsverpflichteten; dieser übermittelt die IDAT des Patienten über das Identitätsmanagement an die jeweilige Datenbank, die die Rückmeldung wie beschrieben veranlasst.

f) Die Depseudonymisierung zur **Datenqualitätssicherung** wird im Kapitel 6.8 über Qualitätssicherung behandelt.

g) Für die **Rekrutierung** von Studienteilnehmern für eine neue Studie – sofern dies aufgrund der Einwilligungserklärung erlaubt ist – wird der gleiche Weg zum verantwortlichen Arzt wie bei der Rückmeldung von Findings beschritten. Von diesem wird die Einwilligung des Betroffenen zur Teilnahme an dieser Studie eingeholt sowie dieser ggf. als Teilnehmer dieser Studie angemeldet.

h, i) Bei einem **Widerruf** der Teilnahme am Forschungsverbund werden, abhängig von der Vereinbarung in der Einwilligungserklärung, über das Identi-

tätsmanagement die entsprechenden Daten in den Datenbanken gefunden und **gelöscht** (Fall h) oder der Fall wird im Forschungsverbund **anonymisiert** (Fall i); das bedeutet hier einfach, dass in der Patientenliste – falls vorhanden, auch in dezentralen Patientenlisten – die IDAT gelöscht werden und somit der Bezug zum Individuum nicht mehr hergestellt werden kann. Im Einzelfall kann es hierbei auch nötig sein, charakteristische Merkmale der MDAT zu vergröbern oder zu löschen. Die Pseudonyme können, wenn mit Sicherheit niemand mehr darüber einen Personenbezug herstellen kann, als anonyme Kennzeichen in den jeweiligen Datenbanken verbleiben; ansonsten sind sie durch eindeutige anonyme Kennzeichen zu ersetzen.

j) Im **Todesfall** eines Patienten oder Probanden sind in der Regel, sobald eine Nachmeldung oder Nacherfassung von Daten nicht mehr zu erwarten ist, alle seine Daten im Forschungsverbund zu anonymisieren; die Regularien des Forschungsverbundes und die Einwilligungserklärung sind zu beachten. Da im Maximalmodell die Dauerspeicherung nur im Forschungsmodul vorgesehen ist, sind noch im Klinischen Modul oder Studienmodul befindliche Daten dorthin zu überführen und überall sonst zu löschen. Proben und Daten im Biobankenmodul können – sofern das vorgesehen und rechtlich abgesichert ist – ebenfalls erhalten bleiben, und ebenso muss die Assoziation zwischen Daten im Forschungsmodul und Biobankenmodul über PSN und LabID bzw. $LabID_{tr}$ erhalten bleiben. In der Patientenliste sind die IDAT und alle nicht mehr benötigten pseudonymen Kennungen zu löschen.

k) Eine **Umpseudonymisierung** (Ersetzen vorhandener Pseudonyme durch neue) kann nötig werden, wenn einzelne Pseudonyme als kompromittiert erkannt werden oder wenn das Pseudonymisierungsverfahren insgesamt als unzureichend oder nicht mehr dem Stand der Technik (s. Kap. 2.6 des Kryptographischen Gutachtens im Anhang[27]) entsprechend eingeschätzt wird (vgl. zugehörigen Anwendungsfall in Kap. 6.4.2.10). Beim Verfahren ist zu unterscheiden, ob die Pseudonyme durch eine Zuordnungsliste oder eine kryptographische Transformation erzeugt wurden. Im Fall einer Zuordnungsliste, die willkürliche Pseudonyme ohne Verwendung eines deterministischen Algorithmus vergibt, ist nur der Fall der Kompromittierung einiger oder aller Pseudonyme relevant. Diese müssen dann durch neu vergebene Pseudonyme ersetzt werden, die auch an die jeweiligen Datenbanken des Forschungsverbundes weitergegeben werden. Die Möglichkeit, dass durch die Kompromittierung bereits Daten an Unbefugte gelangt sind, erfordert Reaktionen auf der organisatorischen Ebene des Forschungsverbundes, die aber das Identitätsmanagement nicht weiter involvieren.

Falls die Pseudonyme durch eine kryptographische Transformation vergeben wurden, ist der bisherige Algorithmus durch einen neuen ausreichender Stär-

27 Anhänge zu diesem Dokument sind unter www.tmf-ev.de/datenschutz-leitfaden verfügbar.

ke zu ersetzen; alle bisher vergebenen Pseudonyme müssen durch die nach dem neuen Algorithmus erzeugten ersetzt werden, vgl. Kapitel 2.6 des Kryptographischen Gutachtens im Anhang.

l) Der **Export medizinischer Daten für die Weitergabe an Forscher** ist ein Anwendungsfall, der alle Datenbanken eines Forschungsverbunds betreffen kann. Dabei sind nach Möglichkeit anonymisierte Daten herauszugeben, so dass die Erzeugung anonymer Datensatz-IDs als Funktion des ID-Managements genutzt werden kann. Wenn die Nutzung der Daten mögliche Implikationen für die betroffenen Probanden hat und die Auswertung pseudonymer Daten durch die Wissenschaftler z.B. durch eine Einwilligung rechtlich abgesichert ist, müssen die vorhandenen pseudonymen Kennzeichen der jeweiligen Datenbank durch für diesen Export spezifische neue pseudonyme Kennzeichen ersetzt werden. Um die Auswertungsergebnisse ggf. später wieder einem Probanden zuordnen zu können, muss der für diesen Export eingesetzte Schlüssel für die Umpseudonymisierung gespeichert werden. Auch diese exportspezifische Pseudonymisieurng kann als Funktion des ID-Managements realisiert werden.

m) Das **Aktualisieren der Kontaktdaten** von Patienten oder Probanden kann logisch einem zentralen ID-Management zugeordnet werden (vgl. Kap. 3.2.3.2). Wichtig ist dies, wenn der Forschungsverbund langfristig und ggf. auch studienübergreifend den Kontakt zu den Patienten oder Probanden halten möchte, z.B. um diese für neue Projekte zu rekrutieren oder über neue Ergebnisse zu informieren. Die mit dieser Aufgabe betrauten Mitarbeiter sollten keinen Zugriff auf medizinische Daten haben und auch die für die Pflege der Kontaktdaten nicht nötigen Pseudonyme nicht einsehen können. Wichtig ist das Vorliegen notwendiger Informationen aus den Einwilligungserklärungen, da diese im Regelfall die Rechtsgrundlage für ein direktes Ansprechen der Patienten oder Probanden aus dem Forschungskontext heraus darstellen. Ggf. kann für diesen Aufgabenbereich auch eine spezialisierte CRM-Software zum Einsatz kommen.

6.1.3 Daten und Datenflüsse

6.1.3.1 Daten der Patientenliste

Die Patientenliste speichert und verwaltet IDAT, PID_K, PID_S und zugehörige SICs sowie andere gegebenenfalls in anderen Modulen des Forschungsverbunds benötigte Kennungen wie LabID oder das Pseudonym einer Bilddatenbank (PID_B), außerdem die Kontextdaten $OrgDAT_{PL}$ (einschließlich ADAT). Sie sieht, kennt und speichert nicht MDAT und PSN. Die Verwaltung der pseudonymisierten $LabID_{tr}$ ist logischer Teil des Identitätsmanagements. Sie kann, wie in Abbildung 8 dargestellt, bei der Patientenliste angesiedelt sein. Als Option besteht auch die im generischen Datenschutzkonzept für Biomaterialbanken [2] vorgesehene Möglichkeit, die Zuordnung zwischen LabID und $LabID_{tr}$ der Probenbank als Aufgabe zu übergeben.

Die Patientenliste *erhält* die Daten IDAT und $OrgDAT_{PL}$ (s. Kap. 6.1.3.3 unten), je nach Szenario empfängt sie auch dezentral erzeugte Identifikatoren, z.B. SICs. Sie *gibt* den PID_K an die Klinische Datenbank *zurück*, den PID_S an die Studiendatenbank und an die Datenquelle (hier: Prüfarzt), je nach Szenario auch den PID_K an die Datenquelle (hier: behandelnder Arzt). Je nach Szenario gibt die Patientenliste auch Einmal-Kennungen (als Zugriffstickets) an einen behandelnden Arzt und die Klinische Datenbank, die temporär zur richtigen Zuordnung von Kommunikationsprozessen benötigt werden.

Es wird empfohlen, in der Patientenliste als PID primär den PID_S zu erzeugen, und zwar in menschenlesbarer Form (8 Buchstaben und Ziffern). Der PID_K – ebenso wie weitere benötigte Kennzeichen – wird daraus durch kryptographische Verschlüsselung gewonnen und ist eine nur maschinenlesbare Bitkette; dies ist angemessen, da der PID_K nur in der Kommunikation von Patientenliste mit Klinischer Datenbank genutzt wird und sonst nirgends sichtbar sein soll.

6.1.3.2 Daten des Pseudonymisierungsdienstes

Der Pseudonymisierungsdienst speichert keine Daten außer dem geheimen kryptographischen Schlüssel, der die Zuordnung zwischen PID_S und PSN vermittelt. Er *erhält* einen PID (im generischen Fall den PID_S) und *gibt* das zugehörige PSN *weiter*. Bei einer Depseudonymisierung ist dies genau umgekehrt (s. Kap. 6.1.3.5 unten).

Der Schlüssel für die Transformation PID ↔ PSN ist unauslesbar auf einer Smartcard oder in einer vergleichbar sicheren Umgebung wie z.B. einem Hardware Security Module (HSM) zu speichern, damit er sicher als Geheimnis bewahrt werden kann. Die kryptographischen Funktionen müssen ebenfalls in der sicheren Umgebung, z.B. auf der Smartcard, ausgeführt werden, damit der Schlüssel diese nicht verlassen muss.

6.1.3.3 Datenflüsse der Patientenliste

Die Anmeldung eines Patienten oder Studienteilnehmers beim Forschungsverbund erfolgt bei der Patientenliste. Mit dem dort erzeugten oder schon vorhandenen PID können dann medizinische Daten (MDAT) an die entsprechende Datenbank zusammen mit dem entsprechenden pseudonymen Kennzeichen übermittelt werden. Wird der PID nicht an die Datenquelle zurückgemeldet, wie es in einigen Szenarien sinnvoll ist, wird für die Datenübermittlung statt dessen ein von der Patientenliste erzeugtes Zugriffsticket (zum einmaligen Gebrauch) verwendet.

Auch bei der Depseudonymisierung wirkt die Patientenliste mit (s. Kap. 6.1.3.5 unten).

6.1.3.4 Datenflüsse des Pseudonymisierungsdienstes

Der Pseudonymisierungsdienst erhält einen PID_S von der Patientenliste und gibt das zugehörige PSN an die Forschungsdatenbank weiter (s. Abb. 11). Dazu werden die MDAT (aus dem Klinischen Modul oder Studienmodul) an die Forschungsdatenbank (FDB) nach einer von zwei Methoden übertragen (s. Abb. 9):

- asymmetrisch verschlüsselt über den Pseudonymisierungsdienst weitergeleitet oder
- mit Hilfe eines temporären Zugriffstickets, das die richtige Zuordnung garantiert, direkt von der Datenquelle.

Beiden Optionen ist gemeinsam, dass der Pseudonymisierungsdienst keine Möglichkeit hat, die MDAT zu lesen.

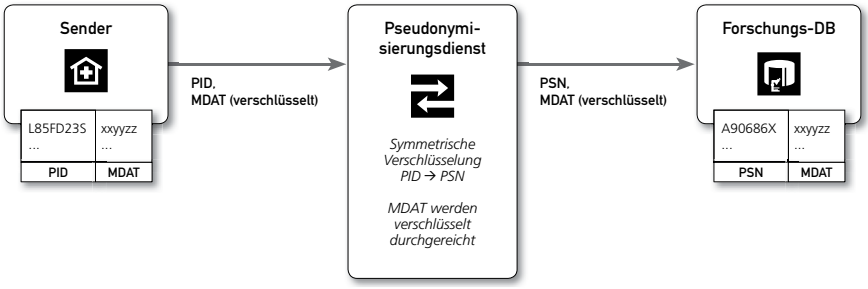

Abb. 11 Workflow des Pseudonymisierungsdienstes; alternativ ist auch eine getrennte Übermittlung der MDAT mit Hilfe eines Zugriffstickets möglich (s. Abb. 9).

In der Forschungsdatenbank werden die MDAT entschlüsselt und mit dem Pseudonym PSN abgespeichert. Der Vorgang ist aus der Sicht des Pseudonymisierungsdienstes unabhängig davon, ob die Daten neu geliefert werden oder ob eine Änderungsmeldung bereits in die Datenbank übernommene Daten korrigiert oder ergänzt. Nur der Betreiber der Datenbank muss und kann die beiden Formen unterscheiden.

Der Rückbezug von Daten aus der Forschungsdatenbank oder daraus abgeleiteten Auswertungen auf den betroffenen Patienten kann daher ausschließlich über den Weg der Depseudonymisierung, d.h. der kryptographischen Rücktransformation des PSN in den PID gewonnen werden (s. Kap. 6.1.3.5 unten).

Hinweis: Ein weiterer Pseudonymisierungsschritt wird für den Export der Daten empfohlen, wenn verhindert werden soll, dass außerhalb der zentralen Datenbank Akkumulationen von Daten erfolgen. Dabei wird das PSN jeweils durch eine weitere kryptographische Transformation oder eine willkürliche Zuordnungsliste in ein projektspezifisches PSN_I umgewandelt, das als Ordnungskriterium für Datenbestände gilt, die an das Forschungsprojekt Nr. *i* exportiert werden. Diese Transformation kann durch einen zentralen Pseudonymisie-

rungsdienst unterstützt werden, der sich auch die projektspezifische Zuordnungsliste oder den verwendeten Schlüssel für mögliche Rückmeldungen merkt. Alternativ kann dies auch im Rahmen der Exportfunktion einer Forschungsdatenbank realisiert werden.

6.1.3.5 Depseudonymisierung

Die Depseudonymisierung kann nur von einer berechtigten Einrichtung bzw. von berechtigten Personen nach dem Regelwerk des Forschungsverbundes veranlasst und nur vom Identitätsmanagement durchgeführt werden.

Technisch ist die Depseudonymisierung im generischen Fall zweistufig angelegt: Die erste Stufe wird auf dem inversen Weg der Pseudonymisierung durch die Transformation eines Pseudonyms PSN in einen Patientenidentifikator PID geleistet. Dazu erhält der Pseudonymisierungsdienst ein PSN, leitet daraus den zugehörigen PID ab und gibt diesen (zusammen mit organisatorischen Daten des Vorgangs) an die Patientenliste weiter; dieser Schritt kann in definierten Anwendungsfällen auch automatisiert ablaufen. In der zweiten Stufe wird der PID in der Patientenliste aufgrund einer Datenbank-Abfrage durch die Identifikationsdaten IDAT ersetzt. Diese werden zusammen mit den organisatorischen Daten des Vorgangs an den in den $OrgDAT_{PL}$ (bzw. ADAT) genannten Verantwortlichen weitergeleitet (s. Abb. 12).

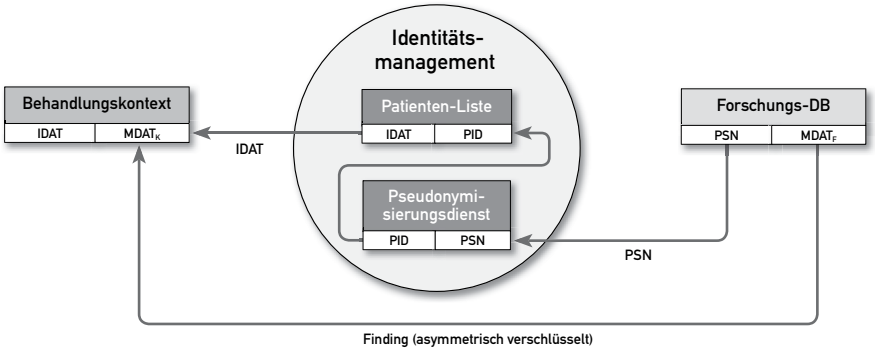

Abb. 12 Workflow der Depseudonymisierung im Anwendungsfall Rückmeldung (PID = PID_S oder PID_K je nach Kontext)

6.1.3.6 Umpseudonymisierung

Die Umpseudonymisierung im Falle einer Zuordnungsliste betrifft die Patientenliste mit dem primär erzeugten PID (im Regelfall PID_S) und erfordert folgenden Ablauf: Jeder zu ändernde PID wird vom Administrator der Patientenliste durch einen entsprechenden neuen ersetzt. Alle daraus erzeugten weiteren Kennungen in Patientenliste und Pseudonymisierungsdienst sind entsprechend

zu ändern und – zusammen mit der alten, zu ändernden Kennung – an die jeweiligen nutzenden Datenbanken zu übermitteln.

Im Falle einer kryptographischen Transformation ist zu unterscheiden, ob nur für einen Einzelfall eine Umpseudonymisierung nötig ist, oder ob wegen Kompromittierung des kryptographischen Verfahrens sämtliche Pseudonyme ausgetauscht werden müssen. Im ersten Fall wird ein neu erzeugter PID zusammen mit dem alten angeliefert und in ein neues PSN umgewandelt, das zusammen mit dem alten an alle relevanten Stellen im Forschungsmodul übermittelt und mit einer Änderungsaufforderung versehen wird.

6.1.3.7 Temporäre Pseudonyme in verteilten Infrastrukturen

Um umfangreiche medizinische Datensätze, wie sie z.B. die Bildgebung oder genetische Sequenzierungsmethoden produzieren, in vertretbarer Zeit und mit ökonomisch vertretbaren Mitteln verarbeiten und auswerten zu können, werden zunehmend verteilte Infrastrukturen, entweder als Grid oder Cloud, eingesetzt. Wenn diese Infrastrukturen nur für die Analyse der Daten und nicht für eine dauerhafte Speicherung eingesetzt werden, was insbesondere bei rechenintensiven Verarbeitungsschritten der Regelfall ist, kann der Schutzbedarf der Daten durch die Verwendung temporärer Pseudonyme noch weiter abgesenkt werden.

Hierfür werden zentral vom ID-Management netzweit eindeutige Pseudonyme bereit gestellt, die für einen Transfer eines Datensatzes in das Grid oder in die Cloud, die dortige Verarbeitung und die Rückübermittlung des Ergebnisses gültig sind. Diese werden ohne Übermittlung identifizierender Daten abgerufen und direkt nach Erhalt der Ergebnisse im ID-Management wieder freigegeben. Lediglich die Daten liefernde Stelle speichert während der Verarbeitung der Daten den Zusammenhang von temporärem Pseudonym und der Identität des zugehörigen Probanden oder Patienten. Im ID-Management wird das herausgegebene Pseudonym bis zur Freigabe durch die abrufende Stelle gesperrt und damit in dieser Zeit nicht erneut herausgegeben.

Bei komplexen Datenstrukturen, wie sie z.B. auch im DICOM-Header von Bildern und Bildserien vorkommen, müssen ggf. multiple Identifikatoren, wie beispielsweise global eindeutige IDs für jedes Bild, das bildgebende Gerät usw. ersetzt werden. In solchen Fällen müssen entweder mehrere temporäre Pseudonyme verwendet werden oder von einem temporären Pseudonym werden weitere abgeleitet, z.B. durch Suffixe oder Präfixe. Dabei ist aber auch zu berücksichtigen, dass für standardisierte Datenformate wie DICOM bestimmte Vorgaben bezüglich der verwendeten IDs eingehalten werden müssen.

Eine weitere Absenkung des Schutzbedarfs kann durch den zusätzlichen Einsatz eines Pseudonymisierungsdienstes erreicht werden, der die temporären Pseudonyme beim Transfer der Daten in das Grid oder in die Cloud durch symmetrisch verschlüsselte temporäre Pseudonyme zweiter Ordnung ersetzt

(s. Kap. 6.1.3.4). Bei der Rückübermittlung der Ergebnisse wird die Umschlüsselung im Pseudonymisierungsdienst wieder rückgängig gemacht, so dass an der Datenquelle die Ergebnisse wieder dem richtigen Patienten oder Probanden zugeordnet werden können. Bei komplexen Datenstrukturen mit multiplen, durch temporäre Pseudonyme ersetzten Identifikatoren müssen alle diese Kennungen beim Verschlüsseln berücksichtigt werden. Da in solchen Anwendungsfällen im Regelfall von größeren Datenmengen auszugehen ist, wird als Implementierungsvariante des Pseudonymisierungsdienstes diejenige empfohlen, bei der die Nutzdaten nicht asymmetrisch verschlüsselt durchgereicht, sondern vollständig an dem Pseudonymisierungsdienst vorbei geleitet und über Tickets korrekt zugeordnet werden (vgl. Kap. 6.1.1.2 und Abb. 9).

Zusätzlich zu den hier beschriebenen speziellen Verfahren im ID-Management sind in solchen Einsatzszenarien auch ergänzende organisatorische Regelungen zu treffen, die u.a. auch ein ausreichendes Schutzniveau im Grid oder in der Cloud garantieren. Weitere Hinweise dazu finden sich in Kapitel 6.6 und in den Ergebnisdokumenten der Projekte PneumoGrid und cloud4health[28]. Die TMF war bzw. ist in beiden Projekten an der Ausarbeitung datenschutzkonformer Umsetzungskonzepte beteiligt.

6.1.3.8 Todesfall

Abhängig davon, wo der Todesfall bekannt wird – in der Regel zuerst im Klinischen Modul, unter Umständen aber auch im Forschungsmodul, wenn dort Nacherhebungen oder Abgleiche mit Melderegistern oder epidemiologischen Registern vorgesehen sind, wird eine Meldung an das Identitätsmanagement gemacht, das die in Kapitel 6.1.2 j) vorgesehenen Löschungen veranlasst und entsprechende Rückmeldungen empfängt.

6.1.4 Nutzer, Rollen und Rechte

6.1.4.1 Patientenliste

Für die Patientenliste gibt es im Allgemeinen Nutzer aus meldenden Einrichtungen, die entweder über ein Web-Formular oder aus einem EDC-System heraus einen Patienten oder Studienteilnehmer melden können; zwischen Neumeldung und Nachmeldung mit geänderten IDAT wird dabei nicht unterschieden. Wird die Patientenliste nur im Batchbetrieb genutzt, gibt es diese externen Nutzer nicht; sie werden durch Sender bzw. Empfänger entsprechender Dateien ersetzt.

Die Patientenliste hat einen Systemadministrator. Dieser hat neben der rein technischen Server-Administration die Aufgaben

28 s. www.pneumogrid.de und www.cloud4health.de

- manuelle Korrektur von Datensätzen bei (z.B. telefonisch) gemeldeten Fehleingaben,
- manuelle Korrektur von Datensätzen bei Zweifelsfällen, in denen die Zuordnung nicht automatisch entschieden werden kann,
- gegebenenfalls Bedienung des Batchbetriebs

zu erfüllen und muss demnach die entsprechenden Rechte zugeteilt bekommen.

Der Eingriff eines Administrators ist auch dann erforderlich, wenn im Rahmen der Depseudonymisierung einem PID die IDAT zugeordnet werden sollen: Hier ist zuerst die Genehmigung zu prüfen; bei positivem Ergebnis muss der Zuordnungsvorgang manuell gestartet werden.

Der Administrator kann bei seinen Aufgaben von einer Dokumentationsfachkraft unterstützt werden; diese benötigt lediglich die Rechte zur manuellen Korrektur von Datensätzen.

6.1.4.2 Pseudonymisierungsdienst

Zur Nutzung des Pseudonymisierungsdienstes siehe Kapitel 6.1.1.2. Er wird bei beabsichtigter Übertragung von Daten an die Forschungsdatenbank durch die entsprechenden Kommunikationskomponenten einer Klinischen oder Studiendatenbank angestoßen, siehe Kapitel 6.1.6.2. Diese müssen die entsprechenden Rechte besitzen, siehe Kapitel 6.2. Direkte Nutzer gibt es für den Pseudonymisierungsdienst nicht; er kann nur als Netzdienst über die definierten Schnittstellen angesprochen werden.

Der Pseudonymisierungsdienst hat einen Systemadministrator mit den üblichen Aufgaben und Rechten. Dieser ist auch für das Anstoßen von Depseudonymisierungsvorgängen nach persönlicher Prüfung der Berechtigung des Vorgangs zuständig, sofern im Regelwerk des Forschungsverbunds für den konkreten Fall nicht ein automatisierter Prozess vorgesehen ist, ebenso für das Anstoßen von Umpseudonymisierungsvorgängen. Dem Systemadministrator obliegt auch die sachgemäße technische Handhabung der Smartcard bzw. des Hardware Security Modules, das den Pseudonymisierungsschlüssel enthält.

6.1.5 Verantwortlichkeiten

Die Gesamtverantwortung für das Identitätsmanagement liegt bei der Leitung des Forschungsverbundes und dem Ausschuss Datenschutz einschließlich dem Datenschutzbeauftragten. Dieser Personenkreis gibt insbesondere Richtlinien und Policies vor. Im Folgenden wird die organisatorische und technische Verantwortung für die Komponenten des Identitätsmanagements im Rahmen dieser Richtlinien beschrieben.

Im generischen Fall wird empfohlen, sowohl die Patientenliste als auch den Pseudonymisierungsdienst zentral für einen Forschungsverbund einzurichten, da so ein hoher Sicherheitsstandard erreicht werden kann und die erforderliche Infrastruktur und das zugehörige Personal nur einmal für den gesamten Forschungsverbund eingerichtet werden muss. Beide Dienste sollten bei unabhängigen vertrauenswürdigen Stellen (Trusted Third Parties, TTP) angesiedelt sein. Wenn diese Dienste separat an unabhängigen Stellen betrieben werden, ist eine zusätzliche wirksame Trennung zwischen den patientennahen Bereichen der Versorgung und der klinischen Forschung und dem patientenfernen Bereich der Forschungsdatenbank gegeben. Varianten dieser generischen Empfehlung werden in den folgenden Kapiteln diskutiert.

6.1.5.1 Patientenliste zentral oder dezentral?

Die Patientenliste kann an drei Stellen angesiedelt sein:

- zentral bei einer eigenen TTP,
- zentral zusammen mit dem Pseudonymisierungsdienst,
- dezentral an den Datenquellen.

Mit der zentralen Einrichtung der Patientenliste wird angestrebt, dass die Kranken- und Behandlungsgeschichten von Patienten mit einer chronischen oder rezidivierenden Erkrankung möglichst langfristig verfolgt werden können. Der Wechsel von Behandlungseinrichtungen und die räumliche Mobilität der Patienten führen dazu, dass Patienten im Lauf der Zeit von verschiedenen Einrichtungen an den Forschungsverbund gemeldet werden. Dann soll auch bei modifizierter Eingabe der IDAT (z.B. durch Schreibfehler bei manueller Erfassung oder Namensänderung) sichergestellt werden, dass der Patient oder Studienteilnehmer im Bestand identifiziert und ihm der bereits vorhandene PID zugewiesen wird.

Es ist zwar möglich, auch dezentrale Patientenlisten so anzulegen, dass mit einem für alle identischen Algorithmus aus identischen Eingabedaten ein identischer PID erzeugt wird, jedoch ist nicht zu vermeiden, dass modifizierte Eingabedaten zu einem neuen PID führen, so dass Synonyme entstehen. Die dezentrale Führung der Patientenliste hat außerdem den Effekt, dass auch die Depseudonymisierung nur dezentral, über die Stellen, die den Patienten persönlich kennen, möglich ist. Eine dezentrale Anordnung ist deshalb nur sinnvoll, wenn die Datenerfassung einmalig ist, wenn mit einem Wechsel des Patienten nicht gerechnet werden muss oder wenn eine Doppelerfassung unerheblich ist. Andererseits fördert die dezentrale Führung von Patientenlisten an den Datenquellen – d.h., bei der Erfassung der Identitätsdaten wird von einer behandelnden Einrichtung auch gleich ein nichtsprechender PID vergeben – in den genannten Fällen die Datensparsamkeit und ist somit vom Datenschutzgesichtspunkt aus unkritisch, zumal dieser Bereich von der ärztlichen Schweigepflicht abgedeckt und damit als vertrauenswürdig anzusehen

ist. In diesem Fall entfällt die Speicherung der ADAT in der Patientenliste, stattdessen kann die Speicherung der ADAT als Teil der MDAT angebracht sein; je nach Reidentifizierungsrisiko und Verhältnismäßigkeit ist auch von der elektronischen Speicherung der ADAT ganz abzusehen und nur eine Papierliste an geeigneter Stelle aufzubewahren.

6.1.5.2 Mehrere Patientenlisten an einem Standort?

Das Hosting mehrerer Patientenlisten von verschiedenen Netzen an einem Standort ist grundsätzlich möglich. Dabei müssen alle Prozesse und Verantwortlichkeiten geklärt und dokumentiert sein – z.B. durch entsprechende SOPs. Zur Umsetzung der notwendigen Mandantenfähigkeit finden sich hilfreiche Hinweise in der entsprechenden Orientierungshilfe der Konferenz der Datenschutzbeauftragten des Bundes und der Länder [37]. Nicht geeignet ist allerdings das Hosting sämtlicher oder sehr vieler Patientenlisten bei einem einzigen Dienstleister, weil hier ein zu großes zentrales Angriffspotenzial entstünde.

Ebenfalls nicht geeignet ist das Konzept einer netzübergreifenden Liste für mehrere verschiedene Forschungsverbünde. Diese könnte zwar die Zugehörigkeit eines bestimmten Patienten zu einem bestimmten Forschungsverbund – und damit seine Diagnose – verschleiern. Da aber die Netzzugehörigkeit eines Patienten auch in einer netzübergreifenden Liste in irgendeiner Form vermerkt werden müsste, da sonst nicht überprüft werden kann, ob eine Anfrage nach den IDAT eines Patienten aus einem konkreten Netz heraus berechtigt ist oder nicht, wäre der Vorteil der Verschleierung auch in einer übergreifenden Liste nicht umsetzbar. Hierfür ist auch unerheblich, ob eine solche Anfrage schon im Ausschuss Datenschutz eines Netzes geprüft worden ist.

6.1.5.3 Sicherheit der Patientenliste

Die Patientenliste ist der sensibelste Teil des Identitäts-Managements und hat damit, wenn sie zentral geführt wird, eine besonders schützenswerte Rolle. Datenschutzrechtlich ist zu berücksichtigen, dass die IDAT, obwohl sie in der Patientenliste nicht mit medizinischen Daten (MDAT) kombiniert werden, den betroffenen Personenkreis als Patienten eines Forschungsnetzes mit einem umschriebenen Krankheitsspektrum ausweisen. Die Patientenliste ist daher unbedingt räumlich und technisch getrennt von den Datenbanken des Forschungsverbundes anzuordnen und auch einer getrennten disziplinarischen Verantwortung zu unterwerfen. Es muss ein praktikables und tragfähiges Sicherheitskonzept vorliegen, das sicherstellt, dass die Unabhängigkeit gewährt ist. Es empfiehlt sich, einer Partnereinrichtung des Forschungsnetzes diese zentrale Aufgabe zu übertragen, während die Datenbanken (KDB, SDB, FDB) bei anderen Partnern angesiedelt werden. Abweichend davon und bei besonders hohen Sicherheitsanforderungen besteht auch die Option, einen externen Datentreuhänder als TTP mit der Betreuung der Patientenliste zu beauftragen.

Hinweis: Der von der TMF bereitgestellte PID-Generator als Software-Implementierung der Patientenliste ermöglicht auch, die IDAT in einweg-verschlüsselter Form statt im Klartext abzulegen. Dies bewirkt einen zusätzlichen Schutz gegen das Reidentifizierungsrisiko, beeinträchtigt aber die manuelle Zuordnung in Zweifelsfällen und verhindert Anwendungen, bei denen eine Kontaktierung des Patienten erforderlich ist. Daher wird die Nutzung dieser Produkteigenschaft im Allgemeinen nicht empfohlen.

6.1.5.4 Lokalisierung des Pseudonymisierungsdienstes

Bei großen Forschungsverbünden – sobald für mehr als ein zeitlich beschränktes Forschungsprojekt pseudonymisiert werden muss – soll der Pseudonymisierungsdienst als zentraler Dienst selbstständig geführt werden. Zur Nutzung durch kleinere Verbünde wird empfohlen, den Pseudonymisierungsdienst als Dienstleistung von dritter neutraler Seite anzubieten, z.B. von der TMF selbst. Damit lässt sich verteilte Verantwortung kostengünstiger organisieren, als wenn jeder Forschungsverbund den Dienst selbst in einem eigenen Organisationsmodul realisiert.

6.1.6 Aspekte der Realisierung

6.1.6.1 Patientenliste

Die Patientenliste umfasst eine Funktion zur Erzeugung und Verwaltung der notwendigen Pseudonyme sowie zur Speicherung der zugehörigen Identitätsdaten (IDAT). Sie soll auf einem dedizierten Rechner geführt und in einem lokalen Netz geschützt aufgestellt werden. Die Kommunikation mit der Außenwelt erfolgt über einen kontrollierten Kanal (per Firewall-Tunnel) unter Nutzung des SSL-Protokolls oder gleichwertiger Lösungen.

Die TMF stellt als eine mögliche Komponente zur Umsetzung einer Patientenliste den PID-Generator zur Verfügung. Dieser kann

- online interaktiv über ein Web-Formular,
- offline im Batchbetrieb mit Datei-Übermittlung oder
- online als Web-Dienst aus einer externen Applikation (RDE-System) heraus

genutzt werden. Für letztere Nutzungsart ermöglicht die SOAP-Schnittstelle des PID-Generators in der jetzigen Implementierung die Client-Server- bzw. Server-Server-Kommunikation zwischen einer externen Applikation und der Patientenliste. Diese wird durch einen Webservice (SubjectList) realisiert und bietet Methoden zur Bearbeitung einer PID-Anforderung (Methode getSubjectID), zur Abfrage von Patientendaten (Methode getSubjectData) und zur Überprüfung der Gültigkeit eines PID (Methode isSubjectIDValid). Die Methode getSubjectID ruft den PID-Generator über die vorhandene CGI-Schnittstelle auf. Die Abfrage der Patientendaten bzw. der Validität eines PID wird direkt über eine SQL-Abfrage der Patientenliste durchgeführt.

Die folgenden Anforderungen, die mehrheitlich aus der hier neu vorgelegten Konzeption resultieren, sind in der bisher verfügbaren Version des PID-Generators noch nicht umgesetzt:

- Erzeugung und Verwaltung mehrerer zusammengehöriger pseudonymer Kennungen einschließlich deren Umwandlung,
- Entgegennahme und Verwaltung auch extern erzeugter Kennungen (z.B. SIC),
- Ausgabe geeigneter Zugriffstickets für die Kommunikation mit KDB, SDB und Pseudonymisierungsdienst,
- Überarbeitung und Erweiterung der Schnittstellen zur Kommunikation mit RDE-Software („SOAP-Schnittstelle"), KDB und SDB bzw. den dort angesiedelten Systemkomponenten des Pseudonymisierungsdienstes (s.u. in Kap. 6.1.6.2).

Die TMF wird zeitnah über mögliche Nachfolgeprodukte des PID-Generators informieren[29].

6.1.6.2 Pseudonymisierungsdienst

Die TMF hat eine Software zur Umsetzung eines Pseudonymisierungsdienstes[30] implementieren lassen, die von den folgenden Voraussetzungen ausgeht:

- Die Daten in Klinischen und Studiendatenbanken sind einfach pseudonymisiert mit einem eindeutigen PID_K oder PID_S.
- Der jeweilige PID ist für ein und dieselbe Person immer gleich, auch wenn diese Person in verschiedenen Einrichtungen behandelt wird oder an unterschiedlichen Studien zu unterschiedlichen Zeiten teilnimmt.
- Die Forschungsdatenbank kann das Attribut PSN speichern.

Über den Pseudonymisierungsdienst werden strukturierte Daten zwischen der Studien- (SDB) und der Forschungsdatenbank (FDB) ausgetauscht. Dazu müssen auf Seiten der SDB und der FDB Schnittstellen eingerichtet werden, um den Pseudonymisierungsdienst aufrufen und Daten in geeigneten Formaten übertragen zu können. Für diese Übertragungen gelten folgende Anforderungen:

Sie müssen je nach Richtung des Informationsaustausches einen PID_S (bei Nachrichten von der SDB an die FDB) oder ein PSN (bei Nachrichten von der FDB an die SDB) enthalten.

Die Nachricht kann über eine definierte Kennung (OrgDAT zum Vorgang) beschrieben werden, die Auskunft darüber erteilt, welche Reaktion von der

29 siehe http://www.tmf-ev.de/datenschutz-leitfaden

30 Im folgenden Textabschnitt wird die Bezeichnung Pseudonymisierungsdienst für das konkrete Softwareprodukt der TMF und nicht das davon unabhängige theoretische Konzept eines Pseudonymisierungsdienstes verwendet. Da sich sowohl für das Konzept als auch das Produkt mittlerweile die Bezeichnung Pseudonymisierungsdienst durchgesetzt hat, werden hier keine neuen separaten Namen eingeführt.

Gegenseite angefordert wird. Solche Reaktionen können sein: „Kontextdaten zu einem PID für die Qualitätssicherung senden“, „Datensatz in FDB abspeichern“, „Patient ein Finding anbieten“ u.a.

Weiterhin kann die Nachricht medizinische Daten (MDAT) enthalten. Der Pseudonymisierungsdienst geht davon aus, dass diese Daten bereits auf Seiten der jeweils sendenden Datenbank verschlüsselt werden und insofern unabhängig davon, ob die Übertragungswege zum Pseudonymisierungsdienst ebenfalls SSL-gesichert sind, niemals im Klartext außerhalb der beiden Datenbanken sichtbar sind. Innerhalb der MDAT dürfen niemals personenidentifizierende Angaben (Namen, PID, PSN, Versicherungsnummern o.ä.) enthalten sein, da die MDAT vom Pseudonymisierungsdienst nicht verändert, sondern lediglich in verschlüsselter Form weitergeleitet werden.

Um diese Voraussetzungen zu gewährleisten, sind auf Seiten der Datenbanken spezielle Komponenten erforderlich, um die Kommunikation mit dem Pseudonymisierungsdienst zu ermöglichen. Diese, als SDB- bzw. FDB-Komponente bezeichnet, sind Teil der Software des Pseudonymisierungsdienstes, werden aber nicht dort, sondern bei der jeweiligen Datenbank implementiert. Abbildung 13 zeigt die Architektur des Pseudonymisierungsdienstes.

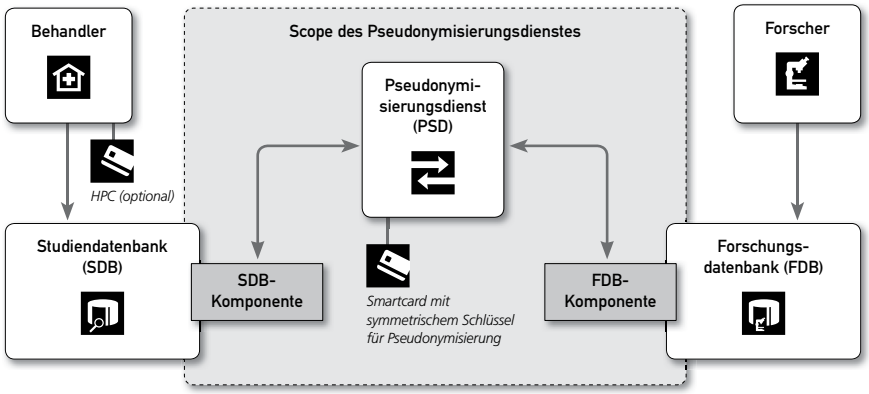

Abb. 13 Vorhandene Komponenten des Pseudonymisierungsdienstes

Der Pseudonymisierungsdienst spezifiziert folgende technische Dienste, die von den Komponenten genutzt werden können:

- PSD-Service: Dies sind die eigentlichen Services des Pseudonymisierungsdienstes, d.h. die Umwandlung eines PID_S in ein PSN und umgekehrt.
- FDB-Service: Dies sind generische und erweiterbare Services, die die FDB-Komponente bereitstellt.
- SDB-Service: Dies sind generische und erweiterbare Services, die die SDB-Komponente bereitstellt.

- Crypter: Mit diesen Services werden Nutzdaten (MDAT, Findings) (außerhalb des Pseudonymisierungsdienstes) zur Kommunikation zwischen SDB- und FDB-Komponente ver- und entschlüsselt.

Um die Vertraulichkeit der Daten zu gewährleisten, kommen die folgenden Schlüsseltypen im Rahmen des Pseudonymisierungsdienstes zum Einsatz:

- https-Keys (asymmetrische und symmetrische Schlüssel) für die SSL-Verschlüsselung,
- MDAT-Keys (asymmetrische Schlüssel) für die Datenverschlüsselung, die dem Pseudonymisierungsdienst selbst nicht bekannt sind,
- Pseudonymisierungsschlüssel (symmetrischer Schlüssel auf Smartcard).

Zur Pseudonymisierung wird ein symmetrischer kryptographischer Algorithmus hoher Sicherheit genutzt. Der Schlüssel ist gegen Auslesen gesichert auf einer Smartcard oder in einem Hardware Security Module gespeichert, deren wenige Exemplare von den für den Pseudonymisierungsdienst verantwortlichen Personen verwahrt und eingesetzt werden. Die Transformation des PID_S in ein PSN wird auf dieser Chipkarte durchgeführt, so dass der geheime Schlüssel die Karte nicht verlässt. Das Gleiche gilt für den umgekehrten Weg, bei dem ein PSN in einen PID_S transformiert wird.

Als Algorithmus wird mindestens DES-3 (Data Encryption Standard) vorgeschlagen, der in vielen marktgängigen Kartenchips implementiert ist; soweit von den Produkten her möglich, sollte der neue AES (Advanced Encryption Standard) genutzt werden (s. Kap. 2.2 des Kryptographischen Gutachtens im Anhang[31]).

Hinweis auf Weiterentwicklungsbedarf: Die Komponente Pseudonymisierungsdienst muss von zwei unterschiedlichen Ausgangskomponenten angesprochen und genutzt werden können. Sowohl aus der Klinischen Datenbank (KDB) wie aus einer Studiendatenbank (SDB) heraus muss eine weitere Pseudonymisierung angestoßen werden können, um Daten an die Forschungsdatenbank zu exportieren. Die aktuelle Implementierung sieht das Szenario eines Exports aus der KDB nicht vor und muss entsprechend erweitert werden. Zudem sollte die Vermittlung der MDAT optional auch ohne vollständige Durchleitung durch den PSD-Service möglich sein. Hierfür sollte ein entsprechendes Handling von Zugriffstickets vorgesehen werden (s. Abb. 9).

Folgende Komponenten müssten angepasst werden:

- PSD-Service: Es muss ermöglicht werden, dass dieser Service nicht nur vom SDB-Service und FDB-Service angesprochen werden kann, sondern auch von dem neu zu konzipierenden KDB-Service. Zudem muss er das Management von Zugriffstickets für die direkte Weitergabe von MDAT vom SDB- oder KDB-Service an den FDB-Service unterstützen.

31 Anhänge siehe www.tmf-ev.de/datenschutz-leitfaden

- KDB-Service: Dieser Service muss neu implementiert werden und analog zum bestehenden SDB-Service Aktionen ausführen. Wegen der unterschiedlichen Handhabung des PID_K (nicht in der KDB bekannt) muss in diese Komponente auch eine Kommunikation mit der Patientenliste, insbesondere die Handhabung eines Zugriffstickets (TKT), eingebaut werden, siehe Abbildung 14.
- SDB-Service: Hier ist eine Erweiterung insofern nötig, als dieser sowohl über einen SIC als auch über einen PID_S angesprochen werden können muss. Zudem ist das Handling von Zugriffstickets bei der direkten Versendung von MDAT an den FDB-Service umzusetzen.

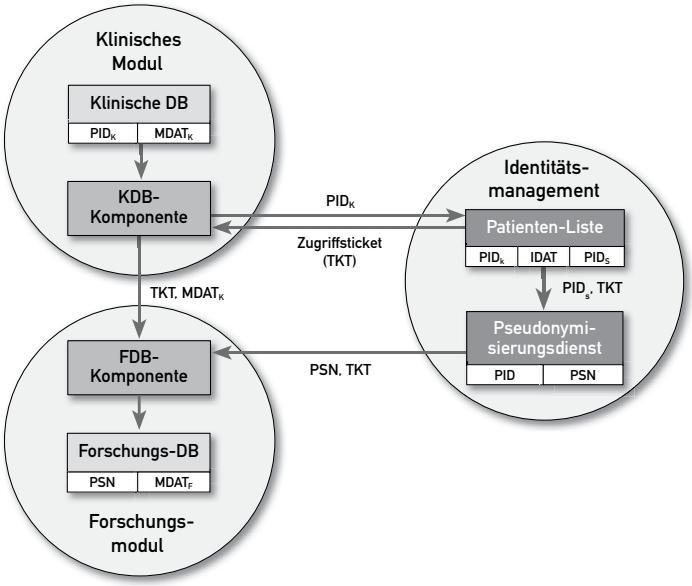

Abb. 14 Die KDB-Komponente des Pseudonymisierungsdienstes. TKT = Zugriffsticket

Ferner sollte die Implementation so gestaltet werden, dass im Maximalmodell von einem Prüfarzt des Studienmoduls, der gleichzeitig als behandelnder Arzt im Klinischen Modul wirkt, Daten aus beiden Modulen in einem Arbeitsgang an die FDB übermittelt werden können.

6.1.6.3 Übertragungssicherheit

Für die Übertragung aller relevanten Datenströme über das Internet ist ein geeignetes kryptographisches Protokoll zu nutzen. Für die meisten Anwendungsfälle (webbasierte Kommunikation oder Web-Dienste) ist SSL/TLS geeignet; es können aber auch sichere Lösungen auf VPN-Techniken aufgesetzt werden, siehe Kapitel 3.7 des Kryptographischen Gutachtens im Anhang.[32]

32 Anhänge siehe www.tmf-ev.de/datenschutz-leitfaden

6.1.7 Einordnung der bisherigen Datenschutzkonzepte der TMF

Die bisherigen Datenschutzkonzepte – die Modelle A und B des generischen Datenschutzkonzepts sowie das Datenschutzkonzept für Biomaterialbanken – ordnen sich in die im Maximalmodell beschriebenen Strukturen so ein, wie es Abbildungen 15–17 skizzieren.

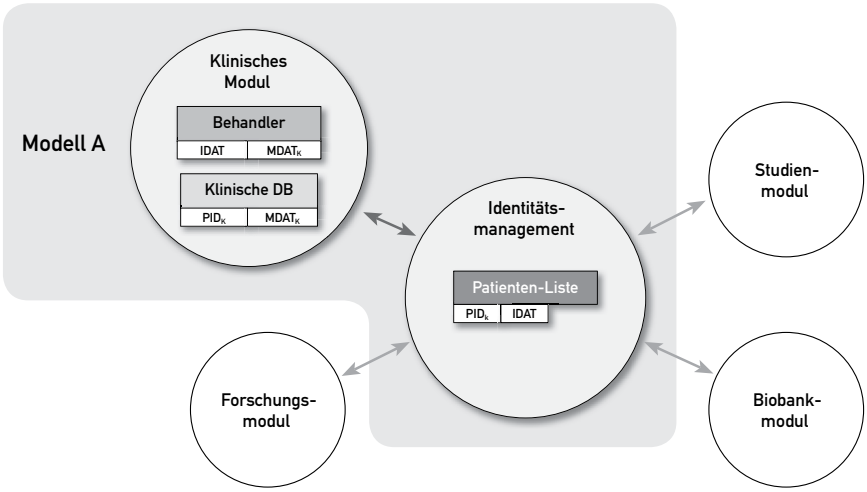

Abb. 15 Das Modell A des bisherigen generischen Datenschutzkonzepts in der übergeordneten Struktur

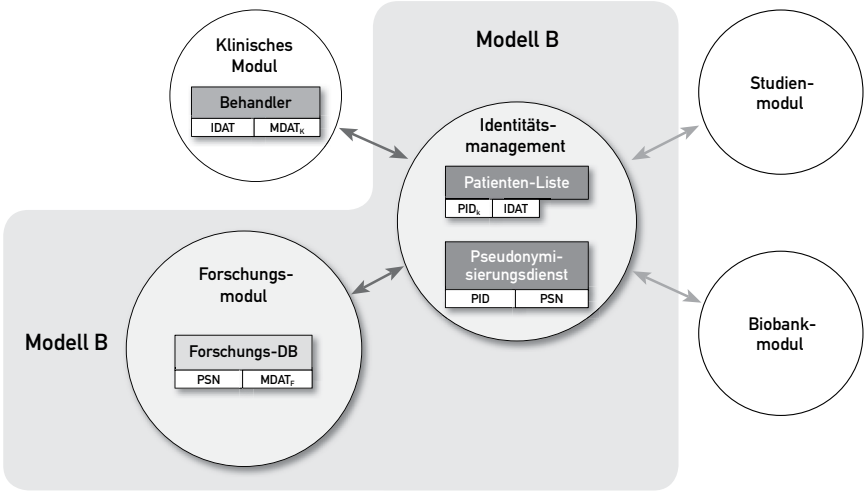

Abb. 16 Das Modell B des bisherigen generischen Datenschutzkonzepts in der übergeordneten Struktur. Für die Dateneingabe ist der Behandler aus dem Klinischen Modul nur beispielhaft dargestellt.

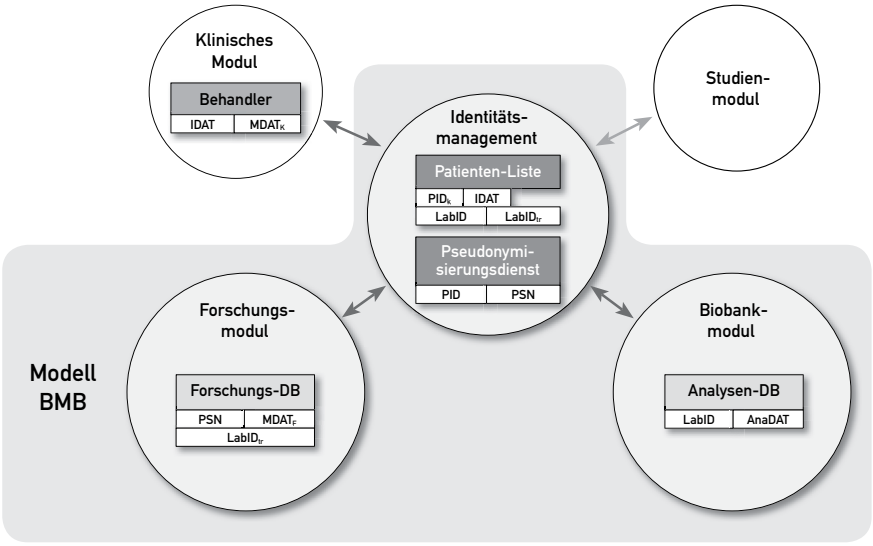

Abb. 17 Das generische Datenschutzkonzept für Biomaterialbanken in der übergeordneten Struktur. Für die Dateneingabe ist der Behandler aus dem Klinischen Modul nur beispielhaft dargestellt.

6.2 Rechtemanagement

Das Rechtemanagement betrifft die Mitarbeiter und Nutzer des Forschungsverbunds und soll u.a. gewährleisten, dass Informationen über Patienten und Studienteilnehmer nicht von Unberechtigten gesehen oder geändert werden können. Das Rechtemanagement bezieht sich auf die im Forschungsverbund eingesetzten IT-Systeme und besteht aus den Teilen:

- Authentisierung, d.h. manipulationssicherer Nachweis von Nutzer-Identitäten sowie
- Autorisierung, d.h. Vergabe von Zugriffsrechten auf Daten und von Ausführungsrechten für Funktionen.

Ist die sichere Authentisierung eines Nutzers gewährleistet, kann seine Autorisierung zur Ausübung von Funktionen im Netz und zum Datenzugriff anhand von Zugriffskontrolllisten und ähnlichen Mechanismen, die in der Regel in einem Datenbank- oder Studiensoftware-System implementiert sind, zuverlässig überprüft werden.

Das Rechtemanagement für die IT-Systeme eines Forschungsverbundes beruht auf dem Regelwerk des Forschungsverbundes, das in Policies ausgedrückt wird. In diesem Kapitel wird nur der technische Aspekt behandelt; auch dafür können nur einige grundsätzliche Aspekte beschrieben werden. Die Details der Umsetzung können sich sehr unterscheiden und sind Gegenstand des Sicherheitskonzepts des Forschungsverbunds. Grundsätzlich wird empfohlen,

- Policies zentral für einen Forschungsverbund und
- konkrete Zugriffsrechte dezentral in den einzelnen Modulen oder Datenbanken des Forschungsverbunds

zu verwalten; diese Aufteilung erscheint sowohl vom Arbeitsaufwand als auch im Hinblick auf die informationelle Gewaltenteilung zweckmäßig.

6.2.1 Zweck und Verwendungsbereich

6.2.1.1 Authentisierung von Nutzern

Authentisierung bedeutet, dass ein Nutzer seine behauptete Identität zweifelsfrei nachweist (sich ausweist); Authentifizierung, dass dieser Nachweis manipulationssicher überprüft wird (s. Abb. 18). Ein bekannter Authentisierungsmechanismus ist die Eingabe eines – zur Benutzerkennung passenden – Passworts, das im System (meist einweg-verschlüsselt) hinterlegt sein muss. Von starker Authentisierung spricht man, wenn stattdessen eine kryptographische Infrastruktur mit der Möglichkeit zur digitalen Signatur verwendet wird (s. Tab. 3); die Passwort-Eingabe wird dabei durch das digitale Signieren eines einmaligen Zufallswertes ersetzt (Challenge-Response-Verfahren), siehe Kapitel 3.6 des Kryptographischen Gutachtens im Anhang[33]. Niemand anders als der Besitzer des privaten Signaturschlüssels kann die korrekte Signatur erzeugen, und ein Angreifer kann mit dem erlauschten Zufallswert und der zugehörigen Signatur nichts anfangen. Dieses Verfahren wird typischerweise mit Smartcards realisiert. Ähnlich sicher kann die Authentisierung mit Hilfe von Hardware-Token gestaltet werden, die Einmal-Passwörter erzeugen.

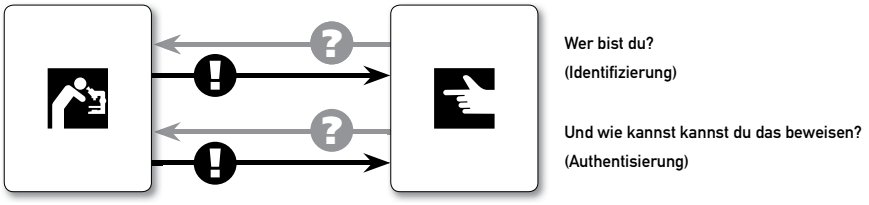

Abb. 18 Authentisierung

Tab. 3 Vergleich von schwacher und starker Authentisierung

	Frage	
Antwort bei …	**Wer bist Du?**	**Wie kannst Du das beweisen?**
Passwortverfahren (schwache Authentisierung)	Name	Passwort
Challenge-Response (starke Authentisierung)	Name (Zertifikat)	digitale Signatur

33 Anhänge siehe www.tmf-ev.de/datenschutz-leitfaden

Andere Authentisierungsverfahren, die auf der Überprüfung von biometrischen Merkmalen beruhen, sind in vernetzten IT-Systemen nicht ohne Weiteres praktikabel, da eine zentrale Datenbank dieser Merkmale mit sehr hohem Sicherheitsanspruch betrieben werden müsste. Für lokale Authentisierungsvorgänge sind aber insbesondere Fingerabdruckscanner durchaus geeignet, z.B. um Nutzer an ihrem Arbeitsplatzrechner oder für die Nutzung ihrer Smartcard zu authentifizieren. Die Marktentwicklung in diesem Bereich sollte beobachtet werden.

6.2.1.2 Rollen und Rechte im Forschungsverbund

Aufgrund einer sicher vollzogenen Authentifizierung eines Nutzers wird seine Autorisierung zur Ausübung von Funktionen im Netz zuverlässig anhand von Rechtedefinitionen festgelegt, die in Policies und Rollenbeschreibungen formuliert und in Zugangskontrolllisten o.ä. abgelegt sind.

Rechte im Forschungsverbund betreffen

- Datenzugriffe und die Verarbeitung von Daten sowie
- die Administration der IT-Systeme und der Infrastruktur mit ihren Komponenten.

Rechte sind in der Regel an Rollen gebunden, wie z.B. „Forscher“, „Systemadministrator der Forschungsdatenbank“, und werden an Einzelpersonen über deren Zuordnung zu Rollen vergeben. Die für die einzelnen Module und Komponenten eines Forschungsverbunds relevanten Rollen und Rechte werden jeweils dort beschrieben.

Ein zentrales Nutzer- und Rollenverzeichnis (z.B. Active Directory) kann für die Rechte- und Rollenverwaltung gute Dienste leisten, erscheint aber in einem verzweigten und heterogenen Forschungsverbund kaum mit angemessenem Aufwand realisierbar. Daher wird empfohlen, Regelwerke in Form von Policies zentral (als Texte) zu verwalten und in jeweils dezentraler Nutzerverwaltung und Rechtevergabe umzusetzen. Die für bestimmte Zugriffsentscheidungen benötigten ADAT werden im generischen Fall in der Patientenliste, u.U. aber auch bei den MDAT gespeichert, siehe Kapitel 6.1.5.1 und 6.5.2.4.

6.2.1.3 Zugriffsentscheidungen

Um eine Zugriffsentscheidung treffen zu können, muss das jeweilige IT-System oder der Netzdienst folgende Informationen zur Verfügung haben:

- Identität des Zugreifenden (authentifiziert),
- Rolle des Zugreifenden,
- Definition der Rechte, die mit diesem Nutzer und dieser Rolle verbunden sind.

Für die Verwaltung der Zugriffsrechte auf Objekte (IT-Systeme oder Netzdienste, Daten oder Prozesse) gibt es prinzipiell verschiedene Ansätze (s. Abb. 19):

1. Objekte (bzw. die sie tragenden IT-Systeme) verwalten sich selbst, d.h., sie prüfen bei einer Anfrage eines authentifizierten Partners (Person oder Prozess) anhand der in ihnen selbst implementierten Regeln, wie sie antworten oder reagieren wollen. Ein solcher dezentraler Ansatz benötigt nur ein Minimum an Vertrauensannahmen (nämlich in die zuverlässige Authentisierung), stößt aber sehr schnell an Komplexitätsgrenzen.
2. Es werden vertrauenswürdige Dienste genutzt, die die Entscheidung auf sichere Weise treffen und übermitteln können; dies kann wiederum auf zwei Weisen geschehen:
   - online durch Abfrage eines TTP-Dienstes, der eine Entscheidung der Art „erlaubt“ oder „nicht erlaubt“ zurückliefert,
   - offline, durch Prüfung eines „Credentials“, also eines von einem TTP-Dienst signierten Attributs (Zugriffsticket), das die Berechtigung ausdrückt und vom Antragsteller präsentiert wird.

Beispiele: Die Patientenliste enthält auch die Information, wer als zugriffsberechtigter behandelnder Arzt für einen Patienten beim Forschungsnetz erfasst ist (ADAT, s. Kap. 6.1.1.1). Diese Information kann z.B. an eine Klinische Datenbank (KDB) weitergegeben werden. Rechte, die in einer Studiendatenbank (SDB) festgelegt sind, können an andere Anwendungen im Forschungsverbund mitgeteilt werden.

6.2.2 Anwendungsfälle

6.2.2.1 Anwendungsfälle und ihre empfohlene Lokalisierung

Tabelle 4 stellt dar, welche Lokalisierungen für die verschiedenen Anwendungsfälle empfohlen werden.

Tab. 4 Lokalisierung von Anwendungsfällen

Anwendungsfall	empfohlene Ansiedlung
Anlegen und Administrieren von Nutzerkonten	dezentral, evtl. in zentralem Verzeichnis
Anlage und Verwaltung von Rollen	zentral oder dezentral nach zentral vorgegebenen Policies
Zuordnung von Nutzern zu Rollen	dezentral
Definition von Policies	zentral
Definition von Rechten	dezentral
Zuordnung von Rechten zu Rollen	zentral oder dezentral
Verteilung der Nutzer- und Rechtedaten an Subsysteme (z.B. einzelne Datenbanken)	von zentral nach dezentral, evtl. auch von dezentral nach dezentral
Prüfung von Rechten	dezentral, evtl. durch zentrale Dienste unterstützt

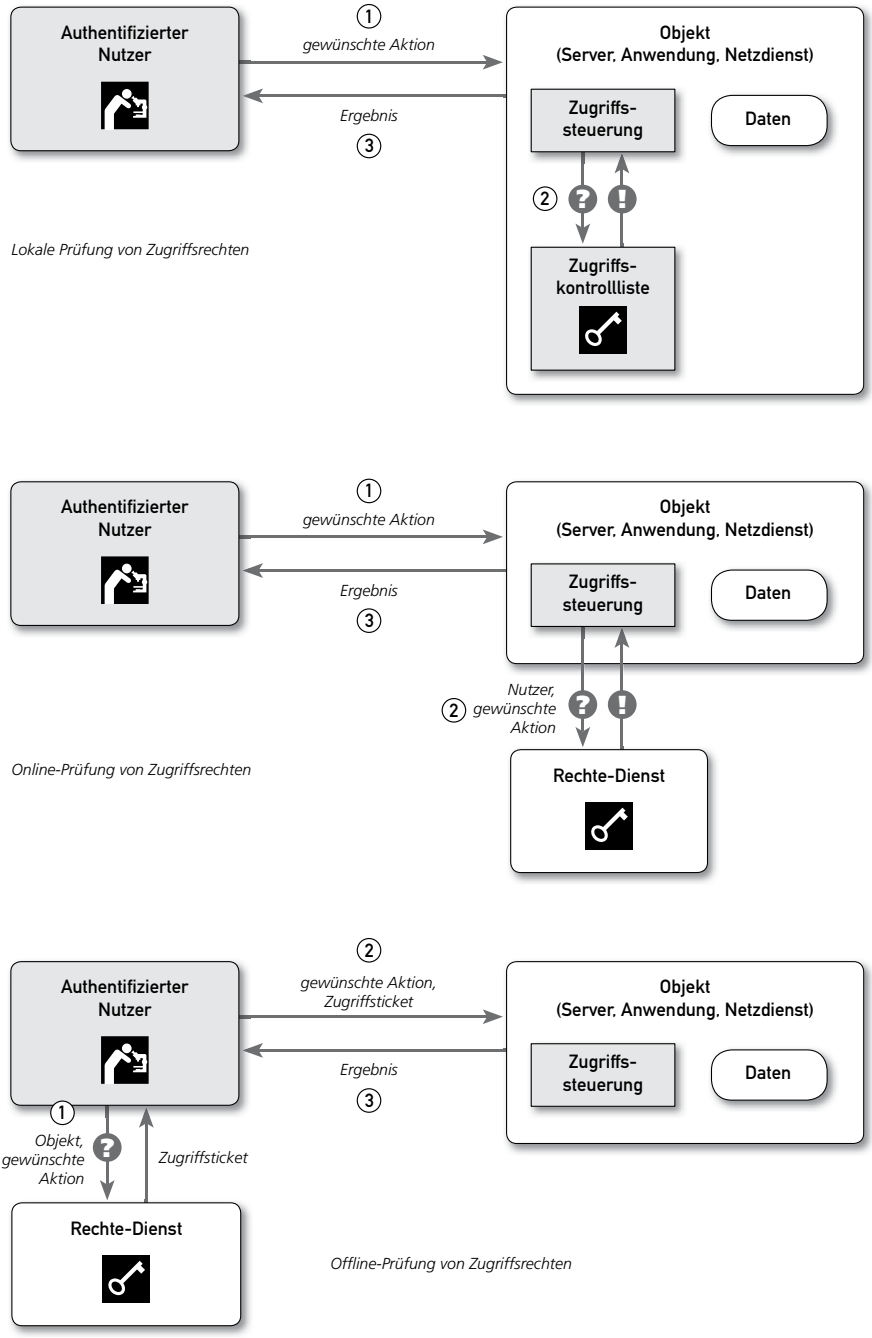

Abb. 19 Drei Modelle der Zugriffsentscheidung

Als Anwendungsfälle kommen, je nach konkreter Implementierung, die Verwaltung von zentralen Diensten sowie der Datenbanken im Forschungsverbund und evtl. einer PKI hinzu.

6.2.2.2 Benötigte Dienste

Falls das Rechtemanagement im Forschungsverbund überhaupt zentral organisiert wird, sind eine Reihe von Sicherheitsdiensten nötig, die als Trusted Third Party Services aufzusetzen sind, d.h. – vom technischen Standpunkt aus betrachtet – als Netzdienste oder Web-Services, die sowohl interaktiv als auch von Prozessen in Anspruch genommen werden können und das Vertrauen aller am Netz Beteiligten genießen. Die wichtigsten davon sind

- Benutzerverwaltung,
- PKI- und Zertifikateverwaltung,
- Authentisierungsdienst,
- Rollenverwaltung, in der Regel in Form von Benutzergruppen, mit dezentraler Zuordnung von Benutzern zu Rollen,
- Policy-Dienste: Definition, Pflege und Interpretation von Sicherheitsrichtlinien, die vor allem Zugriffsberechtigungen betreffen und durch generische Zugriffsregeln ausgedrückt werden,
- Zugriffskontroll- oder Autorisierungsdienste: konkrete Umsetzung der Policies in Zugriffsentscheidungen (auch dynamische und kontextsensitive, Workflow-abhängige),
- Gateways zwischen Modulen oder Teilbereichen mit unterschiedlichen Policies.

6.2.3 Nutzer, Rollen und Rechte

6.2.3.1 Generische Rollen im Forschungsverbund

Die durch die primären Aufgaben des Forschungsverbunds definierten Rollen (behandelnder Arzt, Prüfarzt, Studienleiter, Forscher) sind bei den einzelnen Modulen definiert; siehe auch das Glossar.

Daneben gibt es eine Reihe von implementationsabhängigen, durch die IT-Architektur des Forschungsverbundes definierten Rollen, hauptsächlich Systemadministratoren und Nutzer für die eingerichteten Dienste und Datenbanken; diese werden im jeweiligen Kontext beschrieben.

6.2.3.2 Rollen im Rechtemanagement

Hier gibt es die Rolle der Systemverwalter für alle separat betriebenen zentralen Komponenten, z.B. für ein zentrales Verzeichnis, sowie zur dezentralen Rechteverwaltung auf Servern und für Dienste.

6.2.3.3 Mögliche Rollenkonflikte

Einzelne klinisch tätige Ärzte sind gleichzeitig auch als Wissenschaftler in ihren jeweiligen Forschungsnetzen tätig; dabei besteht das Risiko, dass ein solcher Arzt Daten von einem seiner früheren Patienten, der inzwischen bei einem anderen Arzt in Behandlung war, trotz Pseudonymisierung wieder erkennt und somit unbefugt Informationen erhält. Dieser interpersonelle Konflikt ist jedoch nicht neu, auch bei der konventionellen Papierdokumentation in der klinischen Forschung bestehen ähnliche Probleme, wobei die Wiedererkennbarkeit sogar erleichtert ist. Zudem sind hier die Zugriffsregeln nicht elektronisch einstellbar, so dass die Regel „wenig Zugriff ist voller Zugriff" bei der konventionellen Datenhaltung zutrifft. Im elektronischen Verfahren lässt sich die Behandlung der „Doppelrolle" durch eine rollenbasierte Zugriffsberechtigung, die in diesem Fall zwei unterschiedliche Profile für einen Mitarbeiter vorsieht, regeln. Ein bewusster, vorsätzlicher Missbrauch dieses Konzeptes – wie auch bei der Papierlösung – lässt sich aber naturgemäß nicht restlos verhindern, und ist durch die ärztliche Schweigepflicht sowie durch die Regeln des Forschungsverbundes auszuschließen. Hier sei auch auf das Rechtsgutachten [11] verwiesen. Ähnliches gilt, wenn Arzt und Systemadministrator in einer Person verkörpert werden.

Weiter kann ein Rollenkonflikt entstehen, wenn eigentlich getrennt zu haltende Datenbestände wie z.B. identifizierende Daten (IDAT) und medizinische Daten (MDAT) in derselben Institution gespeichert werden und Zugriffe von Mitarbeitern auf beide Datenbestände nicht sicher genug ausgeschlossen werden können. In solchen Fällen ist kritisch zu prüfen, ob eine solche Vereinfachung nach den Kriterien der Verhältnismäßigkeit (vgl. Kap. 6.7) vertretbar ist. Gerade in großen Forschungsverbünden, die das hier vorgeschlagene Maximalmodell implementieren und je Modul ggf. auch über mehrere Datenbanken verfügen, ist eine besonders sorgfältige Prüfung auf mögliche Rollenkonflikte bei allen Beteiligten unumgänglich.

6.2.4 Verantwortlichkeiten

Für die grundsätzliche Definition von Rechten und Policies ist die Leitung des Forschungsverbundes zuständig. Diese Aufgabe kann an den Ausschuss Datenschutz delegiert werden.

Die Dienste für das Rechtemanagement nach Kapitel 6.2.2.2 können zentral oder dezentral angeordnet werden. Bei zentraler Anordnung besteht noch die Wahl zwischen einer am Netz beteiligten Einrichtung und einem externen Dienstleister. Wie solche TTP-Dienste rechtlich und organisatorisch aufgesetzt werden, hängt vom rechtlichen und organisatorischen Status der Anwendungsumgebung ab. Für ein medizinisches Forschungsnetz wird man rechtlich oder vertraglich verpflichtete Organisationen wählen, die ein hohes Sicherheitsniveau garantieren können.

Das Nutzer- und Rechtemanagement ist in hohem Maße abhängig von bestehenden Infrastrukturen und Workflows in den Forschungsnetzen. Es wird aber jedenfalls von allen Modulen vorausgesetzt und benutzt. Da das Identitätsmanagement an zentraler Stelle im Forschungsverbund steht, liegt es nahe, die zentralen Funktionen des Rechtemanagements ebenfalls hier anzusiedeln. In einem großen Forschungsverbund ist eine Trennung der Funktionen im Sinne der informationellen Gewaltenteilung zu empfehlen. Hier ist aber, abhängig von Größe und Struktur des Forschungsverbundes, die Verhältnismäßigkeit des Aufwandes zu wahren. Unter dem Gesichtspunkt des Datenschutzes können sowohl zentrale als auch dezentrale Lösungen sicher gestaltet werden.

6.2.5 Aspekte der Realisierung

In der Praxis gibt es für das Rechtemanagement und seine Komponenten viele unterschiedliche technische Lösungen; keine davon ist als allgemein etablierter Standard anzusehen. Am weitesten verbreitet dürfte nach wie vor ein separates Rechtemanagement in jeder selbstständig administrierten Komponente eines Forschungsverbundes sein, in der Regel mit Passwort-basierter Authentisierung auf jedem einzelnen Server, sowie die in der jeweiligen Datenbank vorgesehene Regelung von Zugriffsrechten; dies ist aufgrund der Marktdominanz dieser Verfahren bei weitem am einfachsten umzusetzen. Daher ist das Rollenmanagement auf der Seite der konkreten Implementierung in der Regel nicht als Modul oder Komponente abgrenzbar.

Dennoch wird empfohlen, die Verwaltung von Nutzern bei geeigneten Ressourcen möglichst zentral zu organisieren, auf jeden Fall aber die Definition von Policies und möglichen Rollen. Allerdings sollen die Rechte für jeden Teil der Infrastruktur gesondert administriert werden können: Die in diesem Konzept an vielen Stellen geforderte informationelle Gewaltenteilung wird nur wirksam umgesetzt, wenn die disziplinarisch unabhängigen Stellen im Netz über die jeweilige Rechtevergabe selbst wachen. Dies impliziert insbesondere die Zuordnung von Nutzern zu Rollen auf der Ebene der einzelnen Module. Zur Nutzerverwaltung und Rollenvergabe werden also dezentrale Zugriffsrechte benötigt.

6.2.5.1 Nutzung eines Verzeichnisdienstes

Als zentrale Komponente dafür wird ein Verzeichnisdienst (Directory) empfohlen, der aber dezentralen Systemadministratoren für die Verwaltung ihrer jeweiligen Nutzer zugänglich ist. Dieser ermöglicht eine zentrale Authentisierung (Single-Sign-On). Unter dem Gesichtspunkt des Datenschutzes ist aber eine völlig dezentrale Nutzerverwaltung ebenso akzeptabel.

6.2.5.2 Nutzung einer PKI

Die Public-Key-Infrastruktur (PKI) sorgt für ein sicheres Management privater und öffentlicher Schlüssel. Für den privaten Schlüssel – der ja als persönliches Geheimnis zu behandeln ist – ist ein sicherer Aufbewahrungsort vorzusehen, den der Schlüssel möglichst nicht verlassen muss. Ideal geeignet ist eine Chipkarte (Smartcard), die auch in der Lage ist, die kryptographischen Grundfunktionen Verschlüsselung, Signatur und starke Authentisierung auszuführen.

Öffentliche Schlüssel müssen dagegen nicht geheim gehalten werden, aber ihre Authentizität muss gesichert werden. Dazu dienen Zertifikate. Sie setzen voraus, dass eine von allen Teilnehmern anerkannte vertrauenswürdige Zentralinstanz existiert, die durch digitale Signatur den öffentlichen Schlüssel an eine eindeutige Kennzeichnung seines Besitzers bindet. Eine solche Instanz wird Trustcenter oder Certification Authority (CA) genannt und ist ein Beispiel für eine Trusted Third Party (TTP).

Aufbau und Betrieb einer PKI sind Standardaufgaben, zu denen es zahlreiche bestehende Verfahrensvorschriften und Softwareprodukte gibt. Für medizinische Forschungsverbünde wird aber empfohlen, keine eigene PKI aufzubauen, sondern mittelfristig die der zukünftigen Gesundheitstelematik, insbesondere den Heilberufeausweis (HBA) zu nutzen. Die Möglichkeiten hierzu folgen aus dem Rechtsgutachten von Roßnagel und Mitarbeitern [11].

6.2.5.3 Technische Aspekte der Rechtevergabe

Wird mit einem rollenbasierten Ansatz gearbeitet, so ist keine explizite Verteilung von Rechtedaten nötig. Die Zuordnung von Nutzern zu Rollen wird in der Nutzerverwaltung dezentral (selbst wenn es ein zentrales Nutzerverzeichnis gibt) durchgeführt, das rollenbasierte Rechtemanagement wird lokal an den Servern auf der Basis der netzweit geltenden Policies eingestellt.

Für die Verteilung der Rechte-Informationen geeignete Methoden und Werkzeuge werden nachfolgend aufgeführt. Solche Informationen können entweder als Zugriffsentscheidung auf geschütztem Wege von einem zentralen Server an die jeweiligen Dienste oder Datenbanken übermittelt werden, oder der jeweilige anfragende Nutzer erhält diese in Form eines Credentials (Zugriffstickets), d.h. einer vom zentralen Dienst digital signierten „Erlaubnisbescheinigung".

6.2.5.4 Spezifikation von Richtlinien und Regeln

Richtlinien und Regeln eines medizinischen Forschungsverbundes werden in der Regel in Textform beschrieben und dezentral in den einzelnen Modulen und Komponenten entsprechend implementiert. Für große und komplexe Verbünde kommt aber auch die Einführung und Nutzung von technischen Werkzeugen in Betracht, wenn sich der Investitionsaufwand hierfür lohnt.

Wichtige entsprechende Werkzeuge für die Einrichtung von Sicherheitsdiensten sind standardisierte Sprachen, mit denen Richtlinien und Regeln eindeutig spezifiziert und automatisiert verarbeitet werden können; gängige Ansätze hierfür sind z.B.

- SAML = Security Assertion Markup Language, eine XML-basierte Auszeichnungssprache zur Beschreibung von sicherheitsbezogenen Informationen.
- XACML = eXtensible Access Control Markup Language: ein XML-Schema, das die Verwaltung von Policies (im engeren Sinne: Rechtevergaberegeln) definiert. XACML definiert eine Sprache, in der Zugriffsberechtigungen durch Attribute, Bedingungen und Regeln ausgedrückt und zwischen verschiedenen Diensten und Prozessen kommuniziert werden können. Dadurch lassen sich wesentlich komplexere Zugriffsregeln ausdrücken als durch einfache Zugriffslisten (ACL = Access Control List).

6.2.5.5 Weitere mögliche Werkzeuge

Auch hier handelt es sich um eine Aufzählung von Werkzeugen, deren Nutzung einer Aufwands- und Nutzenabschätzung unterliegt und die nicht generell für alle Forschungsverbünde empfohlen wird.

- Kerberos ist ein verteilter Authentisierungsdienst für Computernetze.
- Shibboleth ist eine Sammlung von Diensten, die lokalen Authentisierungs- und Autorisierungsdiensten ermöglicht, fremden Diensten die nötigen Informationen für Zugriffsentscheidungen zur Verfügung zu stellen
- VOMS (Virtual Organization Membership Service) ist ein datenbankgestützter Mechanismus zur zentralen Verwaltung von Rollen und Rechten (globale Autorisierung)

Hinweis: Erfahrungen mit dem Einsatz solcher Werkzeuge liegen bisher nur im Grid-Umfeld vor. Die Nutzbarkeit für medizinische Forschungsverbünde müsste erst noch in einem Pilotprojekt geprüft werden, bevor konkrete Empfehlungen zum Einsatz formuliert werden können.

6.3 Kombinierter Einsatz von Studienmodul und Klinischem Modul

6.3.1 Zweck und Anwendungsbereich

In den Kapiteln 5.1 und 5.2 werden die Konzepte eines Klinischen Moduls zur versorgungsnahen Datenerhebung sowie eines Studienmoduls zur Durchführung einzelner Forschungsprojekte (z.B. klinischer Studien) getrennt voneinander beschrieben. Im Folgenden soll der kombinierte Einsatz eines Klinischen Moduls und eines Studienmoduls innerhalb eines Forschungsnetzwerks dargestellt werden. Die Module werden hierbei auf Basis der im jeweiligen

Kapitel beschriebenen Konzepte eingesetzt. Abweichungen sowie Besonderheiten im Zusammenspiel der Module werden zusätzlich beschrieben.

Im Rahmen seltener sowie chronischer Erkrankungen kann sich die Notwendigkeit ergeben, Patienten längerfristig im Rahmen eines Forschungsnetzes zu behandeln. Hierdurch wird zum einen die Möglichkeit zur kontrollierten longitudinalen Erhebung und Auswertung von Daten geschaffen, zum anderen können Patienten auf Basis der im Rahmen der Versorgung erhobenen Daten für Studien im Rahmen des Forschungsnetzes rekrutiert werden. Es entstehen Überschneidungen zwischen den im Rahmen der Routineversorgung und für die Studiendokumentation benötigten Daten, die sowohl zum Vorteil des Patienten aus der Forschung in die Versorgung übernommen (z.B. Nutzung eines aufwändigen Bildgebungsverfahrens für die Versorgung) als auch zur Vermeidung einer Mehrfacherfassung von der Versorgung in die Forschung übertragen werden können. Voraussetzung ist das Vorliegen der Daten in einer für den jeweiligen Anwendungszweck notwendigen Qualität und Vollständigkeit.

Durch die Verbindung von Studien- und Klinischem Modul ergeben sich zusätzliche Datenflüsse für die Datenübernahme, die sowohl bei der Umsetzung dieser Komponenten als auch beim ID-Management berücksichtigt werden müssen.

Im Rahmen des Zusammenflusses von Studien- und Versorgungsaspekten kann sich auch eine Personeneinheit von Studienarzt und behandelndem Arzt ergeben. Die Zugriffsrechte ergeben sich dabei aus der Vereinigungsmenge der beiden Rollen. Um Doppelerfassungen zu vermeiden, sollte nach Möglichkeit die Eingabe über ein System erfolgen, von dem aus die Daten geschützt an das jeweilige andere System übertragen werden.

Informationssysteme der Routineversorgung (z.B. Krankenhausinformationssysteme, Praxisverwaltungssysteme oder elektronische Patientenakten) können Bestandteile eines Klinischen Moduls sein. Ebenso können Daten aus separat betriebenen Routineversorgungssystemen mit dem Klinischen Modul ausgetauscht werden. Das generische Datenschutzkonzept der TMF deckt in diesen Fällen nur die zweckgebundene Nutzung im Kontext des Klinischen Moduls ab, es soll kein allgemeines Datenschutzkonzept für den Betrieb von Systemen der klinischen Routineversorgung abgebildet werden.

6.3.2 Anwendungsfälle und Prozesse

Die Anwendungsfälle des Studien- und des Klinischen Moduls werden ausführlich in den entsprechenden Kapiteln beschrieben. In diesem Kapitel wird der Fokus auf die Prozesse gelegt, die sich aufgrund der Verbindung des Studienmoduls mit dem Klinischen Modul ergeben.

6.3.2.1 Daten zwischen Studienmodul und Klinischem Modul übermitteln

Wenn ein Patient eines Forschungsverbundes sowohl an einer Studie als auch am Klinischen Modul teilnimmt, so besteht die Möglichkeit, die schon in einem Modul erfassten Daten des Patienten in das andere Modul zu übertragen, um Doppelerfassungen zu vermeiden. Um diese Übertragung zu ermöglichen, bedarf es eines netzweiten Identitätsmanagements (s. Kap. 6.1), das sowohl das Pseudonym des Patienten aus dem Klinischen Modul (PID_K) als auch das aus dem Studienmodul (PID_S bzw. SIC_S für jede Studie) enthält.

Bei der Übertragung aus dem Studienmodul in das Klinische Modul werden zwei Fälle unterschieden:

a) Der verantwortliche Arzt stellt eine entsprechende Anfrage von Seiten des Klinischen Moduls an das Studienmodul (s. dazu auch Abb. 20). Hierzu schickt er eine Nachricht mit den identifizierenden Daten des Patienten an das Identitätsmanagement. Das Identitätsmanagement authentifiziert und autorisiert den Arzt, selektiert anhand der identifizierenden Daten des Patienten die Pseudonyme PID_K und PID_S bzw. SIC und erstellt ein Zugriffsticket (TKT). Das TKT wird mit dem PID_K des Patienten an das Klinische Modul und mit dem PID_S bzw. SIC an das Studienmodul geschickt. Das Studienmodul selektiert anhand des PID_S bzw. dem SIC die medizinischen Daten ($MDAT_S$) des Patienten, entfernt den PID_S bzw. den SIC und schickt die Daten mit dem TKT an das Klinische Modul. Das Klinische Modul ordnet den Datensatz anhand des TKT dem PID_K des Patienten zu und speichert die Daten zu dem entsprechenden PID_K.

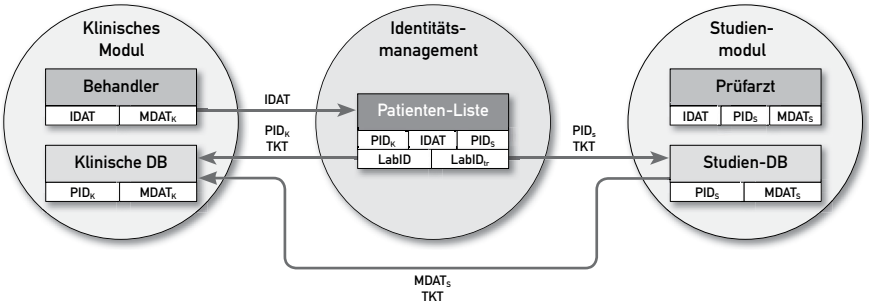

Abb. 20 Transfer von medizinischen Daten eines Patienten aus dem Studienmodul in das Klinische Modul, angestoßen vom Klinischen Modul.

b) Der Datentransfer wird auf Seiten des Studienmoduls angestoßen (dies kann beispielsweise nach Abschluss einer Studie automatisch erfolgen, s. dazu auch Abb. 21). Hierzu muss der PID_S des entsprechenden Patienten an das Identitätsmanagement geschickt werden. Das Identitätsmanagement authentifiziert und autorisiert den entsprechenden Anfragenden, ordnet dem PID_S des Patienten seinen PID_K zu und erstellt ein Zugriffsticket. Das TKT wird mit dem

PID_K des Patienten an das Klinische Modul geschickt und mit dem PID_S an das Studienmodul. Die Selektion und Übertragung erfolgt wie oben bereits beschrieben.

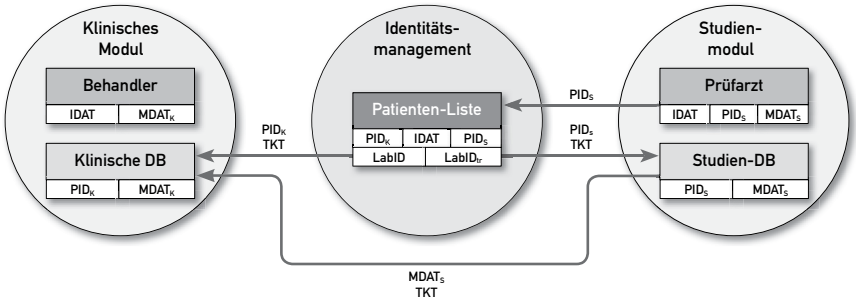

Abb. 21 Transfer von medizinischen Daten eines Patienten aus dem Studienmodul in das Klinische Modul, angestoßen vom Studienmodul.

Die Übertragung aus dem Klinischen Modul an das Studienmodul kann auch aus beiden Richtungen angestoßen werden. Die Mechanismen sind ebenfalls die gleichen. Die Übertragung der Daten des Patienten erfolgt in diesem Fall vom Klinischen Modul an das Studienmodul.

Ziel der Verwendung eines durch das Identitätsmanagement vermittelten Zugriffstickets ist es, eine Verknüpfung von PID_K und PID_S bzw. SIC außerhalb der Patientenliste zu verhindern. Eine direkte Kommunikation der medizinischen Daten des Patienten (in verschlüsselter Form) über die Patientenliste und damit die Möglichkeit, die Pseudonyme direkt in den medizinischen Daten auszutauschen (angelehnt an das Verfahren des Pseudonymisierungsdienstes), wurde nicht gewählt, um die klare räumliche und organisatorische Trennung der medizinischen Daten von den identifizierenden Daten des Patienten aufrecht zu erhalten. Das oben beschriebene Verfahren wird als eine mögliche Variante zur datenschutzkonformen Übertragung medizinischer Daten eines Patienten zwischen Klinischem und Studienmodul gesehen. Sofern die gerade beschriebenen Schutzprinzipien eingehalten werden, sind auch andere Umsetzungsmöglichkeiten vorstellbar.

Eine relevante Standardisierungsinitiative für solche Datentransfers stellt das Profil Retrieve Form for Data Capture (RFD) der Organisation Integrating the Healthcare Enterprise (IHE) dar[34]. Auch wenn dieser Vorschlag auf das notwendige korrekte Handling der Pseudonyme derzeit noch nicht detailliert eingeht, könnten die hier spezifizierten Kommunikationsstandards für künftige Softwaresysteme eine wichtige Anforderung darstellen.

34 siehe http://wiki.ihe.net/index.php?title=Retrieve_Form_for_Data_Capture

6.3.2.2 Patienten in Klinisches Modul oder Studienmodul aufnehmen

Beim Einholen der Einwilligungserklärung sollte mit abgefragt werden, ob der Patient auch an möglichen Studien Interesse hat bzw. später auf entsprechende Studien hingewiesen werden möchte. Des Weiteren sollte abgefragt werden, ob der Patient schon an Studien im Rahmen des Forschungsverbundes teilgenommen hat und ob diese Daten im Klinischen Modul genutzt werden sollen. Ist dies der Fall, so können die Daten wie oben beschrieben aus dem Studienmodul an das Klinische Modul übertragen werden.

6.3.2.3 Datenqualität sichern

Ergänzend zu üblichen Qualitätssicherungsverfahren in einzelnen Studien kann bei Kombination mit einem Klinischen Modul auch ein Abgleich von Daten aus der Versorgung zu einem Patienten durchgeführt werden. Sollte sich dabei z.B. herausstellen, dass sich das Geburtsjahr geändert hat oder bei einem erwachsenen Patienten eine deutlich veränderte Körpergröße festgestellt wurde, so wären dies gerechtfertigte Auslöser für eine weitere Datenüberprüfung. Aus Sicht des Klinischen Moduls können ebenfalls Daten einer Studie für die Qualitätssicherung herangezogen werden.

Sollten bei der Qualitätssicherung im Klinischen Modul Änderungen an Daten erfolgen, die ihren Ursprung im Studienmodul haben und in das Klinische Modul übernommen wurden, so sollte sichergestellt werden, dass die Verantwortlichen des Studienmoduls informiert werden. Hierbei können die entsprechend zu ändernden Daten des Patienten mit dem Hinweis auf den Fehler wie oben beschrieben zwischen dem Klinischen Modul und dem Studienmodul ausgetauscht und entsprechende Fehler korrigiert werden. Bei Korrekturen im Studienmodul wird äquivalent verfahren. Für diese Vorgehensweise müssen Datensätze, die in das jeweils andere Modul übertragen worden sind, entsprechend gekennzeichnet werden.

6.3.2.4 Daten erheben

Werden teilweise die gleichen Daten eines Patienten im Rahmen der Versorgung sowie einer Studie erfasst, so ist es sinnvoll, diese nur in einem System zu erfassen und sie anschließend im Studienmodul und Klinischen Modul zu speichern. Je nach Workflow kann hier das Studienmodul oder das Klinische Modul das erfassende System sein. Die Übertragung der Daten an das andere System kann nach dem oben beschrieben Verfahren erfolgen.

6.3.2.5 Studiendaten auswerten

Nach der Auswertung einer Studie können die entsprechenden Ergebnisse der Studie für den Patienten im Klinischen Modul bereitgestellt werden. Hierzu

werden die Ergebnisse mit dem PID_S des Patienten versehen und nach dem gleichen Verfahren wie die Studiendaten an das Klinische Modul übertragen.

6.3.2.6 Unerwartete Ereignisse managen

Sollten während einer Studie unerwartete Ereignisse zu einem Patienten eintreten, so können diese Informationen mit dem PID_S des Patienten versehen und nach dem gleichen Verfahren wie die Studiendaten an das Klinische Modul übertragen werden.

6.3.2.7 Studiendaten archivieren

Vor der Archivierung von Studien sollten die entsprechenden Daten für die Versorgung des Patienten wie oben beschrieben in das Klinische Modul übertragen werden.

6.3.2.8 Daten sperren, anonymisieren oder löschen

Bei einem Widerruf der Einwilligung eines Patienten ist zu überprüfen, ob dieser Widerruf sowohl für die Daten des Studienmoduls als auch für die Daten des Klinischen Moduls gilt und die entsprechenden Daten in den Modulen gelöscht bzw. anonymisiert werden. Des Weiteren ist eine Löschung der identifizierenden Daten im Identitätsmanagement erforderlich.

Wenn Studiendaten im Rahmen einer klinischen Prüfung nach den Vorschriften des AMG verwaltet werden, so ist bei einem Widerrufen der Einwilligung zu beachten, dass bestimmte Daten aus Sicherheitsgründen nicht gelöscht oder anonymisiert werden dürfen (vgl. Kap. 4.3.1).

6.3.2.9 Machbarkeit einer Studie prüfen und Rekrutierung unterstützen

Für das Prüfen der Machbarkeit einer Studie und mehr noch die Rekrutierungsunterstützung wird vor allem die Klinische Datenbank wichtige Daten liefern können. Zum einen wird sie mehr Datensätze als eine Studiendatenbank beinhalten, wenn in einem Forschungsverbund über einen längeren Zeitraum mehr Patienten behandelt wurden als aktuell in eine der laufenden Studien eingeschlossen sind. Wichtiger aber ist noch, dass die aktuell in Studien eingeschlossenen Patienten nicht für eine weitere Rekrutierung zur Verfügung stehen.

Die Überprüfung der Machbarkeit einer Studie anhand anonymer Fallzahlen aus der Vergangenheit kann grundsätzlich auch unabhängig von einer vorherigen Einwilligung durchgeführt werden. Problematisch wäre jedoch eine anonyme Zusammenführung von Klinischer und Studiendatenbank, die zu doppelten Datensätzen führt. Bei Vorliegen einer entsprechenden Einwilligung der Probanden und einer Genehmigung durch den Ausschuss Daten-

schutz ist grundsätzlich auch eine pseudonyme Zusammenführung der Daten zur Abschätzung der Machbarkeit neuer Vorhaben oder auch zur Rekrutierungsunterstützung möglich. Letzteres jedoch nur bei Vorliegen eines sinnvollen Nutzungsszenarios für die Daten einer Studiendatenbank zu Rekrutierungszwecken. Eine mögliche Kontaktierung des Patienten kann wie im Kapitel 6.3 beschrieben erfolgen.

6.3.3 Nutzer, Rollen und Rechte

Für die Patientenliste ist analog zur Beschreibung im Studienmodul ein Administrator, ggf. unterstützt durch eine Dokumentationskraft, vorzusehen. Für die Administration der Studiendatenbank und der Klinischen Datenbank werden separate Administratoren benötigt, die Zugriff auf die jeweils dort gespeicherten MDAT, nicht jedoch die IDAT in der Patientenliste haben. Das Administrationspersonal von ID-Management und Studiendatenbank/Klinischer Datenbank sollte unter getrennter Verantwortung stehen (organisatorische Gewaltenteilung).

Studienärzte und Dokumentationskräfte erhalten Zugang auf die von ihnen betreuten Patienten im Studienmodul. Behandelnde Ärzte wiederum erhalten Zugriff auf die von ihnen betreuten Patienten im Klinischen Modul. Eine Patientenselbsteingabe kann in das Studienmodul oder Klinische Modul erfolgen, wenn sichergestellt ist, dass ein Patient nur Zugriff auf den jeweils eigenen Datensatz erhält. Bei Personeneinheit von Studien- und behandelndem Arzt ergeben sich die Berechtigungen aus der Vereinigungsmenge der jeweiligen Rollen.

Die Umsetzung der Rollen Monitor und Sponsor erfolgt analog zur Beschreibung des Studienmoduls, wobei Quelldaten, die im Rahmen des Monitorings geprüft werden, auch im Klinischen Modul liegen können. Wenn sich aufgrund des Monitorings Änderungsbedarf an Daten des Studienmoduls ergeben, sollen diese Änderungen im Klinischen Modul nachvollzogen werden.

6.3.4 Verantwortlichkeiten

Die Verantwortlichkeiten innerhalb des Studienmoduls und Klinischen Moduls werden in den Kapiteln 5.2 und 5.1 beschrieben. Die Gesamtverantwortung liegt beim Forschungsverbund. Mit der Führung der Patientenliste sowie der Studiendatenbank /Klinischen Datenbank werden voneinander unabhängige Einheiten des Verbunds beauftragt. Analog zum Studienmodul kann die Gesamtverantwortung mit der Führung der Patientenliste auch an einen zentralen Datentreuhänder übergeben werden. Datenschutzrechtlich sensible Fragen (z.B. Depseudonymisierung) sollten in einem zentralen Gremium (Ausschuss Datenschutz) entschieden werden. Bei einer multizentrischen Erhebung in ein zentrales System müssen Verantwortliche an den beteiligten

Standorten festgelegt werden. Bei Studien gemäß AMG oder MPG sind hier geltende zusätzliche Anforderungen sowie die übergeordnete Verantwortung des Sponsors zu berücksichtigen.

6.4 Kombinierter Einsatz von Studien- und Forschungsmodul

In Kapitel 5.2 wird das technische und organisatorische Konzept für eine Infrastruktur beschrieben, die für die Durchführung einzelner Forschungsprojekte wie z.B. klinischer Studien gemäß der Vorgaben des AMG oder MPG geeignet ist. Dabei wurde lediglich angedeutet, dass auch eine weitere Nutzung der Daten nach Abschluss der Studien für bestimmte Fragestellungen notwendig sein könnte. Ebenfalls bereits beschrieben (Kap. 5.3) sind Aufbau und Rahmenbedingungen langfristig angelegter Forschungsdatenbanken, wie sie typischerweise für epidemiologische Fragestellungen, zur Generierung neuer Forschungshypothesen oder für die Rekrutierungsunterstützung genutzt werden. Dabei wurde jedoch bisher nicht näher beschrieben, wie eine übergreifende Infrastruktur in datenschutzgerechter Weise aufgebaut werden kann, die sowohl die Durchführung einzelner Studien und Forschungsprojekte wie auch die Zusammenführung der Daten nach Abschluss der Studien in einer übergeordnet und langfristig angelegten Forschungsdatenbank unterstützt. Der Schwerpunkt der folgenden Ausführungen liegt somit auf der Verzahnung von Studien- und Forschungsmodul, wie sie jeweils einzeln bereits in Kapitel 5 beschrieben wurden.

6.4.1 Zweck und Anwendungsbereich

Die vom Bundesministerium für Bildung und Forschung (BMBF) seit 1999 geförderten Kompetenznetze in der Medizin hatten und haben neben der Vernetzung von Forschung und Versorgung insbesondere auch die Einbettung einzelner Forschungsprojekte in übergeordnete Infrastrukturen zum Ziel. So sollten nicht nur die Aufwände für die Umsetzung einzelner Studien verringert, sondern aus den gesammelten Daten auch ein größerer Nutzen gezogen werden können. Dieses auch im Interesse der Patienten liegende Ziel führte über die Vermittlung und Unterstützung in der TMF zur Ausarbeitung einer generischen Konzeption, die dann als Modell B in den generischen Datenschutzkonzepten der TMF 2003 bekannt wurde [1]. Der Aspekt der Unterstützung einzelner klinischer Studien nach den engen Vorgaben des Arzneimittelrechts fand damals noch keine ausführliche Berücksichtigung, da gerade die wissenschaftlich motivierten Arzneimittelstudien noch von dem Regulierungskorsett des AMG ausgenommen waren. Dies hat sich mit der 12. Novelle des AMG 2004 grundlegend geändert. Die vorliegende Neukonzeption einer übergreifenden Forschungsinfrastruktur profitiert somit einerseits von den Erfahrungen einiger Kompetenznetze, die im Gefolge der 12. AMG-Novelle bereits Erfahrungen mit dem Aufbau und der Anpassung ihrer Studieninfra-

strukturen gemäß der Vorgaben des AMG gesammelt haben. Andererseits werden hier Erfahrungen in neuer und generischer Form zusammengefasst, die neuen Forschungsverbünden eine deutlich schnellere und ökonomischere Konzeption, Abstimmung und Umsetzung einer zukunftssicheren und langfristig angelegten Studieninfrastruktur erlaubt. Auch wenn die Vorgaben des AMG für die IT-Unterstützung und Durchführung einzelner Studien berücksichtigt werden, ist die vorgeschlagene Infrastruktur nicht auf diese Form der Forschung festgelegt.

6.4.2 Prozesse und Anwendungsfälle

6.4.2.1 Patienten in das Studienmodul aufnehmen

Der Prozess der Aufnahme von Patienten in eine Studie wurde bereits in Kapitel 5.2 beschrieben. Für die spätere Zusammenführung und Nutzung der Daten aus verschiedenen Studien oder Forschungsprojekten ist allerdings entscheidend, dass die Einwilligungserklärungen zentral hinterlegt oder zumindest ausgewertet werden, da gerade die spätere Nutzung der Daten nur in Übereinstimmung mit einer idealerweise abgestuft formulierten Einwilligungserklärung [5, S. 97] geschehen darf. Zudem ist ein zentrales ID-Management (vgl. Kapitel 6.1) notwendig, welches zumindest die personenbezogene Zusammenführung der Pseudonyme einzelner Studien (Subject Identification Codes, SIC) anhand einer übergeordneten ID (PID_S) erlaubt. Die Patientenliste kann dabei für jede Studie separate IDs (SICs) verwalten und anhand eines einheitlichen Pseudonyms für das Studienmodul (PID_S) einander zuordnen, oder es wird von der Patientenliste für jedes Studien- oder Forschungsprojekt immer das gleiche Pseudonym (PID_S) herausgegeben. Wenn je Studie unabhängige SICs verwendet werden, kann die Patientenliste diese entweder von anderen Softwarekomponenten entgegennehmen und dauerhaft zusammen mit dem PID_S verwalten, oder die Patientenliste erzeugt diese IDs selber und gibt sie auf Anfrage heraus. Diese Prozesse sind bereits ausführlich in Kapitel 5.2 zum Studienmodul als Variante mit zentraler Patientenliste sowie in Kapitel 6.1 beschrieben.

Eine deutliche Vereinfachung und Beschleunigung der Rekrutierung lässt sich erreichen, wenn auf die Daten früherer Forschungsprojekte in einer Forschungsdatenbank in Übereinstimmung mit den entsprechend formulierten Einwilligungserklärungen zurückgegriffen werden kann. Auch hierfür ist ein zentraler Zugriff nicht nur auf die für die Ein- und Ausschlusskriterien relevanten Daten, sondern auch auf die durch die Einwilligungserklärungen festgelegten Nutzungsmöglichkeiten entscheidend. Diese sollten hierfür auch im Forschungsmodul mit hinterlegt sein.

6.4.2.2 Studiendaten erheben, auswerten und archivieren

Die Erhebung, Auswertung und Archivierung der Daten innerhalb einer klinischen Studie oder eines einzelnen Forschungsprojekts richtet sich weitestgehend nach den Anforderungen des konkreten Forschungsvorhabens und ist in Kapitel 5 ausführlicher beschrieben. Die nach Abschluss des Einzelprojekts stattfindende Überführung der Daten in eine übergeordnete Forschungsdatenbank hat auf diese Prozesse keinen direkten Einfluss. Dies gilt auch für Aufgaben wie z.B. das Management unerwarteter Ereignisse samt den damit einhergehenden gesetzlichen Meldeverpflichtungen.

6.4.2.3 Datenqualität im Studienmodul sichern

Ergänzend zu üblichen Qualitätssicherungsverfahren in einzelnen Studien (vgl. Kap. 5.2.2.5) kann bei Kombination mit einem Forschungsmodul auch ein Abgleich von Daten aus verschiedenen Studien zu einem Patienten durchgeführt werden. Eine solche Datenüberprüfung kann durch verschiedene unstimmige Daten ausgelöst werden (vgl. Kap. 6.3.2.3).

Bei der Übermittlung von Kontextdaten aus der Forschungsdatenbank an das zentrale Datenmanagement im Studienmodul zum Zwecke der Qualitätssicherung wird nur eine einstufige Depseudonymisierung benötigt. Diese erste Stufe der Depseudonymisierung sollte als reine Maschinenfunktion angelegt werden, die nur von besonders autorisierten Mitarbeitern des Datenmanagements im Studienmodul angestoßen werden kann. Dieser Anwendungsfall setzt die Definition eines studienübergreifenden Kern- oder Basisdatensatzes voraus, der in jeder Studie erhoben und somit zum Abgleich genutzt werden kann. Die Einwilligung der betroffenen Patienten für den konkret auf diesen Kerndatensatz bezogenen Abgleich mit bereits früher erhobenen Daten muss vorliegen.

Ausgangspunkt des Prozesses ist das zentrale Datenmanagement im Studienmodul, welches für alle Probanden, die bereits an einer vorhergehenden Studie teilgenommen haben, die PID_S sammelt und an den Pseudonymisierungsdienst schickt, der diese symmetrisch in PSN umschlüsselt und an die Forschungsdatenbank weiterreicht. Dort werden die zu den PSN zugehörigen Kerndatensätze herausgesucht und wiederum an den Pseudonymisierungsdienst geschickt, der diese nach symmetrischer Umschlüsselung der PSN in PID_S an das Datenmanagement im Studienmodul ausliefert. Dort können dann automatisierte Auswerteroutinen den Abgleich vornehmen und Auffälligkeiten in den Daten für eine weitere Überprüfung kennzeichnen.

Alternativ kann dieser Prozess auch über einen SIC aus der Studiendatenbank gesteuert werden. Hierfür ist die Einschaltung der Patientenliste über ein Ticket-Handling nötig, so dass der PID_S gegenüber der Studienzentrale nicht offenbart werden muss und andererseits die Patientenliste keinen Zugriff auf medizinische Daten erhält.

6.4.2.4 Datenqualität im Forschungsmodul sichern

Es kann die Notwendigkeit bestehen, medizinische Daten, die sich schon in der Forschungsdatenbank befinden und im Regelfall bereits einen aufwändigen Qualitätssicherungsprozess durchlaufen haben, trotzdem noch einmal zu ändern oder zu korrigieren. Im Zusammenhang mit einem Studienmodul wird dies z.B. dann notwendig, wenn im Rahmen der aktuellen Studie festgestellt wird, dass bereits in einem früheren Forschungsprojekt erhobene Basisdaten eines Patienten aktualisiert werden müssen (vgl. Kap. 5.2.2.5 zur Qualitätssicherung). Ein anderer Fall liegt vor, wenn allein in den Daten der Forschungsdatenbank Inkonsistenzen oder Fehler entdeckt werden, die jedoch nur mit Rückgriff auf die Quelldaten behoben werden können.

Wenn der Aktualisierungsbedarf im zentralen Datenmanagement des Studienmoduls entdeckt wird, übermittelt dieses den aktualisierten Datensatz mit dem zugehörigen PID_S an den Pseudonymisierungsdienst im ID-Management, der die Umschlüsselung des PID_S in ein PSN übernimmt und die Daten dann an die Forschungsdatenbank weiterleitet. Hier wird der Datensatz anhand des Pseudonyms PSN selektiert und geändert bzw. überschrieben.

Werden hingegen Fehler oder unplausible Daten im Forschungsmodul selbst festgestellt und ist ein Rückgriff auf die Quelldaten notwendig, ist ein anderes Vorgehen einzuhalten. Zunächst wird zu dem betreffenden Datensatz das Pseudonym PSN ermittelt und zusammen mit einer Fehlerbeschreibung über den Pseudonymisierungsdienst an die für die identifizierenden Daten zuständige Stelle im ID-Management geschickt. Diese nimmt daraufhin Kontakt mit dem aktuell behandelnden Arzt auf und leitet die Anfrage mit der Fehlerbeschreibung weiter. Nach Beantwortung der Anfrage wird ein ggf. korrigierter Datensatz vom behandelnden Arzt an die zentrale Stelle im ID-Management und von dort mit dem PID_S über den Pseudonymisierungsdienst an das Datenmanagement im Forschungsmodul weitergeleitet. Dort kann anhand des PSN eine Korrektur des betreffenden Datensatzes vorgenommen werden.

Aus Sicht des Datenschutzes kann es dem Betreiber der Forschungsdatenbank überlassen werden, ob er die Änderung in Form einer Versionierung oder mit Hilfe eines Audit Trails nachvollziehbar macht. Im Sinne einer hohen Datenqualität sind aber Funktionen, die eine Nachvollziehbarkeit aller Änderungen gewährleisten, auf jeden Fall zu empfehlen.

6.4.2.5 Daten vom Studienmodul in das Forschungsmodul übermitteln

Auch wenn in die hier konzeptuell beschriebene Forschungsdatenbank Daten aus ganz unterschiedlichen Quellen eingespeist werden können, steht in der folgenden Beschreibung die Übernahme von Daten aus klinischen Studien oder anderen, einzeln über ein Studienmodul abgewickelten Forschungsprojekten im Vordergrund. Dabei werden die Daten üblicherweise nach Abschluss

eines Forschungsprojekts oder zumindest nach Abschluss der Qualitätssicherung in die langfristig angelegte Forschungsdatenbank transferiert. Für diesen Prozess sind folgende Voraussetzungen entscheidend:

- Der Patient hat in die pseudonymisierte Speicherung nach Abschluss des konkreten Forschungsvorhabens eingewilligt.
- Der Patient hat in die Auswertung und Nutzung seiner Daten über die konkrete Fragestellung der aktuellen Studie hinaus eingewilligt.
- Der in die Forschungsdatenbank übertragene medizinische Datensatz erlaubt im Zusammenhang mit ggf. dort bereits gespeicherten Daten weiterhin eine pseudonyme Speicherung, d.h. dass auch bei Nutzung aller medizinischen Daten eines Patienten nach wie vor eine Reidentifizierung faktisch ausgeschlossen bleibt.
- Die Kennung der medizinischen Einrichtungen oder der individuellen Ärzte kann in den Forschungsdaten im Klartext oder ebenfalls pseudonymisiert gespeichert werden. Vor einer Speicherung solcher Daten im Klartext muss geklärt werden, dass hierdurch kein relevantes Reidentifizierungsrisiko entsteht.

Wenn diese Voraussetzungen gegeben sind, müssen die Datensätze mit einem langfristig sicheren Pseudonym versehen werden, welches an keiner weiteren Stelle im Zusammenhang mit den identifizierenden Daten der Patienten gespeichert wird. Dies wird durch eine zweite Stufe der Pseudonymisierung in einem Pseudonymisierungsdienst als Teil des Identitätsmanagements umgesetzt.

Als Ausgangspseudonym kann dabei nicht ein studienspezifisches Pseudonym (SIC) verwendet werden, da dann über die daraus gebildeten Pseudonyme der zweiten Stufe keine Zusammenführung von Daten aus mehreren Studien möglich wäre. Somit müssen für den Schritt der zweiten Pseudonymisierung die medizinischen Daten aus der aktuellen Studiendatenbank und die studienunabhängige ID im Studienmodul (PID_s) aus der Patientenliste im Zusammenhang verarbeitet werden, wobei eine physische Zusammenführung der Daten nach Möglichkeit vermieden werden sollte. Insbesondere dürfen die medizinischen Daten nicht an die Patientenliste geschickt werden, da dies die informationelle Gewaltenteilung in Bezug auf identifizierende und medizinische Daten durchbrechen würde. Die folgenden beiden Lösungen werden vorgeschlagen:

1. Die Studiendatenbank enthält bereits den PID_s als übergeordnete ID oder kann diesen problemlos und in Vereinbarkeit mit dem Datenschutzkonzept bei der Patientenliste mit Hilfe eines SIC abfragen. Dann werden die medizinischen Daten mit dem öffentlichen Schlüssel der Forschungsdatenbank asymmetrisch verschlüsselt und zusammen mit dem PID_s an den Pseudonymisierungsdienst geschickt.
2. In der Studiendatenbank ist nur ein studienspezifischer SIC hinterlegt und die übergreifende ID aus der Patientenliste kann aufgrund der

Datenschutzregeln des Forschungsverbunds auch nicht abgefragt werden, bzw. sollte sie der Studienzentrale mit der aktuellen Studiendatenbank nicht zur Kenntnis gelangen. In diesem Falle übermittelt die Studiendatenbank an die Patientenliste nur den aktuellen SIC und erhält von dieser ein Ticket (TKT) als temporäre ID. Daraufhin schickt die Patientenliste den studienunabhängigen PID_S zusammen mit dem gerade erzeugten Ticket an den Pseudonymisierungsdienst. Gleichzeitig werden die medizinischen Daten (MDAT) aus der Studiendatenbank mit dem öffentlichen Schlüssel der Forschungsdatenbank asymmetrisch verschlüsselt und zusammen mit dem von der Patientenliste übermittelten Ticket im Klartext an den Pseudonymisierungsdienst geschickt. Dieser kann dann die verschlüsselten und damit nicht einsehbaren MDAT über das Ticket dem von der Patientenliste empfangenen PID_S zuordnen und gemeinsam verarbeiten. Das Ticketprinzip ist in Kapitel 6.1 in Abbildung 9 veranschaulicht. Der Datentransfer über den Pseudonymisierungsdienst ist zudem in Kapitel 6.1.3.4 beschrieben.

Die Pseudonymisierung selbst wird durch eine symmetrische Umschlüsselung des PID_S in das weitere Pseudonym PSN umgesetzt. Aus Sicherheitsgründen wird der hierfür genutzte Schlüssel unauslesbar auf einer SmartCard oder einer vergleichbar sicheren Umgebung gespeichert. Die medizinischen Daten werden nach dem Umschlüsselungsprozess in unveränderter Form zusammen mit dem PSN an die Forschungsdatenbank im Forschungsmodul übergeben. In der Forschungsdatenbank können die medizinischen Daten (MDAT) anhand des privaten Schlüssels entschlüsselt und mit dem PSN als Ordnungskriterium gespeichert werden.

6.4.2.6 Daten an Forscher weitergeben

Die bei Exporten zu berücksichtigenden Rahmenbedingungen und Prozesse sind bereits in Kapitel 5.3 für isoliert aufgebaute Forschungsdatenbanken beschrieben.

6.4.2.7 Machbarkeit einer Studie prüfen

Um die Machbarkeit einer Studie prüfen zu können, müssen Indizien dazu ausgewertet werden, wie viele den spezifizierten Ein- und Ausschlusskriterien entsprechende Patienten innerhalb einer bestimmten Zeitspanne zu erwarten sind und ggf. in die Teilnahme einer Studie einwilligen würden. Die Daten des hier beschriebenen Forschungsmoduls können hierfür herangezogen werden. Im Regelfall wird für eine solche Analyse kein vollständiger Export von Datensätzen benötigt, sondern es reicht die Herausgabe der Anzahl von Datensätzen innerhalb eines festgelegten Zeitraums, die den spezifizierten Kriterien entsprechen. Das nähere Verfahren ist im Kapitel 5.3 über das Forschungsmodul beschrieben.

6.4.2.8 Rekrutierung unterstützen

Patienten, die bereits an einem früheren Forschungsprojekt teilgenommen haben, können mit Hilfe des Forschungsmoduls auch effektiv für weitere Studien rekrutiert werden, sofern dies durch die Einwilligungserklärung abgedeckt ist. Dies kann insbesondere bei chronischen Erkrankungen von Interesse sein. Anders als bei der Überprüfung der Machbarkeit wird hierfür eine Depseudonymisierung der Datensätze ausgelöst werden müssen, die den gesuchten Ein- und Ausschlusskriterien entsprechen. Idealerweise sollten die hinterlegten Einwilligungserklärungen der ausgewählten Patienten eine direkte Ansprache aus dem Forschungsverbund heraus erlauben. Anderenfalls könnte auch eine Ansprache über die aktuell oder zuletzt behandelnde Einrichtung geregelt sein.

Der Prozess startet im zentralen Datenmanagement des Studienmoduls, wo die Ein- und Ausschlusskriterien zusammengetragen und mit den Metadaten zur Forschungsdatenbank abgeglichen werden. Die aus diesem Abgleich entstandene Abfrage wird vom Ausschuss Datenschutz geprüft und an das Datenmanagement des Forschungsmoduls weitergereicht. Dort werden die PSN der zu dieser Abfrage passenden Datensätze extrahiert und über den Pseudonymisierungsdienst, der eine Umschlüsselung der PSN in die PID_s vornimmt, an das zentrale Datenmanagement des Studienmoduls geschickt. Das zentrale Datenmanagement richtet nun an die für die Speicherung der Identitätsdaten verantwortliche Stelle im ID-Management (z.B. als Treuhänder) die Anfrage, die aktuellen Behandler und ggf. weitere Adressdaten zu den übermittelten PID_s herauszugeben. Die hierfür verantwortliche Stelle prüft die Anfrage und informiert nach Freigabe durch den Ausschuss Datenschutz die aktuellen Behandler über die für eine Rekrutierung in Frage kommenden Patienten. Für den Fall, dass Patienten nicht mehr über den aktuell verzeichneten Behandler angesprochen werden können, sollte eine Alternativlösung vorbereitet werden. In so einem Falle könnte ein Treuhänder beispielsweise die Patienten anhand der Adressdaten direkt ansprechen, siehe auch die Ausführungen zum Kontaktmanagement in den Kapiteln 6.1, 6.5.2.4 (Unterkapitel zu Kontaktdaten) und 6.6.6.

6.4.2.9 Auskunft geben

Wenn ein Patient Auskunft verlangt, welche Daten über ihn gespeichert sind, ist im Falle eines kombinierten Studien- und Forschungsmoduls nicht nur der Datensatz in einer ggf. aktuell laufenden Studie zu berücksichtigen, sondern darüber hinaus eventuell auch im Forschungsmodul gespeicherte Datensätze aus früheren Forschungsprojekten.

Der Patient wendet sich hierzu entweder an seinen behandelnden Arzt oder direkt an eine zentrale, hierfür benannte Stelle des Forschungsverbunds. Diese Stelle kann bei einem zentralen Datentreuhänder angesiedelt sein. Die vom

behandelnden Arzt oder Patienten direkt angefragte Stelle ermittelt den zum Patienten zugehörigen PID_S und übermittelt diesen via Pseudonymisierungsdienst an das Datenmanagement im Forschungsmodul. Zu dem hier ankommenden PSN werden alle zugehörigen Datensätze ermittelt und über den Pseudonymisierungsdienst an die anfragende Stelle zurückgeschickt. Diese reicht die Daten entweder an den Patienten direkt oder an den behandelnden Arzt weiter.

6.4.2.10 Daten im Forschungsmodul umpseudonymisieren

Der Austausch der Pseudonyme einer Forschungsdatenbank kann aus unterschiedlichen Gründen notwendig werden: Z.B. bei Verlust oder Kompromittierung einer zur Pseudonymisierung genutzten SmartCard oder bei drohender Kompromittierung des verwendeten Verschlüsselungsalgorithmus. Liegt die Notwendigkeit eines Austausches der Pseudonyme vor, so muss dieser durch das Identitätsmanagement durchgeführt werden. Hierbei muss durch geeignete Verfahren sichergestellt werden, dass die neuen Pseudonyme dem richtigen Patienten bzw. Datensatz zugewiesen werden.

Das Identitätsmanagement im Forschungsverbund teilt dem Forschungsmodul mit, dass eine Umpseudonymisierung nötig ist. Das Forschungsmodul ermittelt dann alle betroffenen Pseudonyme PSN, ggf. auch aus mehreren Forschungsdatenbanken, und schickt diese an den Pseudonymisierungsdienst im Identitätsmanagement. Dieser transformiert die PSN zunächst mit Hilfe der bisherigen Smartcard zurück in den studienübergreifenden PID_S und im Anschluss mit Einsatz der neuen Smartcard in ein neues Pseudonym. Für diesen Prozess muss das Forschungsmodul die neuen PSNs eindeutig den Datensätzen korrekt zuordnen können, die bisher mit den alten PSNs markiert waren.

6.4.2.11 Daten sperren, anonymisieren oder löschen

Wenn keine Notwendigkeit für die Speicherung pseudonymisierter Daten im Forschungsmodul besteht, sind diese zu löschen oder zu anonymisieren. Wenn eine Anonymisierung möglich ist, wird dieser im Sinne weiterer wissenschaftlicher Verwertung der Daten vor dem Löschen der Vorzug zu geben sein. Auslöser der Anonymisierung können der Wunsch des Patienten, dessen Versterben oder das Verstreichen einer spezifizierten Frist seit der letzten Studienteilnahme sein. In letzterem Falle ist die Frist so zu bemessen, dass nach Ablauf der Frist eine erneute Teilnahme des Patienten an einer Studie im gleichen Forschungsverbund und damit die Erhebung von Follow-up-Daten so unwahrscheinlich ist, dass die weitere Speicherung pseudonymisierter Daten nicht mehr gerechtfertigt erscheint. Die hierfür relevante Frist ist im Datenschutzkonzept des Forschungsverbunds festzulegen und der Patient darüber zu informieren.

Die Anonymisierung wird vom ID-Management des Forschungsverbunds aus gesteuert, welches entweder bei Verstreichen der festgelegten Frist ohne Follow-ups diese selbst initiiert oder von einer behandelnden Einrichtung oder aus dem Umfeld des Patienten über einen Widerruf oder das Versterben des Patienten informiert wird.

In der Patientenliste des ID-Managements wird dann das studienübergreifende Pseudonym PID_s ermittelt und mit der Anonymisierungsanfrage über den Pseudonymisierungsdienst an das Forschungsmodul übermittelt. Der Pseudonymisierungsdienst ersetzt in dieser Anfrage den PID_s durch das symmetrisch verschlüsselte Pseudonym PSN und schickt die Anfrage an das Forschungsmodul weiter.

Im Forschungsmodul muss in jedem Fall sichergestellt sein, dass die Daten in allen beteiligten Forschungsdatenbanken anonymisiert werden. Es ist, z.B. durch die Wahl eines geeigneten Präfixes oder Suffixes, darauf zu achten, dass die anonymen IDs nicht mit pseudonymisierten IDs zu verwechseln sind. Wenn nur wenige Datensätze mit anonymen IDs versehen sind, muss ggf. zur Verhinderung einer Reidentifizierung auf eine eindeutige Markierung anonymer IDs verzichtet werden. Die Anonymisierung kann im Einzelfall und nach Abschätzung des Reidentifizierungsrisikos auch erfordern, dass einzelne charakteristische Merkmale des Falls gelöscht oder vergröbert werden.

Nach der Anonymisierung der Daten im Forschungsmodul muss die Löschung der identifizierenden Daten in der Patientenliste im Identitätsmanagement veranlasst werden. Ggf. sind auch Patientenlisten in den behandelnden Einrichtungen zu löschen.

Wenn das Löschen von Datensätzen im Forschungsmodul erforderlich ist, wird auch dieser Prozess von der Patientenliste im ID-Management gesteuert. Hierzu wird ebenfalls der studienübergreifende PID_s ermittelt und über den Pseudonymisierungsdienst mit der Löschungsaufforderung an das Forschungsmodul geschickt. Im Forschungsmodul werden dann in allen beteiligten Forschungsdatenbanken alle Datensätze mit dem zum PID_s korrespondierenden PSN gelöscht. Im Anschluss sind die IDAT in der Patientenliste und ggf. in den beteiligten Zentren lokal gespeicherte Zuordnungslisten zu löschen.

6.4.3 Nutzer, Rollen und Rechte

Neben der Patientenliste wird für das ID-Management bei Verknüpfung eines Studienmoduls mit einem Forschungsmodul auch ein Pseudonymisierungsdienst benötigt. Somit muss für das ID-Management nicht nur das Personal für die Administration der Patientenliste (s. Kap. 5.2.4) betrachtet werden sondern zusätzlich auch jenes für den Pseudonymisierungsdienst. Die Administration der beiden Dienste im ID-Management sollte unter getrennter Verantwortung stehen, so dass auch getrenntes Personal benötigt wird. Für den

Pseudonymisierungsdienst sollte eine administrative Kraft vorgesehen werden, die vollen Zugriff auf die Software, Einstellungen und das Handling der Smartcards hat. Das Personal im ID-Management wird, von administrativen Ausnahmen abgesehen, vor allem auf Anweisungen des Ausschusses Datenschutz tätig und wird entsprechend den Vorgaben die einfache oder vollständige Depseudonymisierung unterstützen.

Für das Studienmodul mit seinen Studiendatenbanken und behandelnden Einrichtungen sind im vorliegenden Verknüpfungsszenario die gleichen Nutzer und Rollen vorzusehen, wie sie bereits in Kapitel 5.2.4 beschrieben wurden. Insbesondere sind die Studienärzte und unterstützende Kräfte in den beteiligten Zentren, administratives Personal je Studiendatenbank, Monitore und ggf. Personal eines kommerziellen oder universitären Sponsors als Nutzer mit entsprechenden Rechten einzurichten.

Wie im Falle eines einfachen Studienmoduls muss auch geklärt werden, ob Patienten selbst in die Dokumentationsprozesse eingebunden werden. Entsprechend sind die Voraussetzungen hierfür zu schaffen (vgl. Kap. 5.2.4).

Der Zugriff auf eine oder mehrere Datenbanken im Forschungsmodul kann durch den Administrator sowie autorisierte Forscher erfolgen. Der Administrator hat vollen Zugriff auf die Datenbanken und kann entsprechende Selektionen und Exporte veranlassen. Der Forscher kann seinem Antrag entsprechend bestimmte Teile der Forschungsdatenbank einsehen. Während der Administrator die PSN im Forschungsmodul sehen darf, bleiben diese dem Forscher verborgen. Im Falle mehrerer Datenbanken im Forschungsmodul kann es auch getrennte administrative Zuständigkeiten für die einzelnen Datenbanken geben. Dabei muss aber eine einheitliche Schnittstelle zum Pseudonymisierungsdienst des ID-Managements gewährleistet bleiben.

6.4.4 Verantwortlichkeiten

Die Gesamtverantwortung liegt beim Forschungsverbund, was durch eine passende Rechtsform auch einen rechtsverbindlichen Ausdruck bekommt. Der Forschungsverbund beauftragt unterschiedliche Stellen mit den Aufgaben des ID-Managements sowie der Führung des Studien- und Forschungsmoduls. Dabei ist die Aufsicht über die beiden Komponenten Patientenliste und Pseudonymisierungsdienst des ID-Managements an zwei separate und voneinander unabhängige Institutionen zu vergeben. Ebenfalls unabhängig voneinander sollte die Aufsicht über das Studien- und das Forschungsmodul organisiert werden. In begründeten Einzelfällen kann von dieser Form der vollständigen informationellen Gewaltenteilung abgewichen werden. Die hierfür relevanten Entscheidungskriterien werden in Kapitel 6.7 dargestellt und diskutiert.

Die Verantwortlichkeiten innerhalb des Studienmoduls sind grundsätzlich so wie in Kapitel 5.2.5 beschrieben zu regeln. Im vorliegenden Falle ist lediglich

die Einbettung in die beim Forschungsverbund liegende Gesamtverantwortung zu berücksichtigen. Dies gilt auch für die Regelung der Verantwortlichkeiten in den beteiligten Zentren bzw. den behandelnden Einrichtungen. Diese und die Verantwortlichkeiten in der Studienzentrale müssen im Falle gesetzlich geregelter Studien nach AMG oder MPG zudem auch den gesetzlichen Anforderungen genügen. Eine übergeordnete Verantwortlichkeit kommt dann dem Sponsor der Studie zu. Wenn Prozesse, wie z.B. die Archivierung der Daten, ausgelagert werden, ist eine klare Delegationsregelung erforderlich.

Die Einrichtung eines „Ausschusses Datenschutz" als zentrales Gremium für die Beratung und Entscheidung datenschutzrechtlich sensibler Fragen wird dringend empfohlen. Dieses Gremium ist zudem für die Vorgabe von Richtlinien und Policies im Umgang mit den Daten zuständig.

Die Patientenliste ist der sensibelste Teil des Identitäts-Managements und hat damit, wenn sie zentral geführt wird, einen hohen Schutzbedarf. Datenschutzrechtlich ist zu berücksichtigen, dass die IDAT, obwohl sie in der Patientenliste nicht mit medizinischen Daten (MDAT) kombiniert werden, den betroffenen Personenkreis ggf. als Patienten eines Forschungsnetzes mit einem umschriebenen Krankheitsspektrum ausweisen. Im Falle eines stigmatisierenden oder in anderer Hinsicht sensiblen Krankheitsbereichs ist daher eine auch in den Augen der betroffenen Patienten besonders vertrauenswürdige Stelle mit der Führung der Patientenliste zu beauftragen.

Der Zugriff auf die Daten der Forschungsdatenbank kann nur nach Bewilligung durch den Ausschuss Datenschutz gewährt werden. Dabei werden der Forschungsansatz und der dafür benötigte Datensatz geprüft. Das Gremium leitet die Bewilligung an den Administrator des Forschungsmoduls weiter, der die Daten entsprechend der genehmigten Anforderung des Wissenschaftlers selektiert und exportiert oder ggf. dem Forscher eine selektive Sicht auf die Forschungsdatenbank ermöglicht.

6.4.5 Aspekte der Realisierung

Zentral für die Verzahnung eines Studienmoduls mit einem Forschungsmodul ist ein auslagerbarer Pseudonymisierungsdienst, der eine sichere Trennung der einstufig und zweistufig pseudonymisierten Datenbestände ermöglicht. Dieser Pseudonymisierungsdienst darf als Komponente des ID-Managements lediglich Zugriff auf die Pseudonyme selbst und nicht etwa auf MDAT oder IDAT haben. Da die MDAT zwischen den Systemen „vor" und „hinter" dem Pseudonymisierungsdienst transferiert werden müssen, ist für diese ein asymmetrisches Verschlüsselungsverfahren zu nutzen, bei dem der Absender jeweils den öffentlichen Schlüssel des Empfängers zum Verschlüsseln verwendet. So kann sichergestellt werden, dass nur die empfangende Komponente mit ihrem privaten Schlüssel die MDAT entschlüsseln und lesen kann. Alternativ zu diesem Ansatz ist auch ein Ticketsystem möglich, bei dem die MDAT

gar nicht an den Pseudonymisierungsdienst geschickt werden, sondern mit Hilfe eines temporären Tickets direkt vom Sender zum Empfänger geschickt werden können. Beide alternativen Möglichkeiten sind in Abbildung 9 veranschaulicht.

Der erste Weg, der eine asymmetrische Verschlüsselung nutzt, wird bereits in der aktuellen Version des Pseudonymisierungsdienstes der TMF (PSD) unterstützt. Dieser Dienst besteht aus drei Softwarekomponenten, die die gesamte Kommunikation abbilden. Die erste Komponente wird im Sicherheitskontext der zentralen Datenbank des Studienmoduls installiert und nimmt dort unverschlüsselte MDAT und einen PID_S in einer vordefinierten XML-Struktur entgegen. Diese XML-Struktur besteht aus einem vordefinierten Header, der u.a. den Absender und den umzuschlüsselnden PID_S enthält. In dem nicht weiter vordefinierten und frei nutzbaren Body-Teil der XML-Struktur können die MDAT in unverschlüsselter Form hinterlegt werden. Diese Softwarekomponente des PSD sorgt für eine asymmetrische Verschlüsselung der MDAT und schickt dann die XML-Struktur an die zentrale Komponente des PSD, wo mit Hilfe eines auf einer Smartcard sicher gehaltenen Schlüssels eine symmetrische Umschlüsselung des PID_S in das langfristig nutzbare Pseudonym PSN erfolgt. Mit diesem PSN im Header der XML-Struktur werden die Daten dann an die dritte Komponente weiter geleitet, die im Sicherheitskontext des Forschungsmoduls installiert ist. Diese Komponente kann mit dem dort verfügbaren privaten Schlüssel des Forschungsmoduls die MDAT entschlüsseln und zusammen mit dem PSN der Datenbank des Forschungsmoduls zur Verfügung stellen.

Diese Lösung kann auch genutzt werden, um Daten verschiedenen Datenbanken im Forschungsmodul zur Verfügung zu stellen, wobei die Adressierung für den PSD transparent im Sinne von unsichtbar erfolgt, d.h. diese wird innerhalb der MDAT vorgenommen. So können z.B. Bilddaten in eine separate Bilddatenbank transferiert werden, während die klinischen Verlaufsdaten in einem anderen Datenbanksystem landen.

Zusätzlich verfügt der PSD über einen Finding-Manager, der die Möglichkeit bietet, einen im Forschungsmodul ermittelten relevanten Befund auf sicherem Wege zum behandelnden Arzt zu übermitteln. Für die Aufrechterhaltung einer langfristigen Sicherheit der Pseudonyme im Forschungsmodul verfügt der PSD zudem über die Funktion eines Recryptings aller PSN mit Hilfe der alten und einer neuen Smartcard mit neuem Schlüssel.

Eine Erweiterung des PSD ist dahingehend geplant, dass auch ein Ticketsystem zum direkten Transfer der MDAT zwischen den Komponenten in Studien- und Forschungsmodul unterstützt wird. Ebenso eine Erweiterung um die Funktion einer Vermittlung eines SIC in einen PID_S mit Hilfe einer Kommunikation mit der zentralen Patientenliste des ID-Managements.

6.5 Das Maximalmodell eines medizinischen Forschungsverbundes

6.5.1 Zweck und Anwendungsbereich

Dieses Kapitel beschreibt einen Forschungsverbund, in dem ein ganzes Spektrum medizinischer Forschung zu einem bestimmten Krankheitsbild oder zu einer Gruppe zusammengehöriger Krankheitsbilder – von der molekulargenetischen über die klinische (beobachtende und interventionelle) bis zur epidemiologischen Forschung – in Kooperation organisiert wird. Weiterhin wird angenommen, dass im Forschungsverbund die Notwendigkeit zu langfristiger Aufbewahrung von Daten und Proben besteht und es größere Patienten- und Probandenzahlen gibt, es sich also nicht z.B. um eine seltene Erkrankung handelt (s. hierzu auch unter „Verhältnismäßigkeit", Kap. 6.7).

Ein medizinischer Forschungsverbund im Maximalmodell benötigt jedes der Module

- Klinisches Modul,
- Studienmodul,
- Forschungsmodul und
- Biobankenmodul

sowie dazu als zentrale Infrastruktur u.a. die Komponenten

- Identitätsmanagement und
- Rechtemanagement.

Diese Struktur ist in Abbildung 8 dargestellt.

In jedem der Module können unter Umständen auch mehrere Datenbanken gleicher Art angesiedelt sein; insbesondere im Studienmodul ist es oft nicht sinnvoll oder machbar, eine gemeinsame zentrale Datenbank für alle im Forschungsverbund durchgeführten klinischen Studien einzurichten.

Werden im Forschungsverbund auch in größerem Umfang Bilddaten erzeugt und aufbewahrt, sind als Komponenten eine oder auch mehrere eigenständige Bilddatenbanken vorzusehen; diese werden in einem eigenen Modul angesiedelt oder – bei entsprechend eingeschränktem Verwendungszweck – in Klinisches, Studien- oder Forschungsmodul integriert. Auch innerhalb eines solchen Moduls ist bei entsprechender Einschätzung des Reidentifizierungsrisikos unter Umständen die Speicherung von Bildern in einer auch für Daten genutzten, bereits vorhandenen Datenbank möglich oder in einer getrennten Datenbank notwendig. Eine organisatorisch getrennte Speicherung von Bilddaten ist dann nötig, wenn diese – z.B. als Fotografien des Gesichts – die Person leicht erkennen lassen. Kriterien für die einzelnen Optionen werden im Kapitel zur Verhältnismäßigkeit aufgeführt.

6.5.2 Prozesse und Anwendungsfälle

Die für das Maximalmodell relevanten Prozesse umfassen die Gesamtheit der bei den einzelnen Modulen und bei deren Zusammenspiel behandelten Prozesse. Hier wird zunächst für das Maximalmodell der typische Weg eines Patienten durch die Module des Forschungsverbunds beschrieben, danach folgen einige für das Maximalmodell nötige Erweiterungen und Ergänzungen zu den bereits in früheren Kapiteln beschriebenen Prozessen.

6.5.2.1 Patienten aufnehmen

Patienten werden im Maximalmodell bevorzugt in das Klinische Modul aufgenommen und in dessen Rahmen behandelt; ihre Daten werden in einer Klinischen Datenbank gespeichert, die auch als Grundlage für Beobachtungsstudien dient, wie in Kapitel 5.1 beschrieben. Der Forschungsverbund führt auch klinische Studien im Sinne des AMG oder MPG durch; dieses wird in Kapitel 5.2 beschrieben. Für diese Studien können geeignete Patienten aus dem Klinischen Modul, unter Umständen auch aus dem Forschungsmodul oder der Biomaterialbank gewonnen werden. Die Gewinnung von Teilnehmern an neuen Studien ist beispielsweise in Kapitel 5.3.2.5 für eine Forschungsdatenbank beschrieben. Da es sich dabei um Interventionsstudien handelt, muss für die Teilnahme eine neue gesonderte Einwilligungserklärung eingeholt werden; dies ist mit vertretbarem Aufwand möglich, da der Betroffene ja ohnehin kontaktiert werden muss. Hierfür werden die ADAT (s. Kap. 6.5.2.4) benötigt.

Auch nach Aufnahme in das Studienmodul verbleibt ein Patient in der Regel weiterhin im Klinischen Modul; das Zusammenspiel der beiden Module ist in Kapitel 6.3 beschrieben. Patienten, die direkt (primär) als Studienteilnehmer gewonnen wurden, werden, in Abhängigkeit von den Erfordernissen des Forschungsverbunds und der Einwilligung, parallel dazu auch in das Klinische Modul aufgenommen.

Gesunde Probanden, die als Kontrollpersonen am Forschungsmodul teilnehmen, können auch direkt dort aufgenommen werden.

Proben des Patienten oder Probanden werden an das Biobankenmodul übergeben und dort wie in Kapitel 5.4 beschrieben behandelt. Spätestens wenn der Patient nach den in Kapitel 5.1 formulierten Kriterien nicht mehr im Klinischen Modul geführt werden soll oder darf und auch an keiner klinischen Studie des Forschungsverbunds mehr teilnimmt, werden die Daten in das Forschungsmodul überführt, das in Kapitel 5.3 beschrieben wurde; der Übergang vom Studien- in das Forschungsmodul wird in Kapitel 6.4 behandelt. Im Forschungsmodul werden die Daten – abhängig natürlich von der vorliegenden Einwilligung – ebenso wie die Proben und Analyseergebnisse im Biobankenmodul in der Regel langfristig aufbewahrt.

6.5.2.2 Erweiterte Prozessbeschreibungen

Prozesse, die im Maximalmodell mehrere oder alle Module betreffen, sind der Widerruf, der Todesfall und die Qualitätssicherung. Die nötigen Erweiterungen sind:

- **Widerruf**: Widerruft ein Patient oder Proband seine Teilnahme am gesamten Forschungsverbund, so muss – in Abhängigkeit von den Regelungen der Einwilligungserklärung – dafür gesorgt werden, dass seine Daten in allen Modulen des Verbundes gelöscht bzw. anonymisiert werden. Für die Anonymisierung bedeutet dies insbesondere, dass alle noch im Klinischen oder Studienmodul befindlichen Daten in das Forschungsmodul übertragen und die IDAT im Identitätsmanagement gelöscht werden. Hierbei ist in jedem Einzelfall darauf zu achten, dass durch diese Datenzusammenführung kein erhöhtes Reidentifizierungsrisiko entsteht.
- **Todesfall**: Im Todesfall werden, sobald die Datenerhebung zum Fall abgeschlossen ist, die Daten und Proben in allen Modulen des Forschungsverbundes anonymisiert wie im Fall „Widerruf" beschrieben. Anderweitige Vereinbarungen aus der Einwilligungserklärung sind zu berücksichtigen.
- **Qualitätssicherung**: Ergeben sich bei Qualitätssicherungsmaßnahmen in einem Modul Datenänderungen (in der Regel sind das Fehlerkorrekturen), so sind diese den anderen Modulen mitzuteilen, sofern sie dort relevant sind. Für die richtige Zuordnung der pseudonymen Daten ist die Mitwirkung des Identitätsmanagements notwendig (s.a. Kap. 6.8).

Ein Prozess, der im Maximalmodell neu auftritt, betrifft einen Patienten, der an einer klinischen Studie teilnimmt und sowohl in einer Klinischen als auch in einer Studiendatenbank des Forschungsverbunds geführt wird, wenn seine Daten in die Forschungsdatenbank übermittelt werden (*simultane Übermittlung an die FDB*). In dieser Situation ist der Prüfarzt der Studie im Sinne des Studienmoduls in aller Regel gleichzeitig behandelnder Arzt im Sinne des Klinischen Moduls. Hierzu ist nach Möglichkeit die Anwendungssoftware so zu gestalten, dass diese beiden Übermittlungsvorgänge durch eine einzige Aktion gemeinsam gestartet werden; entsprechende Anforderungen wurden in Kapitel 6.1.6.2 formuliert.

6.5.2.3 Bilddaten

Bilddaten können zu verschiedenen Zwecken in einem Forschungsverbund wichtig sein:

- Im Klinischen Modul oder Studienmodul werden Bilder zur Referenzdiagnostik benötigt; hier besteht in der Regel ein Behandlungszusammenhang im Sinne einer konsiliarischen Tätigkeit, da das Ergebnis der Referenzbefundung direkten Einfluss auf die Behandlung des Patienten hat.

- Zur Verbesserung der Versorgung – nicht des abgebildeten Patienten, sondern weiterer Patienten mit ähnlichem Krankheitsbild – dient die Bereitstellung von Bildern als Vergleichsmaterial für den diagnostischen Prozess.
- Ferner können Bilder zu Ausbildungszwecken als Anschauungsmaterial bereitgestellt werden.
- Und schließlich können Bilddaten wie alle anderen Daten zu Auswertungen im Forschungszusammenhang dienen.

Für die Erhebung und Bereitstellung von Bilddaten im Forschungsverbund gilt das allgemeine Ablaufmodell für medizinische Daten mit nur geringfügigen Modifikationen. Folgende Besonderheiten müssen beachtet werden:

Generell enthalten vor allem Schichtbilddaten Informationen, aus denen mit Hilfe moderner dreidimensionaler Rekonstruktionsverfahren morphologische Informationen über einen Patienten rekonstruiert werden können; so kann z.B. das Gesicht einer Person aus einer Computertomographie des Schädels erzeugt werden. Dies birgt die Gefahr, dass trotz Anonymisierung oder Pseudonymisierung und der Löschung der identifizierenden Daten aus dem DICOM-Header[35] die Möglichkeit besteht, solche Rekonstruktionen mit biometrischen Daten aus anderen Quellen abzugleichen und so den Patienten zu identifizieren. Eine solche Rekonstruktion ist aber – zumindest zurzeit noch – mit einem sehr hohen Aufwand verbunden; die Risiko-Einschätzung ähnelt also der für genetische Daten: Man kann mittelfristig nicht von einer wirksamen Anonymisierbarkeit ausgehen. Eine Speicherung und Verwendung der Daten, soweit sie für die Zwecke des Forschungsverbunds unverzichtbar sind und der Forschungsverbund hierfür verbindliche Regelungen getroffen hat, ist in pseudonymisierter Form mit der entsprechenden Einwilligung langfristig möglich.

Eine weitere Besonderheit betrifft das so genannte Einbrennen von Patienten identifizierenden Daten in das Bildmaterial selbst. Solche Daten finden sich vor allem auf gescannten Röntgenbildern oder auch in Datensätzen aus Ultraschallgeräten. Hier ist ein nachträgliches Löschen der Daten, z.B. durch eine (semi-)automatisierte „Schwärzung" der betroffenen Bereiche, erforderlich. Da dieser Vorgang zum Teil sehr kompliziert, in bestimmten Datenformaten sogar kaum zu lösen ist, muss bereits im Vorfeld einer Studie geklärt werden, welche Geräte für die Datenerhebung eingesetzt werden sollen, um ihre spezifischen Eigenschaften prüfen und entsprechende Löschprozeduren implementieren zu können. Alternativ muss das Einbrennen der Daten von vornherein verhindert werden.

Bilder, insbesondere Fotografien von Gesichtszügen oder besonderen persönlichen Merkmalen, auf denen der Patient leicht zu erkennen oder wieder zu

35 Für die technische Ausführung des Löschvorgangs sei auf das Datenschutzkonzept „TMI-Server" verwiesen. Dieses kann bei der Geschäftsstelle der TMF angefragt werden (www.tmf-ev.de/datenschutz).

erkennen ist und bei denen die Erkennbarkeit nicht zu entfernen ist, sollten in der Einwilligung explizit erwähnt und besonders restriktiv gehandhabt werden. Um die Reidentifizierung der zugehörigen klinischen Daten zu verhindern, ist hier insbesondere eine getrennte Speicherung und ein separates Pseudonymisierungsschema (pseudonymer PID_B) vorzusehen.

Die so aufbereiteten Bilder werden – nach einem evtl. nötigen zusätzlichen Qualitätssicherungsprozess – in eine Bild-Datenbank oder, wie in Kapitel 6.5.1 beschrieben, eine andere Datenbank des Forschungsverbunds übertragen.

6.5.2.4 Organisatorische Daten

Organisatorische Daten (OrgDAT) gehören zu einer höheren Abstraktionsstufe (bzw. einer niedrigeren Prozessschicht) des IT-Modells. Ihre Notwendigkeit und ihr Informationsgehalt ergeben sich aus den Anforderungen der Datenprozessierung im Netz, sie sind nicht von vornherein durch die fachlichen Anforderungen des Forschungsverbunds definiert, so wie etwa IDAT und MDAT. Daher sind sie sehr von der konkreten Implementation der Prozesse im Forschungsverbund abhängig; generisch können nur einige allgemeine Aussagen gemacht werden.

OrgDAT begleiten andere Datenarten (z.B. MDAT in verschiedenen Kontexten) als eine Art von Metadaten und erfüllen folgende Zwecke:

Zugriffsregelung

OrgDAT dienen z.T. der Zugriffsregelung, vor allem im Klinischen Modul, u.U. auch im Studienmodul. Dann ist ihr *logischer* Platz im Rechtemanagement, z.T. auch im Identitätsmanagement. So enthält die Patientenliste auch die Information, wer als behandelnder Arzt (ADAT) für einen Patienten beim Forschungsnetz erfasst ist und damit Zugriffsberechtigung auf die Daten eines Patienten im Klinischen Modul hat; im Studienmodul kann das analog geregelt werden, sofern dort nicht die Zugriffe ohnehin innerhalb der Studiensoftware zufrieden stellend abgesichert werden können. Diese ADAT sind daher zusammen mit den IDAT zu speichern, das heißt, die behandelnden Ärzte sind den Patienten zugeordnet. Die Arztdaten können auch aus einem (pseudonymen) Verweis auf eine separat geführte Arztdatenbank bestehen. Die Informationen zur Zugriffsberechtigung auf einzelne Patientendatensätze befinden sich also auf dem Server der Patientenliste. Zu den OrgDAT gehören auch Zugriffstickets, die aber nur temporär sind und daher nicht mit anderen Daten permanent gespeichert werden.

Kontaktdaten

Bei der Anmeldung eines Patienten bei der Patientenliste werden das Kennzeichen der meldenden Klinik und das Datum der Meldung übertragen und in

der Liste gespeichert (als Teil des ADAT-Satzes). Dies gilt auch dann, wenn einem Patienten bereits ein PID zugewiesen wurde und dieser einer neu meldenden Klinik übermittelt wird. Kennzeichen und Datum werden nicht als Historie geführt, sondern durch die jeweils aktuelle Meldung überschrieben. Die Daten werden benötigt, damit die Stelle, welche die Patientenliste führt, erkennen kann, welche Klinik oder welcher Arzt informiert werden muss, wenn ein Patient in einem der dafür vorgesehenen Anwendungsfälle depseudonymisiert wird.

Dokumentation des Patientenwillens

Zu MDAT (in welchem Modul auch immer) sowie zu Proben und daraus gewonnenen AnaDAT gehören OrgDAT mit den Informationen, was im Rahmen der Patienteneinwilligung mit den zugehörigen Nutzdaten oder Proben gemacht werden darf. Hierzu gehören auch Kontaktinformationen, also in der Regel ein Verweis auf den behandelnden Arzt (ADAT). Um dadurch nicht einen erleichterten Abgleich bei unbefugter Kenntnisnahme von IDAT und MDAT zu ermöglichen, sollen die ADAT allerdings nicht an mehreren unabhängigen Stellen mitgeführt werden; in der Regel ist die Patientenliste der geeignete Speicherort für die ADAT, während die direkten Angaben zur Regelung in der Patienteneinwilligung mit den MDAT bzw. AnaDAT zusammen gespeichert werden.

Prozessunterstützung

Hier fallen OrgDAT etwa als Auftragsbeschreibungen bei der Kommunikation (z.B. Befundanforderung und Rückmeldung) an: Auftraggeber und Adressat, Umfang des Auftrags, Datum, Fristen, Besonderheiten. OrgDAT benötigt man auch zur Definition des Status der zugehörigen Daten oder Proben (z.B. für Qualitätssicherung und Monitoring oder, bei Proben, als Hinweise auf Aliquots).

Qualitätssicherungsaspekte können es erforderlich machen, die Daten vor ihrer Überführung in die Forschungsdatenbank, d.h. vor der zweiten Stufe der Pseudonymisierung, zu prüfen (s. 6.8). In diesem Fall wird eine temporäre Datenbank TempDB eingerichtet. In dieser werden die Daten: PID, MDAT, OrgDAT, $LabID_{tr}$ zusammen mit dafür nötigen OrgDAT kurzzeitig gespeichert. Hier können sie in einem definierten kurzfristigen Zeitraum zur Qualitätssicherung genutzt werden. Bei der Pseudonymisierung und Übertragung in die Forschungsdatenbank werden die Daten in der TempDB gelöscht.

OrgDAT werden in größerem Umfang im Biobankenmodul benötigt. Hier sind sie Begleitdaten einer Probe, die an unterschiedlichen Stellen entstehen und verwendet werden. So erfasst z.B. die Proben gewinnende Stelle die Probenart und gegebenenfalls die Informationen zu Probenentnahme und Präanalytik.

In der Probenbank werden die Begleitdaten einer Probe mit weiteren Informationen wie z.B. den Umständen von Konservierung, Lagerung und Qualität gespeichert. Für weitere Details sei auf das Datenschutzkonzept für Biomaterialbanken [2] verwiesen.

Auch in „angehefteten Dokumenten" können OrgDAT enthalten sein, z.B. in eingescannten Formularen oder in Röntgenbildern. Hier ist auf den Gehalt solcher Dokumente an identifizierenden Daten zu achten. (Z.B. würde die eingescannte Patienteneinwilligung den Namen des Patienten enthalten. Oder die in einem Ultraschallbild enthaltene Gerätekennung identifiziert den Arzt.)

OrgDAT sollen, wie andere Daten auch, nur so lange gespeichert werden, wie sie für die definierten Zwecke benötigt werden. Temporäre OrgDAT wie Zugriffstickets werden direkt nach Benutzung ohne Spuren gelöscht; in den zur Zugriffsüberprüfung mitgeführten Protokollen werden sie nicht benötigt. „Permanente" OrgDAT sind solche, die zumindest eine Zeitlang benötigt werden; hier ist besonders auf mögliche Reidentifikationsrisiken zu achten. Das bedeutet insbesondere, dass die einzelnen Bestandteile der OrgDAT jeweils nur in einer einzigen Datenbank des Forschungsverbunds gehalten werden sollten.

6.5.2.5 Zusammenwirken der Module

Die Daten im Maximalmodell setzen sich zusammen aus den Daten der einzelnen Module. Auch die Datenflüsse wurden bereits weitgehend beschrieben: Das Zusammenspiel von Klinischem Modul und Studienmodul wurde in Kapitel 6.4 beschrieben, das Zusammenspiel von Studienmodul und Forschungsmodul in Kapitel 6.3. Weitgehend ähnlich dazu ist das Zusammenspiel von Klinischem Modul und Forschungsmodul, wobei die unterschiedliche Handhabung von PID_K und PID_S (bzw. SIC) zu beachten ist.

Das Zusammenspiel von Biobankenmodul und anderen Modulen wurde bereits im TMF-Datenschutzkonzept für Biomaterialbanken beschrieben und in Kapitel 5.4 zusammengefasst.

Der Gebrauch von Pseudonymen in den verschiedenen Modulen wurde zum großen Teil in Kapitel 6.1, insbesondere Unterkapitel 6.1.1 beschrieben.

6.5.3 Nutzer, Rollen und Rechte

Nutzer, Rollen und Rechte sind auch im Maximalmodell in der Regel einem einzelnen Modul zugewiesen. Übergreifende Rollen sind im IT-Bereich nicht vorgesehen und auch nicht nötig. Im organisatorischen Bereich ist vor allem die übergreifende Verantwortung für die verbundweit gültigen Richtlinien (Policies) und – soweit modulübergreifend sinnvoll – Verfahrensanweisungen

(SOPs) zu nennen; diese Rollen werden aber in der Regel nicht in der IT-Struktur direkt abgebildet.

Auch das Datenmanagement ist für die einzelnen Module, ja sogar Datenbanken, personell getrennt. Hat der Forschungsverbund zusätzlich einen zentral verantwortlichen Datenmanager oder CIO, ist darauf zu achten, dass dieser nicht Daten zur Kenntnis erhält, die ihn zur Umgehung der Pseudonymtrennung oder gar zu einer Reidentifizierung befähigen.

6.5.4 Verantwortlichkeiten

Modulübergreifende Verantwortlichkeiten betreffen

- die Definition zentraler Policies sowie
- die Genehmigung von Datenweitergaben an Forschungsprojekte.

Diese werden vom Ausschuss Datenschutz des Forschungsverbunds wahrgenommen. Ansonsten ist die Verantwortung für die einzelnen Module organisatorisch getrennt.

6.5.5 Aspekte der Realisierung

Die Anforderungen eines Forschungsverbunds im Maximalmodell an die IT-Infrastruktur sind so vielfältig, dass sie mit einem einzelnen zentral ausgerichteten EDC-System nicht zu realisieren sind. Die einzelnen Module sind in der Regel unabhängig voneinander mit geeigneten Softwaresystemen auszustatten, die aber die benötigten Kommunikationsbeziehungen und zentralen Dienste unterstützen müssen. Ein Beispiel ist das in den Kapiteln 6.5.2 und 6.1.2 b, c) beschriebene Zusammenwirken zwischen KDB, SDB und FDB.

Auch die Gewinnung, Aufbereitung, Verwaltung und Bereitstellung von Bildern erfordert eigene Software-Systeme. Enthalten die generierten Datensätze dabei im Bildmaterial selbst Daten, welche Patienten, Institutionen und Geräte identifizieren und die für die Aufnahme in die Forschungsdatenbank geschwärzt werden müssten, so muss mit den Herstellern dafür eine Änderung ihrer Software ausgehandelt werden. Für die Betrachtung und Nutzung der Bilder sollten geeignete Viewer in die Software der jeweiligen Datenbank integriert sein.

6.6 Organisatorische Regelungen

Ein Datenschutzkonzept muss immer technische und organisatorische Regelungen umfassen. Der Grundsatz „Verhindern ist besser als Verbieten“ spricht dafür, möglichst weitgehende technische Vorkehrungen zur Unterstützung des Datenschutzes zu implementieren. Aber technische Maßnahmen können

nur in einem definierten organisatorischen Rahmen ihre Wirksamkeit entfalten, der z.B. die Verantwortlichkeit klar regelt. Außerdem können technische Maßnahmen nicht alle Datenschutzanforderungen umsetzen, sondern werden immer viele Lücken lassen; hier müssen organisatorische Absicherungen, z.B. Verbote, ergänzend eingreifen.

Viele dieser organisatorischen Aspekte wurden in den bisherigen Kapiteln bereits beschrieben. In diesem Kapitel werden die nötigen Rahmenbedingungen und Regelungen eines Forschungsverbundes zusammengefasst und systematisch dargestellt.

6.6.1 Rechtsform – Forschungsverbund als juristische Person

Für eine rechtssichere Umsetzung der Regeln zu Datenschutz und Datensicherheit ist es unerlässlich, dass sich der Forschungsverbund den Status einer juristischen Person gibt, siehe auch Kapitel 4.2.3. In dieser Eigenschaft kann er für zentrale Dienste Aufträge vergeben und mit Nutzungsordnungen verbinden, welche die organisatorisch und datenschutzrechtlich relevanten Regelwerke darstellen. Die möglichen verschiedenen Rechtsformen wurden in dem Rechtsgutachten ausführlich analysiert, das die TMF im Rahmen des Biomaterialbanken-Projektes hat anfertigen lassen und das den Kern der zugehörigen Publikation [23] bildet; eine zusammengefasste Darstellung ist in [2] enthalten. Diese ist zu großen Teilen auch für medizinische Forschungsverbünde im Allgemeinen gültig und wird in entsprechend angepasster Umformulierung im Folgenden wiedergegeben.

Im akademischen Umfeld entstehen Forschungsprojekte üblicherweise durch die persönliche Initiative eines oder mehrerer Wissenschaftler. Die Trägerschaft ist dann aber in der Regel nicht an diese Person gebunden, sondern an die entsprechenden Universitäten und Kliniken. Diese beschäftigen das Personal für den Forschungsverbund und stellen Räumlichkeiten und Infrastruktur zur Verfügung. Die in diesen Einrichtungen vorhandene Infrastruktur ist einerseits ein Garant für die fachgerechte Durchführung eines Projekts, insbesondere die qualifizierte Betreuung von Daten- und Probensammlungen, andererseits besteht unter dem steigenden Kostendruck der Universitäten und Kliniken aber auch die Gefahr, dass das Forschungsvorhaben nicht weiter unterstützt wird, wenn die Leitung der Universität bzw. Klinik andere fachliche Schwerpunkte setzt oder der Initiator die Einrichtung wechselt. Auf Dauerhaftigkeit ausgerichtete Forschungsverbünde sind daher im Regelfall in einen privatrechtlichen Rahmen zu überführen und dort mittels einer geeigneten Rechtsträgerschaft zu verstetigen.

Grundsätzlich kommt jede denkbare Rechtsform für den Träger eines Forschungsverbundes in Frage. Typischerweise eignen sich in der Wissenschaft die Rechtsformen eingetragener Verein, GmbH und privatrechtliche Stiftung besonders gut für einen Forschungsverbund. Die TMF bietet mit ihrer Arbeits-

gruppe Netzwerkkoordination ein Austauschforum an, in dem dieses und weitere verwandte Themen diskutiert und Erfahrungen dazu weitergegeben werden können.

6.6.2 Allgemeine Regelungen

Satzung bzw. Gesellschaftervertrag legen die Grundlagen für die Organisation des Forschungsverbundes fest. Zu diesen Grundlagen gehören die Verantwortlichkeiten und die Ermächtigungsgrundlagen für Geschäftsordnungen. Innerhalb der Satzung oder des Gesellschaftervertrages sind die grundsätzlichen Zuständigkeiten festzulegen. Die detaillierte Ausgestaltung kann im Rahmen zusätzlicher Geschäftsordnungen erfolgen. Dabei sollte eine Unterscheidung zwischen allgemeiner Geschäftsführung und Geschäftsführung für spezielle Aufgabenfelder bezüglich Datenschutz und Forschung vorgenommen werden. Derartige Statuten zur Festlegung von Zuständigkeiten innerhalb des Forschungsverbundes sind in jedem Fall erforderlich, auch wenn er sich in öffentlich-rechtlicher Trägerschaft befindet. Ferner schließt der Träger des Forschungsverbundes Verträge bzw. Vereinbarungen mit Datenzulieferern, externen Teilnehmern und Forschungsinstitutionen sowie allen Dienstleistern, die für den Forschungsverbund tätig werden.

Wie der Verbleib des Vermögens eines Vereins muss auch der Verbleib der Daten des Forschungsverbunds in der Satzung oder im Gesellschaftervertrag geregelt sein. Für den in öffentlich-rechtlicher Hand befindlichen Forschungsverbund sollten schon zu Beginn Überlegungen angestellt werden, ob eine Übertragung der Daten an andere Institutionen oder eine Überführung in eine privatrechtliche Organisation für einen späteren Zeitpunkt vorgesehen werden soll, und entsprechende Regelungen in den Statuten getroffen werden. Hierfür können z.B. je nach Anwendungsfall Fachgesellschaften oder auch Patientenverbände in Frage kommen. Jeder Forschungsverbund sollte in seinem Regelwerk Bedingungen für seine Auflösung festlegen. Für den Fall einer Übertragung ist die Zustimmungspflicht durch die Ethikkommission, möglicherweise auch eine zuständige Fachgesellschaft zu regeln. Werden andere Regelungen gewählt, müssen diese entsprechend begründet sein.

6.6.3 Der Ausschuss Datenschutz

Die Satzung des Forschungsverbundes sieht als wichtiges Gremium neben dem Vorstand den Ausschuss Datenschutz vor, der die Regelung aller mit dem Datenaustausch und dem Datenzugang zusammenhängenden Fragen verantwortet. Die Satzung muss eine Besetzung dieses Gremiums vorsehen, die Interessenkonflikten entgegenwirkt. Sollte der Forschungsverbund als juristische Person auch über einen Datenschutzbeauftragten verfügen, sollte dieser möglichst auch Mitglied sein. Gleiches kann für Datenschutzbeauftragte der

im Forschungsverbund beteiligten Institutionen gelten. Der Ausschuss Datenschutz wird durch den Vorstand des Forschungsnetzes mit folgenden Aufgaben berufen:

- Er entscheidet mit Hilfe eines formalisierten Antragsprozesses über Art und Inhalt der Weitergabe medizinischer Daten oder wissenschaftlicher Proben an die Antrag stellenden Wissenschaftler. Mit der Bewilligung ist zu definieren
  - der auf die Forschungsaufgabe zugeschnittene Datensatz,
  - die anzuwendenden Selektionsfilter sowie
  - der Zugang zu pseudonymisierten oder anonymisierten Daten.
- Er entscheidet, ob die Benachrichtigung eines Patienten über die gewonnenen Erkenntnisse durch den zuletzt behandelnden Arzt zulässig ist. Bei besonders schwierigen Fragen kann der Ausschuss Datenschutz eine Ethikkommission zur Beratung hinzuziehen.
- Er verabschiedet die Regelwerke (Policies), die für jeden für Datenschutz und Datensicherheit relevanten Prozess zu formulieren sind, und ist verpflichtet, ihre Einhaltung im Forschungsnetz zu überprüfen und sie bei Bedarf fortzuschreiben.

Im Einzelnen sind Aufgaben des Ausschusses Datenschutz in den Kapiteln

- 4.2.3 (Verantwortlichkeiten),
- 5.2.4 (Studienmodul – Nutzer, Rollen und Rechte),
- 5.2.5 (Studienmodul – Verantwortlichkeiten),
- 5.3.2.10 (Forschungsmodul – Ergebnisse mitteilen),
- 5.3.5 (Forschungsmodul – Verantwortlichkeiten),
- 6.1.5 (ID-Management – Verantwortlichkeiten),
- 6.1.5.2 (ID-Management – Mehrere Patientenlisten an einem Standort?),
- 6.2.4 (Rechtemanagement – Verantwortlichkeiten),
- 6.3.2.9 (Kombinierter Einsatz von Studienmodul und Klinischem Modul – Machbarkeit einer Studie prüfen und Rekrutierung unterstützen),
- 6.3.4 (Kombinierter Einsatz von Studienmodul und Klinischem Modul – Verantwortlichkeiten),
- 6.4.2.8 (Kombinierter Einsatz von Studienmodul und Klinischem Modul – Rekrutierung unterstützen),
- 6.4.3 (Kombinierter Einsatz von Studien- und Forschungsmodul – Nutzer, Rollen und Rechte),
- 6.4.4 (Kombinierter Einsatz von Studien- und Forschungsmodul – Verantwortlichkeiten),
- 6.6.5.2 (Organisatorische Regelungen – Regeln für die Datenverwendung) sowie
- 6.7.4.2 (Kriterien der Verhältnismäßigkeit – Rechtemanagement)

beschrieben. Die Entscheidungen des Gremiums sollten nach dokumentierten Kriterien oder eine entsprechenden Leitlinie getroffen werden.

6.6.4 Informationelle Gewaltenteilung

Informationelle Gewaltenteilung bedeutet, dass Daten so auf verschiedene Datenspeicher mit wechselseitig nicht weisungsbefugter Administration aufgeteilt werden, dass die einzelnen Teile nicht zu einer unbefugten Reidentifikation von betroffenen Personen führen können. Beispiele hierfür sind das Identitätsmanagement, das unabhängig von den im Forschungsverbund vorhandenen Modulen betrieben wird, oder auch die Aufteilung eines Forschungsverbunds in unabhängige Module.

Erleichternd sei bemerkt, dass verschiedene Einrichtungen einer Universitätsklinik wechselseitig nicht weisungsbefugt sind, so dass etwa eine Ansiedlung zweier Datenbanken und Dienste am Klinikrechenzentrum und einem Medizininformatischen Institut in der Regel die Anforderungen an die informationelle Gewaltenteilung erfüllt. Je nach Beurteilung der Gefährdungslage eines Forschungsverbundes, siehe Kapitel 6.7 (Kriterien der Verhältnismäßigkeit), kann aber auch die Einschaltung eines externen Datentreuhänders als wirksamerer und deutlicherer Ansatz zur informationellen Gewaltenteilung angesehen werden; hierzu siehe auch Kapitel 4.2.5 (Elektronische Datentreuhänderschaft).

Einzelne Aspekte der informationellen Gewaltenteilung werden beschrieben in den Kapiteln

- 4.2.5 (Elektronische Datentreuhänderschaft),
- 4.6 (Grundprinzipien datenschutzgerechter Lösungen),
- 5.3.5 (Forschungsmodul – Verantwortlichkeiten),
- 5.4.3 (Biobankenmodul – Daten und Datenflüsse),
- 6.1.5 (ID-Management – Verantwortlichkeiten),
- 6.2 (Rechtemanagement – Einleitung),
- 6.2.4 (Rechtemanagement – Verantwortlichkeiten),
- 6.2.5 (Rechtemanagement – Aspekte der Realisierung),
- 6.4.4 (Kombinierter Einsatz von Studien- und Forschungsmodul – Verantwortlichkeiten),
- 6.7.3 (Kriterien der Verhältnismäßigkeit – Modellvarianten) sowie
- 6.7.4 (Kriterien der Verhältnismäßigkeit – Beispiele).

6.6.5 Regelwerke

Zur Konkretisierung der datenschutzrechtlichen Vorschriften, des Strafgesetzbuches, der Berufsordnung und der sonstigen berufsethischen Normen sind Regelwerke zu schaffen, auf die alle Beteiligten vertrauen können und an die das medizinisch behandelnde und forschende Personal in der Nutzung der Systeme rechtsverbindlich gebunden wird:

1. Für einen Patienten geschieht dies im Rahmen des Behandlungsvertrags mit den Ärzten oder der Klinik sowie durch die Aufklärung und eine informierte Einwilligung, Daten für den Forschungsverbund zur Verfü-

gung zu stellen. Gesunde Probanden sind analog aufzuklären und um eine Einwilligung zu bitten.
2. Für behandelnde Ärzte und klinisches Personal gelten in erster Linie die Regeln, die von den jeweiligen Kliniken unter der Verantwortung des leitenden Arztes vorgegeben sind.
3. Auch das forschende medizinische und nicht-medizinische Personal kann an die Regeln der jeweils verantwortlichen Klinik gebunden werden. Manche der Tätigkeiten, wie die Erhebung und Weiterleitung von Forschungsdaten, überschreiten die Grenzen der Klinik und müssen an Regelwerke gebunden sein, die für den gesamten Forschungsverbund verbindlich sind. Für die rechtliche Verbindlichkeit ist die Regelung der Verantwortlichkeiten durch eine geeignete Rechtsform, siehe Kapitel 6.6.1 oben, wesentliche Voraussetzung.
4. Für die zentralen Dienste – z.B. Führung der Datenbanken, Patientenliste, Qualitätssicherung und Pseudonymisierungsdienst – sind geeignete Nutzungsordnungen und SOPs mit den datenschutzrechtlich relevanten Regelwerken aufzustellen und Verträge zu schließen, welche alle Beteiligten rechtsverbindlich an die Regelwerke binden.

Insgesamt müssen die Regelwerke so gestaltet sein, dass sich aus ihnen die nötigen Rechtedefinitionen für ein wirksames Rechtemanagement (Kap. 6.2) ableiten lassen.

6.6.5.1 Verträge

Der Forschungsverbund als juristische Person schließt Verträge, um die Beteiligten an die Regelwerke zu binden:

1. mit den dokumentierenden Ärzten und ihren Mitarbeitern zur Festlegung der Anforderung an die Forschungsdaten und ihre Überlassung an den Forschungsverbund;
2. mit den Wissenschaftlern zu den Verfahren, die ihnen Zugang zu den Forschungsdaten verschaffen und sie an die regelgerechte Verwendung von Daten und biologischen Proben binden;
3. mit den zentralen Diensten und beteiligten Rechenzentren zur Regelung der Aufgaben und Pflichten, die mit dem Auftrag zur Datenverarbeitung verbunden sind. In den Verträgen soll auch die Unabhängigkeit von Datenbank-Administratoren vom forschenden Personal sichergestellt werden. Ebenso muss die wechselseitige Unabhängigkeit der verschiedenen Datenbank-Administratoren voneinander gewährleistet sein.

6.6.5.2 Regeln für die Datenverwendung

Der Wissenschaftler darf die zur Verfügung gestellten Daten ausschließlich im Rahmen der Zielsetzung seiner Arbeit und der durch das Forschungsnetz ausgesprochenen Genehmigung verwenden. Die Weitergabe der exportierten

Daten an Dritte ist generell untersagt. Für die wissenschaftliche Zusammenarbeit über die Grenzen des Forschungsnetzes hinaus sind getrennte und spezifische Regelungen mit dem Ausschuss Datenschutz des Forschungsnetzes herbeizuführen.

6.6.5.3 Sicherheitspolicy – Nutzungsordnungen

Als Regelwerke für die zentralen Dienste stellt der Forschungsverbund Nutzungsordnungen bereit, die das Sicherheitspotenzial der beschriebenen technischen Instrumente im organisatorischen Bereich verankern. Die Betreiber und die Nutzer werden über die notwendigen Maßnahmen und Abläufe informiert und zu einem planmäßigen, regelgerechten Handeln verpflichtet.

6.6.5.4 Zusammenstellung von Musterdokumenten

Ein Forschungsverbund benötigt insgesamt eine nicht unbeträchtliche Zahl von datenschutzrechtlich relevanten Regel- und Vertragswerken. Dazu gehören u.a. Policies, Verpflichtungserklärungen, SOPs und Service Level Agreements (SLA). Ein Überblick über die bei der TMF vorhandene Sammlung von einschlägigen Musterdokumenten ist im Anhang zu finden.[36]

6.6.6 Einwilligungsmanagement

Der sachgerechte Umgang mit den Patienteneinwilligungen erfordert dann einige technische Überlegungen, wenn die dort getroffenen Regelungen sich bei verschiedenen Patienten oder Probanden unterscheiden können (s. Kap. 4.2.2 „Datenschutzrechtliche Grundlagen" – „Grenzen von Einwilligungsszenarien"). Die Festlegungen, die mit der Einwilligung getroffen werden, müssen bei der jeweiligen Verwendung der Daten auf möglichst unkomplizierte Weise zur Verfügung stehen.

Zunächst wird die Einwilligung in Papierform vom aufnehmenden Arzt eingeholt und sollte in dieser Form dort auch aufbewahrt werden; bei großen und langzeitigen Forschungsverbünden kommt auch die Hinterlegung bei einem Datentreuhänder in Betracht. Falls es keine Variationsmöglichkeiten bei der Einwilligung gibt, wie es bei klinischen Studien oft der Fall ist, ist darüber hinaus kein Einwilligungsmanagement erforderlich.

Anders sieht es aber aus, wenn patientenindividuelle Festlegungen zu berücksichtigen sind. Bei jeder beabsichtigten Datenverwendung muss in jedem Einzelfall feststehen, was erlaubt ist. Daher muss ein Feld der OrgDAT dafür vorgesehen werden, detaillierte Informationen zur Einwilligung abzubilden, den Wunsch nach Wissen oder Nichtwissen und bei einer abgestuften Ein-

36 Anhang siehe unter http://www.tmf-ev.de/datenschutz-leitfaden

willigungsmöglichkeit etwa die gewählte Stufe (s. Kap. 4.4.2). Auch Ausschlüsse müssen dort festgehalten werden, was in den meisten Fällen ein Freitextfeld erforderlich macht. Diese Erweiterung der OrgDAT ist im Klinischen, Studien-, Forschungs- und Biobankenmodul relevant.

Zu beachten ist, dass das Mitführen dieser Daten in Einzelfällen ein erhöhtes Reidentifizierungsrisiko bedeuten kann; z.B. könnte, wenn in einer Forschungsdatenbank und in einer Probenbank eine einzigartige identische Einwilligungsregelung festgehalten wird, die Trennung der Pseudonyme unwirksam werden. Für eine solche Situation ist die Einrichtung eines zentralen Einwilligungsmanagements, etwa kombiniert mit dem Identitätsmanagement, zu empfehlen.

Hinweise zum Einwilligungsmanagement sind in den vorangegangenen Kapiteln

- 5.1.2.1 (Klinisches Modul – Patienten in das Klinische Modul aufnehmen),
- 5.1.2.10 (Klinisches Modul – Rekrutierung unterstützen),
- 5.2.2.1 (Studienmodul – Patienten aufklären und Einwilligung einholen),
- 5.3.2.4 (Forschungsmodul – Machbarkeit einer Studie prüfen),
- 6.3.2.2 (Kombinierter Einsatz von Studien- und Klinischem Modul – Patienten in Klinisches Modul oder Studienmodul aufnehmen),
- 6.4.2.1 (Kombinierter Einsatz von Studien- und Forschungsmodul – Patienten in das Studienmodul aufnehmen),
- 6.4.2.3 (Kombinierter Einsatz von Studien- und Forschungsmodul – Datenqualität im Studienmodul sichern),
- 6.4.2.7 (Kombinierter Einsatz von Studien- und Forschungsmodul – Machbarkeit einer Studie prüfen),
- 6.4.2.8 (Kombinierter Einsatz von Studien- und Forschungsmodul – Rekrutierung unterstützen),
- 6.5.2.1 (Maximalmodell – Patienten aufnehmen) sowie
- 6.5.2.4 (Maximalmodell – Organisatorische Daten – Dokumentation des Patientenwillens)

zu finden. Unproblematisch in technischer Hinsicht ist die Rücknahme einer Einwilligung. Wird eine Einwilligung aber nachträglich abgeändert, müssen die entsprechenden Änderungen natürlich an allen einschlägigen Stellen der OrgDAT nachgetragen werden.

6.6.7 Besonderheiten bei der Umsetzung

Für den Aufbau der organisatorischen Strukturen sind nach den Grundsätzen der Verhältnismäßigkeit verschiedene Varianten möglich; siehe hierzu die detaillierten Kriterien in Kapitel 6.7. Als Beispiel sei ein Vorschlag für den Ausschuss Datenschutz, insbesondere bei seltenen Erkrankungen, erwähnt: Dieser muss in kleineren Verbünden nicht ein gesondertes Gremium sein, sondern könnte vom Vorstand bzw. Leitungsgremium unter Einbeziehung des

Datenschutzbeauftragten verkörpert werden. Dies ist insbesondere dann angemessen, wenn im Verbund, z.B. in einem Register, nur vorher festgelegte Auswertungen vorgesehen sind.

Die von der TMF als Muster angebotenen Vorlagen für Verträge usw. sind im Anhang zusammengestellt und werden online angeboten[37].

Für das Kontaktmanagement könnte der Einsatz kommerzieller CRM-Software von Interesse sein; Erfahrungen hiermit liegen im TMF-Umfeld aber noch nicht vor.

6.7 Kriterien der Verhältnismäßigkeit

6.7.1 Redundanz und Aufwand

Datenschutzmaßnahmen sind unter der Maßgabe der Verhältnismäßigkeit zu sehen. Auf technischer Ebene können Sicherheitsmaßnahmen sehr aufwändig, damit aber auch sehr teuer gestaltet werden. Unbeliebter Aufwand entsteht insbesondere durch die Schaffung von Redundanzen. Redundanz ist aber ein wichtiger Aspekt in Sicherheitskonzepten, wenn es um hochsensible Daten geht: Wenn eine Sicherheitsmaßnahme unwirksam wird, soll eine „sichere Rückfallposition" erreicht werden. Unwirksam kann eine Sicherheitsmaßnahme aus verschiedenen Gründen werden, beispielsweise:

- nicht regelkonformes Verhalten einzelner Beteiligter,
- unbefugte Kooperation verschiedener Beteiligter oder eines Beteiligten mit einem Externen,
- Ausfall einer technischen Komponente oder
- Kompromittierung einer technischen Komponente.

Beispiele für Redundanzen von Bedeutung für dieses Datenschutzkonzept sind:

- Kombination eines Verbots (z.B. in einer vertraglichen Regelung) mit einer technischen Barriere (z.B. durch Zugriffskontrolle) oder Überprüfung (z.B. durch Protokollierung von Handlungen),
- mehrfache unabhängige Pseudonymisierung,
- Zugriffsschranken für Ärzte trotz der dreifachen Absicherung der ärztlichen Schweigepflicht durch die Androhung strafrechtlicher, zivilrechtlicher und standesrechtlicher Folgen oder
- trotz Pseudonymisierung verschlüsselte Übermittlung von Daten über das Internet.

Verhältnismäßigkeit bedeutet in der Regel nicht, dass ein kontinuierlicher Sicherheitsparameter mehr oder weniger hoch angesetzt wird, sondern dass Redundanzen vermehrt oder abgebaut werden.

37 http://www.tmf-ev.de/datenschutz-leitfaden

Unter den für medizinische Forschungsverbünde vorgesehenen Maßnahmen führt eine sehr feingliedrige Trennung der Verantwortung für einzelne Funktionen der Daten-, Proben- und Rechteverwaltung zu solchen erwünschten Redundanzen. Sie stößt aber dort auf Grenzen der Angemessenheit oder sogar der Durchführbarkeit, wo Forschungsprojekte von relativ kleinen Organisationseinheiten durchgeführt werden. Ein „kleines" Forschungsnetz kann mit wenig Redundanz in technischen und organisatorischen Datenschutzmaßnahmen betrieben werden, wenn es als Angriffsziel weniger attraktiv ist, weniger Angriffspunkte bietet, weniger „Geheimnisträger" hat, organisatorisch übersichtlich ist und mit nur wenigen Komponenten auskommt, in deren Zusammenspiel sich Sicherheitslücken verbergen könnten.

Grundsätzlich sind bei allem Aufwand immer Fälle einer unberechtigten Reidentifizierung konstruierbar. Es muss hier der mögliche Schaden mit dem Aufwand ins Verhältnis gesetzt werden. Daher sind Abwägungen zu treffen zwischen dem Umfang der gespeicherten Daten, dem Risiko einer Reidentifizierung, der Komplexität der Organisation und dem möglicherweise bestehenden Interesse für einen Übergriff. Für alle Forschungsverbünde gilt aber, dass mangelnde Ressourcen kein Argument für mangelhafte Datenschutzmaßnahmen sein können. Insbesondere müssen sich die Zuordnungen von Pseudonymen zu Personen (Identitätsmanagement) und die Forschungsdaten mit ganz wenigen Ausnahmefällen in getrennter Verantwortung befinden.

6.7.2 Parameter für die Risikoabschätzung

Die für die Risikoabschätzung relevanten Aspekte eines medizinischen Forschungsverbundes werden hier in vier Dimensionen gegliedert, die nicht notwendig unabhängig voneinander sind. Es kann keine einfache Formel geben, die aus konkreten Werten für die Parameter die Höhe des Risikos berechnet. Manche Parameter können sich sogar gegenläufig auswirken, indem sie an einer Stelle das Risiko erhöhen, es aber an anderer Stelle senken. Sinn dieser Parameterliste ist vielmehr, für einen konkreten Forschungsverbund potenzielle Schwachstellen aufzudecken. Eine Auswirkung der Risikoabschätzung könnte z.B. sein, dass für den einen Forschungsverbund redundante Sicherheitsmaßnahmen als notwendig angesehen werden, für einen anderen dagegen die Redundanz verringert werden kann; oder dass eine Abschwächung an einer Stelle durch zusätzliche Maßnahmen an anderer Stelle kompensiert wird.

6.7.2.1 Risikodimension „Größe des Forschungsverbundes"

Diese wird durch folgende vier Parameter ausgedrückt:

1. **Fallzahl**: Es ist wohl schwierig, hier explizite allgemein gültige Grenzen zu ziehen. Ein Register oder eine Forschungsdatenbank mit wenigen 100 Patienten ist sicher als klein, eines mit über 10.000 Patienten sicher

als groß einzustufen. Es gibt auch gegenläufige Effekte: Mit der Fallzahl steigt die Attraktivität für einen unbefugten Zugriff auf den Datenbestand des Forschungsverbunds; andererseits sinkt das individuelle Reidentifizierungsrisiko aus MDAT und AnaDAT, da es weniger einzigartige Merkmalskombinationen gibt.

2. **Einzugsbereich und Anzahl der Daten- oder Probenquellen**: Ein Forschungsverbund, der deutschlandweit von Hunderten von Arztpraxen Daten sammelt, ist sicher anders zu bewerten als eine Probensammlung in einem Kliniklabor, die nur Blutproben von Patienten einer bestimmten Fachabteilung enthält. Eine einfachere Logistik bietet weniger Angriffspunkte; bei weniger Datenquellen bestehen bessere Möglichkeiten zur dezentralen Organisation, z.B. der Patientenliste, was je nach Umständen auch ein Sicherheitsgewinn sein kann.
3. **Finanzielle Ausstattung und Zahl der Beschäftigten**: Ein sparsam gefördertes öffentliches Forschungsprojekt ohne kommerzielle Ambitionen oder Aussichten hat sicher wenig Möglichkeiten, komplizierte Schutzmaßnahmen umzusetzen; dadurch steigt die Wahrscheinlichkeit von Sicherheitslücken. Hier ist eine besonders sorgfältige Prüfung unter dem Gesichtspunkt der Verhältnismäßigkeit nötig; mangelnde finanzielle Ausstattung darf kein Argument zur Absenkung des Datenschutzstandards sein.
4. **Komplexität**: Mit steigender Komplexität wächst die Wahrscheinlichkeit für unbeabsichtigte Wechselwirkungen, Sicherheitslücken und Datenlecks.

6.7.2.2 Risikodimension „Brisanz des Forschungsverbunds"

Diese Risikodimension ist hoch mit der potenziellen Attraktivität für einen Angreifer korreliert und kann durch folgende sechs Parameter beschrieben werden:

1. **Art der Erkrankung**: In unserer Gesellschaft werden z.B. psychiatrische Erkrankungen und HIV als stigmatisierend angesehen. Hier spielt auch die öffentliche Beachtung des Forschungsprojekts eine Rolle, da sie sich unmittelbar auf das Vertrauen der Patienten auswirkt. Krankheiten mit hoher Morbidität könnten z.B. für Krankenversicherer interessant sein.
2. **Vollständigkeit der Erfassung**: Je vollständiger die Erfassung, desto größer die Wahrscheinlichkeit, dass eine bestimmte Person erfasst ist, desto geringer aber auch die Wahrscheinlichkeit für einzigartige Merkmalskombinationen. Beispielhafte Fragen: Wird nur eine (zufällig ausgewählte) Kohorte erfasst oder handelt es sich um ein Register mit dem Anspruch auf Vollzähligkeit? Werden alle Probanden mit einer seltenen Erkrankung erfasst? Werden alle Probanden aus einer bestimmten Region erfasst, vielleicht alle Patienten einer Klinik?

3. **Umfang der Datenerhebung**: Je umfangreicher die gespeicherten Datensätze sind, desto mehr unterscheiden sich die Einzelfälle, desto höher ist also das Reidentifizierungsrisiko. Beispielhafte Fragen: Werden nur wenige medizinische Daten erfasst? Werden nur wenige Analysedaten erzeugt, z.B. keine genetischen Daten? Welche soziodemografischen Daten werden erfasst? Werden Angaben erfasst, die Angehörige betreffen?
4. **Forschungsziele**: Diese bestimmen die Brisanz eines Forschungsvorhabens wesentlich mit. Beispielhafte Fragen: Sind genetische Analysen vorgesehen, die ja auch Angehörige der Probanden oder ganze ethnische Gruppen betreffen? Dadurch steigt sowohl die Attraktivität für einen unbefugten Zugriff samt der Zahl der dadurch Betroffenen als auch das Reidentifizierungsrisiko. Sollen Forschungsergebnisse in wichtigen Fällen an die Patienten oder Probanden rückgemeldet werden? Sind Langzeitbeobachtungen der Patienten geplant? In diesen beiden letzteren Fällen muss der „Rückweg" zum Patienten durch geeignete Pseudonymisierung offen gehalten werden, was u.U. zusätzliche Angriffspunkte schafft und das Reidentifizierungsrisiko erhöht.
5. **Art der gespeicherten Daten oder des gelagerten Materials**: Wie einfach ist damit eine Reidentifizierung möglich? Zu berücksichtigen sind z.B. soziodemographische Daten, feingranulare Anamnesedaten, charakteristische Bilddaten von offensichtlichen Missbildungen, Proben oder extrahierte DNA.
6. **Einzigartigkeit von Daten oder Proben**, z.B. durch eine Monopolstellung des Forschungsverbunds in einem bestimmten Bereich.

6.7.2.3 Risikodimension „Organisation des Forschungsverbunds"

Die hier beschriebenen neun Parameter wirken sich ganz wesentlich auf die Beurteilung der Verhältnismäßigkeit aus und sind z.T. relativ leicht zu beeinflussen, wenn das Vorhaben sorgfältig geplant wird.

1. **Beschlagnahmesicherheit**: Man muss nach Dierks [20] davon ausgehen, dass eine rechtlich beschlagnahmefeste Aufbewahrung zentral gespeicherter Daten bei verteilter Datenerhebung nur in wenigen Ausnahmefällen möglich sein wird. Andererseits bedingt das Schutzprinzip der informationellen Gewaltenteilung (vgl. Kap. 4.6), dass entweder medizinische oder identifizierende Daten außerhalb der behandelnden Einrichtung aufbewahrt werden müssen und so der Beschlagnahme unterliegen können.
2. **Präzision der Aufklärung und Einwilligung**: Je weniger bestimmt die Forschungsziele benannt werden können, desto mehr ist durch Verstärkung der Sicherheitsmaßnahmen oder weitergehende Informationstrennung zu kompensieren.
3. **Verteiltheit der Zulieferung** (vgl. auch Kap. 6.7.2.1 Nr. 2): Hier gibt es gegenläufige Effekte: mit der Verteiltheit steigt einerseits die Informationstrennung, andererseits auch die Zahl der Angriffspunkte.

4. **Verteiltheit der Datenspeicherung und Probenlagerung.** Auch hier gilt: Mit der Verteiltheit erhöht sich die Informationstrennung und steigt gleichzeitig die Zahl der Angriffspunkte.
5. **Dauer der Datenspeicherung und Probenlagerung:** Das Risiko von Angriffen ist direkt proportional zu dieser Dauer.
6. **Umfang geplanter Nacherhebungen.** Ist eine erneute oder gar häufig wiederholte Kontaktierung der Patienten oder Probanden vorgesehen? Das erfordert eine komplexere Logistik und erhöht die Zahl der Angriffspunkte sowie die Gefahr von Datenlecks und unbefugter oder sogar unbeabsichtigter Reidentifizierung.
7. **Qualität der Policies und SOPs sowie der vertraglichen Regelungen mit Externen:** Hier sind Abwägungen zwischen technischen und organisatorischen Maßnahmen zu treffen und mögliche oder nötige Redundanzen zu diskutieren.
8. **Vertrauenswürdigkeit einer datenspeichernden oder -verarbeitenden Stelle:** Z.B. ist eine Bundesbehörde, deren Mitarbeiter strengen und öffentlich kontrollierbaren Regeln unterliegen, u.U. vertrauenswürdiger im Sinne eines Datenschutzkonzepts als ein privatwirtschaftlich organisierter Betrieb, dessen Regeln sich bei einer Geschäftsübernahme kurzfristig ändern können oder dessen Datenbestand im Konkursfall auf nicht vorhersagbare Weise weitergegeben wird.
9. **Vorgesehene Monitoring- oder anderweitige Kontrollverfahren:** Eine institutionalisierte und genau festgelegte Nachprüfung aller Verfahrensschritte (z.B. ein Monitoring-Verfahren) kann mit anderen Datenschutzmaßnahmen redundant sein und diese eventuell ersetzen.

6.7.2.4 Risikodimension „Verbindung mit externen Daten"

Hierfür sind zwei Parameter relevant, die kaum beeinflusst, nicht einmal vollständig kontrolliert werden können:

1. **Abgleichmöglichkeit oder -pläne mit anderen Datenquellen oder Registern:** Hier ist einem eventuell erhöhten Reidentifizierungsrisiko durch technische oder organisatorische Maßnahmen zu begegnen.
2. **Vorhandensein von Referenzdateien:** Solche Dateien, z.B. mit genetischen Fingerabdrücken oder soziodemographischen Daten, können zu einer unmittelbaren Reidentifizierung von Daten des Forschungsverbundes führen, so dass die Zusammenführung mit technischen oder organisatorischen Maßnahmen verhindert werden muss; beim Datenexport sind entsprechende Fragen zu stellen und Regelungen zu treffen.

Die Möglichkeiten zum externen Datenabgleich können niemals vollständig und für alle Zukunft beurteilt werden; sie betreffen aber genau das Hauptanliegen eines Datenschutzkonzepts, das Reidentifizierungsrisiko zu vermeiden. Daher ist bei diesen Parametern eine besonders vorsichtige Einschätzung notwendig. Nicht erreichbar ist in der Regel *k*-Anonymität [34]. Abzuwägen sind die Möglichkeiten zur vollständigen, faktischen oder nur formalen Anonymi-

sierung – für die Abgrenzung und Problematik der Verwendung dieser Begriffe sei auf das Glossar in diesem Werk verwiesen – und ihre Konsequenzen bzw. kompensatorische Maßnahmen.

6.7.3 Modellvarianten

Bei der Beschreibung der Module und ihrer Komponenten wurden an verschiedenen Stellen bereits Modellvarianten und abweichende Organisationsformen, sogar Zusammenlegungen von im Standard-Konzept getrennten Funktionen oder Datenspeichern als Möglichkeiten aufgeführt. Bei Abweichungen vom Standardkonzept und insbesondere Vereinfachungen der technischen und organisatorischen Maßnahmen ist immer eine Einzelfallprüfung unter Anwendung der Kriterien erforderlich.

Erleichternd für die Zulässigkeit von Abweichungen ist die Etablierung eines Monitoring- oder Audit-Verfahrens. Solche Verfahren gelten ohnehin als die beste Methode, Regelverstöße von Insidern aufzudecken [38].

Stufen für die Datentrennung sind:

1. getrennte Datenbank-Tabellen,
2. getrennte Datenbanken,
3. getrennte Datenhoheit sowie
4. externer Datentreuhänder.

Stufe 1 ist nur bei monozentrischen klinischen Studien oder im Behandlungszusammenhang angemessen und in dieser Form heute oft in Krankenhausinformationssystemen vorzufinden; sie schützt im wesentlichen davor, dass ein Systemverwalter Identitätsdaten und medizinische Daten zusammen sieht, ohne es zu wollen. Außerdem lässt sich für alle anderen Datenbanknutzer auf dieser Basis leicht eine differenzierte Zugriffsregelung aufbauen.

Als Beispiel für Stufe 2 ist bei institutionsinterner Langzeitforschung (Beispiel: Datawarehouse im Krankenhaus) der Aufbau einer getrennten Datenbank mit einfacher Pseudonymisierung ausreichend, obwohl es sich vom Charakter der Datensammlung her um eine Forschungsdatenbank handelt. Ebenso kann Stufe 2 für eine kleine institutionsinterne Biomaterialbank angemessen sein [2]. Diese Stufe der Datentrennung schützt zusätzlich vor einem Angreifer, der Zugang zu einer der Datenbanken hat, z.B. auf einem unzulänglich gelöschten ausrangierten Datenträger.

Stufe 3 ist der empfohlene Normalfall für die meisten medizinischen Forschungsverbünde und führt bei geeigneter organisatorischer Regelung zu einer angemessenen informationellen Gewaltenteilung.

Stufe 4 kann bei besonders sensiblen Erkrankungen verhältnismäßig sein, etwa um dem Misstrauen von Patienten oder Patientenverbänden gegen die medizinische Forschung entgegenzuwirken.

Das Maximalmodell ist angemessen, wenn in einem großen Forschungsverbund vielfältige Projekte aller Art mit komplexer Datenlogistik durchgeführt werden sollen. Hierfür relevante Kriterien sind:

- Notwendigkeit der langfristigen (Pseudonymisierung) Aufbewahrung von Daten oder Proben (über Behandlungs- oder Studienkontext hinaus), langfristige Forschungsvorhaben wie Spätfolge- oder Lebensqualitätsstudien, Kohortenstudien und
- größere Patienten- oder Probandenzahlen (z.B. keine seltene Erkrankung).

In der Mehrheit der Fälle ist allerdings eine „kleinere" Lösung angemessen. Viele Forschungsverbünde, z.B. Netzwerke für seltene Erkrankungen, benötigen nur eine Klinische Datenbank, eventuell in Kombination mit einer Biomaterialbank. Verbünde, bei denen im Wesentlichen epidemiologische Fragen verfolgt werden, können mit einer Forschungsdatenbank auskommen. Hier sind jeweils die bei der Beschreibung der Einzelmodule vorgeschlagenen Lösungen mit eventuell nötigen, sachgerecht begründbaren Modifikationen angemessen. Bei der Wahl des passenden Modells spielen – auch im Sinne des Datenschutzes – Überlegungen zur Praktikabilität eine wichtige Rolle.

6.7.4 Beispiele

Um die in den vorigen Kapiteln auf einer eher abstrakten Ebene angestellten Überlegungen für die praktische Anwendung nutzbar zu machen, werden hier zahlreiche konkrete Anwendungsbeispiele vorgestellt. Entscheidend ist bei allen Varianten, dass das Reidentifizierungsrisiko nicht erhöht wird.

6.7.4.1 Identitätsmanagement

Ist die Aufteilung des ID-Managements in Patientenliste und Pseudonymisierungsdienst nötig? Wird für die Patientenliste ein externer Treuhänder eingesetzt, kann dieser den Pseudonymisierungsdienst auch zusätzlich übernehmen, wenn dieser auf einem eigenen Rechner mit räumlicher und personeller Trennung von der Patientenliste organisiert wird. Bei PID-Vergabe an der Datenquelle – bei kleineren Projekten sinnvoll – kann der PID als Pseudonym dienen. Ein zusätzlicher Pseudonymisierungsdienst ist dann verzichtbar.

Darf ein PID an der Datenquelle bekannt sein? Ja, wenn er dort erzeugt wird. Bei klinischen Studien ist das so vorgesehen (SIC oder evtl. PID_S). Auch wenn der Forschungsverbund kein Klinisches Modul betreibt, sondern seine Daten direkt im Forschungsmodul speichert, ist die Kenntnis des PID an der Quelle unschädlich, wie auch im „alten" Modell B vorgeschlagen [1].

Weitere Hinweise zu einzelnen Punkten finden sich wie folgt:

- Soll die Patientenliste zentral oder dezentral geführt werden? Siehe Kapitel 6.1.5.1.
- Wo soll die Patientenliste angesiedelt sein? Siehe Kapitel 6.1.5.2.

- Wo soll der Pseudonymisierungsdienst angesiedelt sein? Siehe Kapitel 6.1.5.4.
- Soll ein SIC zentral oder dezentral erzeugt werden? Siehe Kapitel 6.1.1.1.

6.7.4.2 Rechtemanagement

Die folgenden Fragestellungen zum Rechtemanagement sind zu berücksichtigen:

- Sollen Arzt-Identitäten (ADAT) pseudonymisiert werden? Argumente hierzu stehen in einem zusätzlichen Hinweis in Kapitel 6.1.1.2.
- Können die ADAT bei den MDAT gespeichert werden? Das ist in der Regel (vgl. Kap. 6.1.5.1) nicht ratsam, da es Hinweise für eine Reidentifizierung geben kann. Eine Ausnahme wäre denkbar, wenn jeder meldende Arzt für sehr viele Patienten zuständig ist, z.B., wenn nur große Schwerpunktkliniken Daten liefern.
- Sollen Nutzdaten (MDAT) durch das Identitätsmanagement oder an ihm vorbei geleitet werden? (vgl. Kap. 6.1.2 b). Das ist vom Datenschutz aus gesehen – bei adäquater Implementierung – äquivalent und kann daher nach Praktikabilität und Performanz entschieden werden.
- Ist für das Nutzer- und Rechtemanagement ein Verzeichnisdienst notwendig? Das wurde in den Kapiteln 6.2.1.2 und 6.2.5.1 diskutiert.
- Welche Rollenkonflikte können bei Ärzten im Forschungsverbund auftreten und wie soll man mit ihnen umgehen? Hierzu macht das Kapitel 6.2.3.3 einige Aussagen.
- Wo soll das Rechtemanagement angesiedelt sein? Dazu sei auf Kapitel 6.2.4 verwiesen. Unabhängig von der technischen Implementierung unterliegt es der zentralen Verantwortung unter Kontrolle des Ausschusses Datenschutz.
- Ist die Nutzung einer PKI im medizinischen Forschungsverbund anzuraten? Dazu sei auf Kapitel 6.2.5.2 verwiesen.
- Welche Werkzeuge sollen zur Spezifikation von Richtlinien und Regeln eingesetzt werden? Dazu wurden in Kapitel 6.2.5.4 Hinweise gegeben.
- Welche Werkzeuge sollen zur Rechteverwaltung eingesetzt werden? Gesichtspunkte dazu wurden in Kapitel 6.2.5.5 erörtert.

6.7.4.3 Bilddaten

Für die Ansiedlung von Bilddaten gibt es verschiedene Varianten:

- zum Klinischen Modul, wenn eine Referenzbefundung mit möglicher Rückwirkung auf den Patienten vorgesehen ist;
- zum Studienmodul, wenn Bilder direkt in einer klinischen Studie benötigt werden, insbesondere zur Referenzbefundung (Befundung von Bildern durch Referenzradiologen);

- zum Forschungsmodul, wenn die Bilder nur zu Vergleichszwecken bei der Befundung oder als Referenzmaterial für Forschung und Lehre dienen sollen – nur bei *qualitätsgesicherten* Bildern (oder Daten);
- in einem Extra-Modul, wenn übergreifende Zwecke verfolgt werden und die Ansiedlung in einem der anderen Module ein zu hohes Reidentifizierungsrisiko bedeutet.

Hauptkriterium für eine eigenständige im Gegensatz zu einer integrierten Bilddatenbank ist die Erkennbarkeit einer Person im Bildmaterial, wenn dieses außerhalb des Behandlungszusammenhangs gespeichert wird; hierzu sei auch auf die Diskussion von Bilddaten in den Kapiteln 6.5.1 und 6.5.2.3 hingewiesen.

Eine weitere Abwägung der Verhältnismäßigkeit ist erforderlich, wenn Bilddaten für die weitere Nutzung zugänglich gemacht werden sollen (Referenzmaterial, wissenschaftliche Auswertung). Für die Frage allerdings, ob der Export von Bildern einem direkten Zugriffsrecht vorzuziehen ist, sind neben dem Datenschutz auch technische Argumente zu berücksichtigen (Dateigröße); oft, insbesondere zu Referenzzwecken, ist schon wegen der Dateigröße oder der Performanz ein Export nicht sinnvoll. Es soll aber auch verhindert werden, dass mit exportierten Daten eine externe Datenbank aufgebaut wird. Daher ist es meistens besser, einen Online-Zugriff für „Forschungsprojekte" einzurichten, der über passende Zugriffsregelungen gestaltet wird.

6.7.4.4 Biomaterialbanken

Beispiele zu Abwägungen der Verhältnismäßigkeit im Zusammenhang mit Biomaterialbanken im Forschungsverbund sind im generischen Datenschutzkonzept für Biomaterialbanken beschrieben [2].

6.7.4.5 Sonstiges

Für den Zugriff auf Forschungsdaten durch externe Wissenschaftler bis hin zu einem Public-Use-File ist die Hierarchie von Möglichkeiten – in Anlehnung an die Regelungen des statistischen Bundesamtes – zu berücksichtigen, die in Kapitel 5.3 vorgestellt wurde. In der Regel werden für Public-Use-Files nur sehr stark vergröberte Daten bereitgestellt werden können, mit denen man Fragen beantworten kann, die für den Forschungsverbund selbst keine Relevanz mehr besitzen, die aber im öffentlichen Interesse sein könnten.

Sollen Daten beim Versand über das Internet zusätzlich verschlüsselt werden, auch wenn sie pseudonymisiert sind? Ja, denn einerseits kann das Reidentifizierungsrisiko pseudonymisierter Daten bei unbekannten Angreifern nicht eingeschätzt werden. Andererseits wird durch die verschlüsselte Übermittlung die Pseudonymisierung keinesfalls überflüssig, da sie ja den Personenbezug vor dem Empfänger schützen soll. Außerdem würde die Pseudonymisierung

die Daten auch noch schützen, wenn das Verschlüsselungsverfahren, das für die Kommunikation verwendet wird, kompromittiert wird; selbst wenn ein Angreifer die Daten in verschlüsselter Form gespeichert hätte, wären sie dann immer noch vor ihm geschützt.

Beim Studienmodul lassen sich die Varianten „zentrales Datenmanagement" auf der einen, „separate Studiendatenbank für jede Studie" auf der anderen Seite und entsprechend die Verwendung von PID_S bzw. nur SICs vom Datenschutzgesichtspunkt aus beide zufrieden stellend umsetzen, siehe die Diskussion in Kapitel 5.2. Die Entscheidung für eine der Varianten kann also unter Praktikabilitätsgesichtspunkten getroffen werden.

Wann die Einrichtung einer temporären Datenbank zur Qualitätssicherung angemessen ist und was dabei zu beachten ist, wird in Kapitel 6.8 beschrieben. Entscheidend ist hier, dass es sich um ein etabliertes Verfahren handelt, das innerhalb des Forschungsverbundes reguliert ist, und dass keine längerfristige Datenspeicherung vorgesehen ist; d.h., die temporäre Zusammenführung von sonst getrennten Daten wird durch Regelungen kompensiert.

6.7.5 Seltene Erkrankungen

Da Netzwerke für seltene Erkrankungen ein wichtiger Schwerpunkt der Forschungsförderung sind, diese aber einerseits durch geringe Ressourcen, andererseits durch extrem kleine Fallzahlen und vielfältige Fragestellungen gekennzeichnet sind, werden Empfehlungen für solche Forschungsverbünde hier explizit zusammengefasst.

In Europa bezeichnet man eine Krankheit als selten, wenn sie weniger als einen unter 2000 Menschen im Laufe seines Lebens trifft [39]. Das bedeutet, dass in Deutschland auch über längere Zeiträume hinweg oft nur wenige hundert Fälle einer bestimmten Krankheit auftreten. Von den ungefähr 30.000 bekannten Krankheiten werden 5.000 bis 7.000 zu den seltenen Erkrankungen gerechnet. Zählt man diese zusammen, sind sie allerdings kein seltenes Phänomen; in Deutschland leiden rund 4 Millionen Menschen an einer seltenen Erkrankung. Häufig handelt es sich um schwere Krankheiten, die eine aufwändige Behandlung und Betreuung erfordern, für die Betroffenen und ihre Familien mit hoher Belastung verbunden sind und oft schon im Kindes- oder Jugendalter mit dem Tod enden. Ein typisches Beispiel ist der kindliche Lebertumor, der im Zeitraum zwischen 1980 und 2004 nur bei 382 Kindern im Alter von bis zu 15 Jahren auftrat. Für die schwere aplastische Anämie waren es im gleichen Zeitraum in der gleichen Population 280 Fälle [40].

Bei vielen seltenen Erkrankungen ist die ihnen zugrunde liegende Ursache ungeklärt. Man nimmt an, dass bei etwa 80% genetische Veränderungen ursächlich sind, allerdings sind die jeweils betroffenen Gene häufig noch nicht identifiziert. Für einige Erkrankungen gibt es bisher noch nicht einmal Ansätze zur Erforschung der Krankheitsursachen [7].

Folglich ist in vielen Fällen die medizinische Versorgung der Kranken noch unbefriedigend. Um aber in der klinischen Forschung valide Ergebnisse zu erzielen, sind Patientenzahlen erforderlich, die einzelne Fachleute und Zentren kaum erreichen können. Zur Verbesserung der Versorgung der Patienten ist es daher unumgänglich, die Forschung zur Klärung der Krankheitsursachen sowie zur Entwicklung, Validierung und Etablierung von Diagnoseverfahren und Therapiekonzepten zu konzentrieren und zu intensivieren [39]. Gleiches gilt für die epidemiologische Forschung, insbesondere wenn Varianten einer Erkrankung untersucht und multifaktorielle Ursachen geklärt oder regionale Unterschiede und zeitliche Trends erkannt werden sollen. Sofern Behandlungsdokumentationen und Proben nicht systematisch und flächendeckend gesammelt werden, besteht keine Chance, eine seltene Krankheit erfolgreich zu erforschen. Daher ist die Vernetzung zwischen behandelnden und forschenden Ärzten sowie medizinischen Einrichtungen eine wesentliche Voraussetzung für den wissenschaftlichen Fortschritt.

Als Musterbeispiel können hier die großen Erfolge im Bereich der Pädiatrischen Onkologie und Hämatologie dienen [41], die durch eine solche systematische Rückkopplung der Forschung in die Versorgung erreicht wurden. Die im Netzwerk vorhandene diagnostische und therapeutische Spitzenkompetenz für die jeweilige Krankheit steht für die Behandlung aller teilnehmenden Patienten zur Verfügung, so z.B. die zentrale Referenzdiagnostik (Labor, Pathologie, Radiologie) oder Konsiliardienste durch die Studienleitung für Therapie-Entscheidungen. Andererseits profitieren diese führenden Fachleute durch die wesentlich verbesserte Datenlage und die Zuarbeit aller Behandler für die weitere Forschung.

Im vitalen Interesse der Patienten selbst liegt es auch, einen möglichst großen Kreis von Fachleuten einzubeziehen. Und für sie ist es ebenfalls wichtig, dass ihre Daten und Proben nicht in abgegrenzten Projekten „vergeudet" werden, sondern in einem gemeinsamen Pool möglichst effizient verwertet werden.

Bei den meisten seltenen Erkrankungen (dies gilt insbesondere für sehr seltene Erkrankungen) ist die rechtlich gebotene Trennung zwischen den Daten zur Behandlung, zur klinischen Forschung und zur epidemiologischen Forschung kontraproduktiv für alle Beteiligten: Die Behandlungsdaten sind für die Forschung ebenso wichtig, wie die Forschungsdaten die Behandlung unterstützen können, und jeder Patient ist immer auch zugleich für die Forschung von Bedeutung. Auch die Patienten selbst haben oft ein großes Interesse an Forschungsprojekten, da neue Behandlungsoptionen aufgrund der notwendigen systematischen Evaluation nur im Rahmen einer Studie zur Verfügung stehen. In einer solchen Situation kann die Dokumentation im Rahmen eines Forschungsprojekts einer elektronischen Patientenakte gleichkommen, die auch für künftige Auswertungen genutzt werden kann.

Zusammengefasst sind die Versorgungs- und Forschungsziele eines Forschungsverbundes für seltene Erkrankungen:

- Koordination der Forschung zu einer Erkrankung mit Akkumulation ausreichender Patientenzahlen, um statistisch sinnvolle Auswertungen, die Untersuchung von Sonderfällen und Varianten, von regionalen Unterschieden und zeitlichen Trends sowie genetische Analysen und die Rekrutierung für künftige klinische Studien zu ermöglichen.
- Langzeitbegleitung der Patienten durch eine möglichst vollständige und standardisierte Dokumentation.
- Optimierung der Therapie und der Betreuung durch den Aufbau eines Behandlungsnetzes unter Beteiligung der führenden Fachleute (horizontale Vernetzung), Konsultationssystem, Erarbeitung von Leitlinien zur Diagnostik und Therapie, Kooperation der Fachzentren untereinander und mit niedergelassenen Ärzten, die in der Regel sehr selten oder erstmalig mit einer seltenen Erkrankung konfrontiert sind.
- Sammlung von Referenzfällen.
- Förderung der direkten Kommunikation von Experten untereinander sowie mit weniger erfahrenen Ärzten, z.B. in einem Expertenforum (s. Kap. 5.1.2.5).
- Bereitstellung von Informationen für Patienten (vertikale Vernetzung) über Ursachen, Diagnostik, Verlauf, Therapiemöglichkeiten; Einbeziehung von Selbsthilfegruppen, insbesondere auch Förderung der Patienten-Community.

Eine Vollerfassung von Patienten mit bestimmten Behinderungen oder lebensbestimmenden Erkrankungen ist aber – auch vor dem Hintergrund der geschichtlichen Erfahrungen in Deutschland – gesellschaftspolitisch heikel. Eine mögliche Stigmatisierung ist, je nach Krankheitsbild, nicht auszuschließen. Erschwerend kommt hinzu, dass viele seltene Erkrankungen mit auffälligen, nicht zu verbergenden körperlichen Erscheinungsformen einhergehen, die eine wirksame Anonymisierung oder Pseudonymisierung der Daten erschweren, z.B. körperliche Fehlbildungen.

Aus diesen Rahmenbedingungen ergeben sich folgende Überlegungen für einen auch aus Sicht des Datenschutzes adäquaten Aufbau eines Forschungsverbundes für seltene Erkrankungen:

Grundsätzlich ist ein solcher Forschungsverbund ein typischer Anwendungsfall für ein Klinisches Modul, wobei hier zweckmäßigerweise nur eine einzige zentrale Klinische Datenbank (meist als Register bezeichnet) aufgebaut werden sollte. In der Regel wird aber auch eine zugehörige Biomaterialbank benötigt. Das Zusammenspiel dieser beiden Komponenten ist im generischen Datenschutzkonzept für Biomaterialbanken [2] beschrieben. Üblich und sinnvoll ist es dabei, eine Gruppe verwandter Krankheiten in einem gemeinsamen Forschungsverbund zu untersuchen.

Wichtig und in der Regel von allen Beteiligten gewünscht ist gerade bei seltenen Erkrankungen die enge Kooperation mit Patientenorganisationen und Selbsthilfegruppen, soweit diese schon vorhanden sind; andernfalls könnte es gerade ein Ziel des Forschungsverbunds sein, solche Gruppen ins Leben zu rufen und zu unterstützen. Da oft Kinder die Betroffenen sind, ist zunächst die Einwilligung der Eltern relevant, die, sobald das Kind die nötige Einsichtsfähigkeit erlangt hat, durch dessen eigene Einwilligung zu ersetzen ist. In den (aller Erfahrung nach seltenen [7]) Fällen, wo diese verweigert bzw. zurückgezogen wird, sind geeignete Anonymisierungsmaßnahmen durchzuführen, oder gar, wenn eine wirksame Anonymisierung unmöglich ist, die entsprechenden Falldaten zu löschen.

Auf jeden Fall sollte die Patientenliste unabhängig von der Datenbank geführt werden; sie könnte auch die LabIDs verwalten, wenn diese nicht ohnehin von den Laboren vergeben werden. Hierfür bilden sich zur Zeit im Umfeld der TMF zentrale Dienstleister an Universitätskliniken heraus, die auch die Patientenlisten verschiedener Forschungsverbünde verwalten; wichtig dabei ist aber, dass es genügend viele davon gibt und somit eine angemessene Verteilung der damit verbundenen informationellen Gewalt gewährleistet ist. Die Frage, ob es sinnvoll ist, um die Zugehörigkeit einer Person zu einer bestimmten Diagnosegruppe zu verschleiern, die Patientenlisten mehrerer seltener Erkrankungen zusammenzufassen, wurde in Kapitel 6.1.5.2 diskutiert und negativ beantwortet

Die Pseudonymisierung beim Export von Daten zur statistischen Auswertung ist in diesem Modell eine Funktion der Klinischen Datenbank, ebenso wie die Funktionen zur „Rekrutierung“, d.h. der Gewinnung von Patienten für neue klinische Studien. Die Gestaltung dieser Funktion wird in den Kapiteln 5.1.2.9 und 5.1.2.10 beschrieben.

Software-Produkte „von der Stange“, die zum Betrieb einer Klinischen Datenbank oder eines Registers, insbesondere für seltene Erkrankungen, direkt geeignet sind, sind bisher nicht erhältlich. Hier besteht noch Entwicklungsbedarf, entsprechende Arbeiten und Projekte sollten von der TMF koordiniert werden. Dies kann, ebenso wie die zentrale Bereitstellung von Dienstleistungskapazität für den Betrieb von Patientenlisten, wesentlich dazu beitragen, der Ressourcenknappheit der Forschungsverbünde für seltene Erkrankungen zumindest teilweise zu begegnen.

6.8 Qualitätssicherung

Unter Qualitätssicherung wird in diesem Text die Sicherstellung der Datenqualität verstanden. Andere Aspekte des Qualitätsmanagements wie Struktur-, Prozess- und Ergebnisqualität sind zwar für medizinische Forschungsverbünde auch von Bedeutung, spielen für das Datenschutzkonzept aber kei-

ne unmittelbare Rolle. Die Datenqualitätssicherung dagegen muss notwendigerweise oft mit personenbezogenen Daten arbeiten und ist daher datenschutzrelevant.

Damit die medizinische Forschung aus den verfügbaren Daten valide Ergebnisse gewinnen kann, müssen die Daten hohe Anforderungen an Genauigkeit, Vollständigkeit und Korrektheit erfüllen. Daher ist die Datenerhebung und Datenverarbeitung in medizinischen Forschungsverbünden in der Regel mit einer oder mehreren Stufen der Qualitätssicherung verbunden; deren Umfang und Komplexität sind für das einzelne Projekt durch das Studiendesign und die Normen, denen es sich unterwirft, definiert, für den gesamten Forschungsverbund durch das Zusammenspiel der Module und Komponenten. Dabei geht es immer um die Ergänzung und Korrektur fehlerhafter, fehlender, unvollständiger und unplausibler Daten. Bei epidemiologischen Studien setzen die Forscher selbst bei der Studienplanung die Anforderungen und Verfahren fest. Bei klinischen Studien sind die Anforderungen in „Standard Operation Procedures“ (SOPs) durch generelle Richtlinien festgelegt.

Typische Datenfehler sind

- Schreibfehler, wie Zahlendreher,
- Einträge in falschen Datenfeldern,
- fehlende Einträge,
- inhaltliche Irrtümer, wie Fehldiagnosen.

Manche dieser Fehler können automatisch durch die Erfassungssoftware abgefangen werden: Ein Monat 13 existiert z.B. nicht; eine Zahl im Namensfeld muss ein Irrtum sein. Fehldosierungen von Medikamenten können zumindest einen Warnhinweis auslösen. Schreibfehler in Namen oder Fehldiagnosen sind dagegen automatisch kaum zu erkennen. Daher ist neben möglichst guten Fehlererkennungsalgorithmen der Software in der Regel auch eine Nachkontrolle durch Menschen nötig. Hierfür ist zunächst das Datenmanagement der jeweiligen Datenbank zuständig, je nach Herkunft oder geplanter Verwendung der Daten sind weitere Kontrollen mit mehr oder weniger aufwändigen Verfahren notwendig.

Die Prozesse der Qualitätssicherung sind in den allgemeinen Richtlinien des Forschungsverbundes zu beschreiben und in Form von SOPs genau festzulegen. Insbesondere ist dort anzugeben, wo personenbezogene Daten benötigt werden und wie mit diesen umgegangen wird. Hinweise dafür geben die folgenden Kapitel.

6.8.1 Klinisches Modul

Im Behandlungskontext werden Daten primär zu Abrechnungszwecken erhoben; eine darüber hinaus gehende Dokumentation ist wegen des damit ver-

bundenen Arbeitsaufwandes in der Regel nicht möglich. Das führt dazu, dass z.B. Nebendiagnosen, die für die Abrechnung keine Rolle spielen, nicht notiert werden; möglicherweise wird sogar die Hauptdiagnose im Hinblick auf den Erlös „optimiert". Solche Daten sind für die medizinische Forschung nahezu unbrauchbar; nicht einmal grobe Krankheitsstatistiken können damit zuverlässig erstellt werden. Sollen die Daten für Auswertungen irgendwelcher Art verwendet werden, ist hier bereits eine erste Stufe der Qualitätssicherung vonnöten. Daher sind direkt bei der Dateneingabe im Klinischen Modul eines Forschungsverbunds qualitätssichernde Maßnahmen einzuführen, die als Nebeneffekt auch die klinische Befundkommunikation verbessern. Solche Maßnahmen bestehen aus Algorithmen zur Überprüfung der Vollständigkeit und Plausibilität der klinischen Daten, die unmittelbar bei der Eingabe durch die Erfassungssoftware ausgeführt werden, sowie aus anschließenden im Behandlungsprozess vorhandenen qualitätssichernden Maßnahmen. Da Online-Eingabe und Abfrage ausschließlich durch den behandelnden Arzt erfolgen kann, sind asynchrone Mechanismen zur Datenüberprüfung zunächst entbehrlich, ein Rückgriff auf die Klinik (oder Rückfrage-Management) beim Export von Forschungsdaten ist oft nicht notwendig, insbesondere, wenn die Klinische Datenbank die einzige zentrale Datenbank des Forschungsverbunds ist, wie es z.B. bei chronischen oder seltenen Erkrankungen häufig zutrifft und im Modell A des alten generischen TMF-Datenschutzkonzept beschrieben wurde [1].

Das Studiendesign eines versorgungsnahen Registers oder einer Beobachtungsstudie kann aber auch ein geeignetes Monitoring-Verfahren einschließen, um verbleibende Fehler zu eliminieren. Ferner ist ein Rückmeldeverfahren notwendig, wenn die Daten für andere Module oder Projekte exportiert und dort Fehler entdeckt werden. Im Klinischen Modul kann also auch, wie in einem Studienmodul (s.u.), ein ausgefeiltes Qualitätssicherungsverfahren mit Rückfrage-Management, Monitoring und Auditing vorgesehen werden.

6.8.2 Studienmodul

In klinischen Studien, bei denen durch die Studienplanung eine bestimmte Datenqualität verbindlich vorgeschrieben ist, ist ein komplexes, zu Beginn der Studie detailliert festgelegtes Qualitätssicherungsverfahren die Regel. Dieses besteht aus konsiliarischer Beratung, Rückfrage-Management, Monitoring, Safety-Management und Audit.

Für die Nutzung in klinischen Studien nach AMG entsprechen fertig entwickelte Softwaresysteme üblicherweise hinsichtlich Funktionsumfang, Qualitätssicherung und Dokumentation den umfangreichen gesetzlichen Vorgaben bzw. den Kriterien der Good Clinical Practice (GCP). Hierzu gehört z.B. die Funktion eines umfassenden Audit-Trails, in dem alle Änderungen an Datensätzen nachvollziehbar gespeichert werden und der dem Monitor zugänglich sein muss.

6.8.2.1 Konsiliarische Beratung

Ein wichtiges Element bei klinischen Studien ist die Begleitung der Behandlung durch einen erfahrenen Studienarzt. Alle Daten des Patienten werden durch ihn bezüglich einer richtigen und adäquaten Behandlung begutachtet. Der Studienarzt kann den behandelnden Arzt gegebenenfalls konsiliarisch beraten. Wenn diese Beratung vom Studienarzt oder anderem ärztlich geführten Personal mit Patientenkontakt durchgeführt wird, geschieht dies im Regelfall in Kenntnis der Identität des betroffenen Patienten. In jedem Fall steht die konsiliarische Beratung in engem Zusammenhang mit der Behandlung des Patienten. Hierbei anfallende Datenänderungen sind in der Studiendatenbank zu dokumentieren.

6.8.2.2 Rückfrage-Management

Als zweite Stufe der Qualitätssicherung klinischer Forschungsdaten (nach der Eingabekontrolle) ist ein Rückfrage-Management vorgesehen. Dieses wird vom Datenmanagement ausgelöst und ist mit einer Kommunikation pseudonymer Daten verbunden, wobei ein Kommunikationspartner die Zuordnung des Pseudonyms (hier SIC oder PID_S) zum Patienten kennt und der andere im Regelfall nicht. So können im zentralen Datenmanagement Rückfragen zu den Daten eines Patienten formuliert werden, ohne dass die Identitätsdaten des Patienten hierfür benötigt werden. Wenn diese Rückfragen vom Studienarzt oder anderem ärztlich geführten Personal mit Patientenkontakt bearbeitet werden, geschieht dies im Regelfall in Kenntnis der Identität des betroffenen Patienten.

6.8.2.3 Monitoring

In einer dritten Stufe werden Monitore eingesetzt, die die eingegebenen Daten vor der Auswertung noch einmal auf Vollständigkeit und Plausibilität überprüfen. Dabei wird auch der Dateneingabevorgang hinterfragt und die Übereinstimmung mit den Quelldaten optisch nachvollzogen. Daher müssen Monitore sowohl Zugang zu den von ihnen zu überprüfenden zentralen Patientendaten wie auch zu den Quelldaten in den beteiligten Zentren und behandelnden Einrichtungen haben. Da dieser Nutzerkreis außerhalb des Behandlungsverhältnisses steht und gleichzeitig auch Zugriff auf personenbezogene Unterlagen mit Klartext-Identitätsdaten hat, müssen Patienten vor Aufnahme in die Studie darüber aufgeklärt werden und können nur nach einer entsprechenden Einwilligung teilnehmen.

6.8.2.4 Safety-Management

Im Rahmen der allgemeinen Qualitätssicherung in klinischen Studien werden alle unerwarteten Studienereignisse je nach ihrem Schweregrad an den Spon-

sor und an die zuständigen Behörden gemeldet. Hierbei gibt es behördliche Verpflichtungen gemäß dem AMG. In der Regel erfolgen diese Meldungen ausschließlich mit pseudonymisierten Daten. Eine Überprüfung der Quelldaten erfolgt dann im Rahmen des allgemeinen Monitoring.

6.8.2.5 Audit

Als Audit werden allgemein Untersuchungsverfahren bezeichnet, die dazu dienen, Prozessabläufe hinsichtlich der Erfüllung von Anforderungen und Richtlinien zu bewerten. Die Audits werden im Regelfall von speziell hierfür geschulten, unabhängigen und externen Fachleuten durchgeführt. Im Kontext von medizinischen Forschungsprojekten, insbesondere klinischen Studien ist ein Audit eine weitere Stufe der Qualitätssicherung. Hierbei werden sämtliche Prozesse auf Übereinstimmung mit Studienplan, Richtlinien, SOPs und anderen verbindlichen Festlegungen – auch hinsichtlich Datenschutzmaßnahmen – geprüft. Ein Zugriff auf konkrete Daten ist, im Gegensatz zum Monitoring, allenfalls stichprobenartig nötig; ein Personenbezug muss dabei nicht offenbart werden.

6.8.3 Forschungsmodul

Wird eine Forschungsdatenbank aus einer ausreichend qualitätsgesicherten Klinischen Datenbank oder einer Studiendatenbank gespeist, kann man in der Regel eine genügende Datenqualität annehmen. Kommen Daten auf anderem Wege in die Forschungsdatenbank, ist vor der Übernahme der Informationen ein vorgeschaltetes Qualitätssicherungssystem erforderlich, welches im Feedback zur Klinik oder Datenquelle Mängel in der Plausibilität und Vollständigkeit der Daten minimiert. Dieses muss besondere Datenschutzanforderungen erfüllen, da viele qualitätssichernde Prozesse nach einer Pseudonymisierung der Daten nicht mehr sinnvoll durchgeführt werden können und daher nur mit einem Personenbezug möglich sind (s.u.). Kompliziert wird das Verfahren durch die Möglichkeit, dass spätere Daten zum gleichen Fall in der Regel einen neuen Qualitätssicherungsprozess für den Gesamtdatensatz auslösen.

Zusätzlich zu diesen Routine-Verfahren ist auch ein Korrektur-Prozess vorzusehen, der später nachgereichte Korrekturen übernimmt, die von der Datenquelle, aus einem anderen Modul des Forschungsverbunds oder von Betroffenen selbst in Wahrung seines datenschutzgesetzlichen Berichtigungsrechts angestoßen werden.

6.8.3.1 Allgemeine Anforderungen an das Verfahren (bei Daten aus dem Behandlungszusammenhang)

Die Qualitätssicherung erfordert einen unbehinderten Austausch von Informationen zwischen dem dokumentierenden Arzt, der die Daten beim Patien-

ten und aus seiner Kranken- und Behandlungsgeschichte erhebt, und der prüfenden Stelle. Die prüfende Stelle kann die Stelle sein, welche die Daten sammelt und speichert, oder eine dazwischen geschaltete Stelle, die das Monitoring übernimmt; es kann aber auch eine eigene, unabhängige Stelle dafür eingerichtet werden. Sie wird im Folgenden als QS-Service bezeichnet.

Liegt die Datenquelle im Behandlungszusammenhang, so gibt der dokumentierende Arzt bzw. die erhebende Klinik die Daten mit einem Patientenidentifikator (PID) weiter, der auch vom QS-Service gespeichert und benutzt wird, um den entsprechenden Datensatz in der Kommunikation mit der Klinik zu identifizieren. Die datenschutzrechtliche Zulässigkeit beruht darauf, dass dem QS-Service ein Zugriff auf die entsprechenden IDAT, die in der Klinik liegen, verwehrt wird. Sie entspricht damit einem bei der Speicherung von Forschungsdaten in Studienzentren oder übergeordneten Forschungsdatenbanken weithin genutzten Standard.

6.8.3.2 Workflow für die Qualitätssicherung von der Behandlung in die Forschungsdatenbank

Zusammengefasst läuft folgender Prozess ab, in dem die verschiedenen Zustände der Daten begrifflich unterschieden werden (s. Abb. 22):

- Der QS-Service erhält von der Klinik die mit dem PID verknüpften „Erhebungsdaten".
- Er benötigt auch den Vergleich mit früher erhobenen Daten, die bereits in der Forschungsdatenbank gespeichert sind; diese werden als „Kontextdaten" temporär zur Verfügung gestellt. Dazu übergibt er eine Liste der aktuellen PIDs an den Pseudonymisierungsdienst, der diese in eine entsprechende PSN-Liste umwandelt.

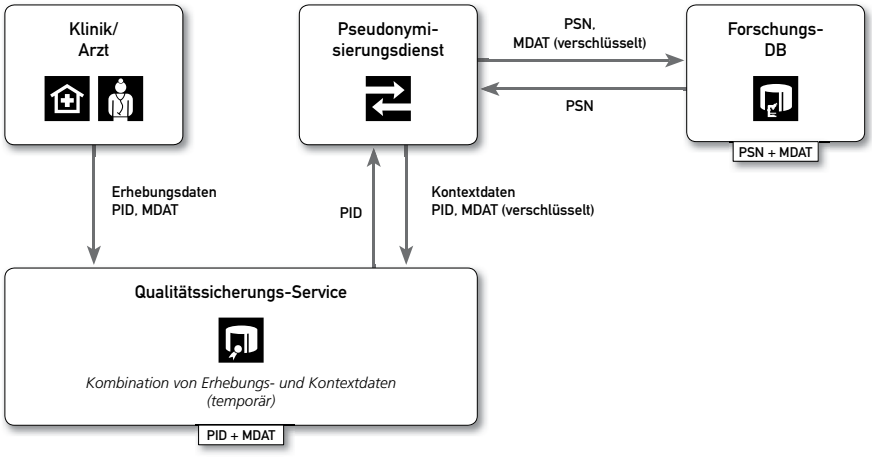

Abb. 22 Der QS-Service übernimmt Erhebungsdaten und Kontextdaten.

- Der Pseudonymisierungsdienst sendet die PSN-Liste an die Forschungsdatenbank.
- Die Forschungsdatenbank sendet die zugehörigen Kontextdaten (mit dem öffentlichen Schlüssel des QS-Service verschlüsselt) an den Pseudonymisierungsdienst.
- Der Pseudonymisierungsdienst ersetzt die PSN wieder durch die entsprechenden PIDs und sendet diese, zusammen mit den immer noch verschlüsselten Kontextdaten an den QS-Service zurück.
- Der QS-Service führt in Kommunikation mit der Klinik die Qualitätssicherung für die mit PID gekennzeichneten Daten durch. Die Behandlungseinrichtung kann dabei auf die Patientenakte und die Original-Dokumentation zugreifen.
- Der QS-Service übergibt die korrekten oder korrigierten Erhebungsdaten über den Pseudonymisierungsdienst der Forschungsdatenbank, wo sie als Forschungsdaten dauerhaft gespeichert werden.
- Danach wird der temporäre Bestand an Erhebungsdaten und Kontextdaten gelöscht.

Der QS-Service führt für seine Aufgaben eine temporäre Datenbank, die in sequentiellen Verfahren mit neuen Erhebungsdaten gefüllt wird, die im Rahmen der Qualitätssicherung laufend abgearbeitet werden. Plausible oder durch Korrektur plausibel gemachte Datensätze werden nach Abschluss der Teilprozesse in Forschungsdaten transformiert und aus der temporären Datenbank gelöscht.

6.8.3.3 Daten aus externen Quellen

Im Forschungsmodul sind, insbesondere bei epidemiologischen Fragestellungen, oft auch Datenabgleiche mit externen Datenquellen bis hin zu Meldeämtern, Gesundheitsämtern und Standesämtern vorgesehen. Diese sind im Rahmen des Qualitätssicherungsprozesses zu definieren und müssen natürlich die rechtlichen Rahmenbedingungen (s. Kap. 4.3.4) einhalten. Im Gegensatz zu dem in Kapitel 6.8.3.2 beschriebenen Verfahren muss hierbei auf Identitätsdaten zurückgegriffen werden. Um die für den QS-Service definierten Regeln nicht aufzuweichen, ist zu empfehlen, dass die externe Kommunikation über die Patientenliste oder eine weitere, eigens für diesen Zweck beauftragte Datentreuhänderstelle abgewickelt wird.

Enthält der Forschungsverbund auch ein Klinisches Modul oder ein Studienmodul, so können die Ergebnisse eines solchen externen Datenabgleichs auch wieder Rückmeldungen oder Rückfragen in dieses auslösen.

6.8.3.4 Einrichtung eines QS-Service

Der QS-Service als besonderer Dienst muss nur in Forschungsverbünden extra aufgesetzt werden, die Daten direkt in eine Forschungsdatenbank aufnehmen,

ohne dass diese zuvor die Qualitätssicherungsprozesse eines Klinischen oder Studienmoduls durchlaufen haben. Er benötigt eine eigene Datenbank, in der Daten während des laufenden Prozesses mit einem PID gekennzeichnet gehalten und nach Beendigung dieses Prozesses (Übergabe qualitätsgesicherter Daten an die Forschungsdatenbank) gelöscht werden. Der Datenbestand ändert sich also laufend, wobei die einzelnen Datensätze nur temporär vorgehalten werden.

Der temporäre Bestand wird auf diese Weise ständig nach der Zahl der Datensätze und in den Inhalten modifiziert; da er niemals einen Zustand erreicht, der nach anderen Regeln als denen der Qualitätssicherung als konsolidiert gelten könnte, ist es nicht möglich, die Daten in anderer als der vorgesehenen Art und Weise zu nutzen. Es besteht kein Anreiz, den Bestand regelwidrig z.B. für irgendwelche Forschungsfragen zu gebrauchen.

Es ist ein Regelwerk festzulegen, nach dem der zulässige Gebrauch des temporären Datenbestands beschrieben und die Einhaltung der Regeln von Dritten (z.B. von Seiten des betrieblichen Datenschutzes) geprüft und bestätigt werden kann.

6.8.4 Patientenliste

Auch die Führung einer Patientenliste im Forschungsverbund kann als Teil der Qualitätssicherungsbemühungen angesehen werden (s. Kap. 6.1.2. a). Hier ist die eindeutige Identifikation des Patienten durch einen fehlertoleranten Record-Linkage-Algorithmus gemeint. Da dieser nicht perfekt arbeiten kann, ist die regelmäßige (z.B. einmal jährlich, je nach Zeithorizont des Forschungsverbunds oder einzelner Projekte) Überprüfung der Patientenliste auf Synonyme (clerical review) vorzusehen; entsprechende Korrekturen sind an die einzelnen Module weiterzugeben.

6.8.5 Rückmeldungen von Datenfehlern

Eine Datenkorrektur in einem Modul des Forschungsverbundes oder in der Patientenliste zieht unter Umständen die Notwendigkeit von Korrekturen in anderen Modulen nach sich. Gleiches gilt, wenn ein Betroffener sein Berichtigungsrecht wahrnimmt. Für diese Korrekturen sind geeignete Rückmeldungsprozesse zu definieren.

6.8.5.1 Nutzung der Daten einer Forschungsdatenbank zum Zwecke der Qualitätssicherung

Wird das Forschungsmodul mit anderen Modulen (z.B. dem Studienmodul oder Klinischen Modul) gekoppelt, so können schon vorhandene Daten eines Patienten aus dem Forschungsmodul zum Zwecke der Qualitätssicherung ge-

nutzt werden. Eine genaue Beschreibung befindet sich im Kapitel 6.4 zum kombinierten Einsatz von Studien- und Forschungsmodul.

6.8.5.2 Kombination von Studienmodul und Klinischem Modul

Die entsprechenden Rückmeldungsprozesse zur Korrektur von Datenfehlern wurden in Kapitel 6.3 ausführlich beschrieben.

6.8.5.3 Kombination von Klinischem Modul und Forschungsmodul

Hier lassen sich die nötigen Prozesse analog zur Kombination von Studien- und Forschungsmodul beschreiben.