Das TMF-Datenschutzkonzept für medizinische Datensammlungen und Biobanken

Klaus Pommerening, Johannes Drepper, Thomas Ganslandt, Krister Helbing, Thomas Müller, Ulrich Sax, Sebastian Semler, Ronald Speer

Institut für Medizinische Biometrie, Epidemiologie und Informatik
Universitätsmedizin Mainz
55101 Mainz
pom@imbei.uni-mainz.de

Abstract: Medizinische Forschung dient der Weiterentwicklung diagnostischen und therapeutischen Wissens und nützt den Patienten durch Verbesserung der Behandlungsqualität. Nötig dafür sind große Mengen an Daten, zunehmend auch Proben und genetische Daten. Allerdings gehören medizinische Daten und Proben zu den sensibelsten persönlichen Informationen und müssen sorgfältig geschützt werden; sie können nicht einfach für Forschungsprojekte zur Verfügung gestellt werden. Um zu zeigen, wie medizinische Forschung unter diesen Randbedingungen möglich ist, entwickelte die Telematikplattform für medizinische Forschungsnetze (TMF) ein generisches Datenschutzkonzept für die Prozessierung von Daten und Proben in medizinischen Netzen und Biobanken. Die wichtigsten Methoden dafür sind informationelle Gewaltenteilung durch eine verteilte Netzarchitektur mit Datentreuhänderdiensten, ein auf Pseudonymen basierendes Identitätsmanagement sowie Mustervorlagen für Einwilligungserklärungen, Policies und Verträge. Eine gründliche Revision des Konzepts auf Grund der bisherigen Erfahrungen aus abgeschlossenen und versuchten Implementationen ist zurzeit in Arbeit.

1 Einleitung

1.1 Medizinische Versorgung und Forschung

Die Verbindung von medizinischer Forschung und Versorgung ist eine der zentralen Herausforderungen des E-Health, insbesondere mit dem Ziel, die Behandlung der Patienten am aktuellen Stand der Forschung auszurichten. Dazu muss ein effizienter Wissenstransfer aus dem Forschungsbereich in die Krankenversorgung gewährleistet sein. In der umgekehrten Richtung ist die medizinische Forschung auf zuverlässige Daten und Proben hoher Qualität von den Patienten angewiesen. Somit hängt der Fortschritt der Medizin entscheidend vom reibungsarmen Fluss von Fragestellungen, Ideen, Daten und Ergebnissen zwischen Krankenversorgung und Forschung ab. Eine wesentliche Voraussetzung dafür ist eine effiziente gemeinsame Infrastruktur.

Neben der medizinischen Grundlagenforschung, die kaum auf Patientendaten, wohl aber zunehmend auf Biomaterialien und genetische Informationen angewiesen ist, sind typische medizinische Forschungsprojekte die klinischen und epidemiologischen Studien.

Klinische Studien sind Studien direkt an Patienten, um etwa neue Therapieverfahren zu prüfen, und unterliegen in den meisten Fällen strengen Regularien aus dem Arzneimittelgesetz (AMG) oder dem Medizinproduktegesetz (MPG). Multizentrisch sind diese Studien, wenn sie nicht nur in einer Einrichtung, etwa einem Krankenhaus, sondern in Kooperation mehrerer Stellen durchgeführt werden; die Motivation hierfür ist meist das Erreichen genügend großer Fallzahlen; insbesondere gilt dies bei seltenen Erkrankungen, für die es bisher noch kaum wirksame Therapien gibt. Bei epidemiologischen Studien ist das Ziel die Erkennung von Krankheitsursachen und -trends im Bevölkerungsbezug und von Langzeiteffekten therapeutischer Maßnahmen, so dass hier der Aufbau langfristig verfügbarer Datensammlungen besonders erforderlich ist.

Typischerweise findet die Datenverarbeitung in der medizinischen Forschung oft weit entfernt vom Patienten statt, so dass eine unmittelbare Kontrolle durch die Betroffenen kaum möglich ist; hinzu kommt das – u. a. vom nationalen Ethikrat postulierte – Recht auf Nichtwissen, z. B. wenn in einem Forschungsprojekt problematische genetische Veranlagungen entdeckt werden. Daher liegt der Schutz der sensiblen medizinischen Informationen weitgehend in fremder Hand. Das erfordert zur Sicherung der Grundrechte starke Maßnahmen, z. B. vertrauenswürdige Organisationen und Systeme, möglichst gute technische Lösungen und rechtlich-organisatorische Rahmenbedingungen, die Restrisiken auffangen.

1.2 Beispiele forschungsorientierter medizinischer Netze in Deutschland

Als typisch kann man ein Szenario aus der Pädiatrischen Onkologie und Hämatologie (Kinderkrebsbehandlung) ansehen, wo die Zusammenarbeit zwischen den Kliniken und den führenden Forschern im Kompetenznetz für die Pädiatrischen Onkologie und Hämatologie (KPOH) [POH09] organisiert ist und für fast alle Krankheitsbilder deutschland- oder europaweite multizentrische Studien durchgeführt werden. Ein Fallbeispiel: Bei einem kleinen Mädchen findet der Kinderarzt einen Verdacht auf eine Krebserkrankung. Er überweist das Kind an eine Schwerpunktklinik. Dort sichert ein Spezialist die Diagnose mit der Hilfe von zentralen Referenzinstitutionen (Pathologie, Radiologie); er nimmt das Kind in die entsprechende multizentrische Therapiestudie auf und behandelt es nach den Vorgaben des Studienprotokolls. Der Leiter der Studie wirkt bei der Therapie immer wieder beratend mit. Das behandelnde Krankenhaus überträgt die anfallenden Studiendaten zur Studienzentrale und zusätzlich einen Basisdatensatz an das Deutsche Kinderkrebsregister. Gewebe- und Blutproben werden an eine studienspezifische Biobank oder ein zentrales Tumorregister weitergeleitet; in vielen Fällen werden auch Bilddaten zu Referenzzwecken in einer zentralen Bilddatenbank gespeichert.

Die klinischen Studien, in deren Rahmen die Behandlungen stattfinden, dienen dem Ziel, die Behandlungspläne zu optimieren, und tragen so fast unmittelbar zur Verbesserung der Krankenversorgung bei. Die Daten des Krebsregisters dienen dagegen eher langfristigen Fragestellungen; mit ihnen werden etwa Studien zu Spätfolgen, Zweiterkrankungen und Lebensqualität durchgeführt, deren Ergebnisse sich erst in der Zukunft auf die Versorgung auswirken können.

Die Situation im Kompetenznetz AHF [AHF09] (angeborene Herzfehler) ist in groben Zügen ähnlich; in diesem Netz gibt es allerdings eine gemeinsame zentrale Studiendatenbank anstatt getrennter Datenbanken für die einzelnen Studien sowie eine Forschungsdatenbank, wo die gesamten Studiendaten langfristig aufbewahrt werden, zusätzlich zu einem Register, das eine lebenslange Patientenakte von kleinem Umfang, aber nicht die detaillierten Forschungsdaten vorhält und die Behandlungskontinuität für den einzelnen Patienten sicherstellen soll.

Wichtig zur Beurteilung der Wirksamkeit von Datenschutzmaßnahmen ist die „Kompartmentisierung" der Forschung: Die Netze sind krankheitsspezifisch; eine gemeinsame Datenhaltung oder ein netzübergreifender Informationsaustausch ist nur in wenigen, besonders begründeten Fällen vorgesehen (bei verwandten Erkrankungen oder Folgeerkrankungen). Ins einzelne Netz werden oft nur wenige 100, maximal bis zu 50000 Patienten aufgenommen.

1.3 Die Architektur eines medizinischen Forschungsnetzes

Ein typisches medizinisches Forschungsnetz enthält einige oder alle der folgenden Komponenten, manchmal sogar in mehreren Instanzen:

- Patientendatenbank für die unmittelbare Versorgung,
- klinische Datenbank für einrichtungsübergreifende Zusammenführung und Auswertung von Behandlungsdaten (im Sinne von Beobachtungsstudien),
- Studiendatenbank für das zentrale Datenmanagement einer oder mehrerer klinischer Studien nach den Regularien des Arzneimittelgesetzes,
- Forschungsdatenbank für die langfristige Sammlung von Forschungsdaten,
- Register mit oder ohne direkten Versorgungsbezug,
- medizinische Bilddatenbank,
- Biobank,
- elektronisches Archiv für die dauerhafte Aufbewahrung von Daten, die nicht mehr aktuell für Behandlungs- oder Forschungszwecke benötigt werden, die aber nach den Regeln guter wissenschaftlicher Praxis aufgehoben werden müssen, um Forschungsergebnisse nachvollziehbar zu dokumentieren.

Zwischen all diesen Datenbanken existieren vielfältige Datenflüsse; daher kann ein Forschungsnetz eine ziemlich komplizierte Architektur haben.

Die Prozesse in einem solchen Netz zu modellieren, ist eine anspruchsvolle Aufgabe und beginnt mit der Spezifikation von Anwendungsfällen (use cases), von denen jeder eine externe Sicht auf eine der Datenbanken mit ihren Verbindungen und Randbedingungen wiedergibt. Ein Datenschutzkonzept muss aber auch eine interne Sicht auf die Details der unterliegenden Prozesse berücksichtigen, um etwa mögliche Datenlecks zu erkennen.

1.4 Die Telematikplattform für medizinische Forschungsnetze (TMF)

In der TMF [TMF09] arbeiten medizinische Forschungsnetze und -projekte zusammen, um gemeinsame Anforderungen an die Infrastruktur zu identifizieren und technische, rechtliche und organisatorische Probleme an der Schnittstelle zwischen Versorgung und Forschung zu lösen, Standards und Terminologien zu entwickeln, rechtlichen und ethischen Rahmenbedingungen gerecht zu werden, Qualitätsmanagement und Technikfolgenabschätzung auf gleichmäßig hohem Niveau zu betreiben und die Öffentlichkeitsarbeit zu fördern.

TMF-Mitglieder sind die Kompetenznetze der Medizin, die Koordinierungszentren für klinische Studien (KKS), Netze seltener Erkrankungen, infektionsepidemiologische Netze, das nationale Genomforschungsnetz und verschiedene andere vernetzte medizinische Forschungsprojekte und -organisationen.

2 Datenschutz in Krankenversorgung und medizinischer Forschung

2.1 Einige Prinzipien

Patientendaten und Biomaterialien (z. B. Blutproben) gehören zu den sensibelsten persönlichen Informationen und müssen sorgfältig geschützt werden gemäß den ethischen Regeln und der ärztlichen Schweigepflicht sowie dem nationalen und internationalen Datenschutzrecht.

Soweit es sich um direkte Krankenversorgung handelt, besteht zunächst ein Behandlungskontext, der sich aus einem (meist impliziten) Behandlungsvertrag herleitet. Hier ist der Patient mit Namen bekannt und erwartet auch, persönlich mit diesem angesprochen zu werden. Die Vertraulichkeit ist in diesem Bereich in erster Linie durch die ärztliche Schweigepflicht geregelt und durch das Strafrecht abgesichert; die Datenschutzgesetze spielen hier nur eine subsidiäre Rolle [PB97].

Sobald die Daten (oder Proben) des Patienten sekundär verwertet werden, beginnt der Forschungskontext. Dieser ist vom Behandlungskontext strikt abzugrenzen [PR04]. Eine solche sekundäre Verwertung von Daten kann ein Forschungsprojekt im eigentlichen Sinn sein, es kann sich aber auch um einfache Auswertungen ökonomischer Natur handeln. Typische Aspekte dieser Datenverwertung sind:

- Die Daten verlassen den Behandlungskontext und sind nicht mehr durch die ärztliche Schweigepflicht, sondern nur noch durch die allgemeinen Datenschutzregelungen geschützt.
- Die Identität des Patienten ist irrelevant; es besteht auch kein direkter Kontakt zu ihm.

In einem solchen Kontext ist die Nutzung anonymisierter Daten – so sie denn wirksam anonymisiert sind – unbedenklich und keiner datenschutzrechtlichen Einschränkung unterworfen. Daher ist Anonymisierung die erste Wahl für die Datenauswertung.

Für die hier beschriebenen Forschungsnetze sind anonyme Daten aber meistens nutzlos: Hier müssen Daten zum selben Patienten aus verschiedenen Quellen und aus verschiedenen Zeiträumen korrekt zugeordnet werden können. In manchen Szenarien ist auch der Weg zurück zur konkreten Person erforderlich: Es könnte im Interesse des Patienten liegen, vielleicht sogar lebenswichtig für ihn sein, Ergebnisse eines Forschungsvorhabens, etwa eine dabei entdeckte genetische Disposition, zurückgemeldet zu bekommen; oder ein Forschungsprojekt benötigt die vorhandene Datensammlung, um geeignete Fälle für eine neue klinische oder epidemiologische Studie ausfindig zu machen.

2.2 Methoden und Werkzeuge zum Schutz von Daten

Die Nutzung von Daten aus der Krankenversorgung für die medizinische Forschung unter Wahrung der Rechte der Patienten, ethischer Prinzipien und Datenschutzregelungen erfordert sorgfältige Anwendung geeigneter Methoden und Werkzeuge wie

- informationelle Gewaltenteilung mit Aufteilung von Pflichten und Verantwortlichkeit, z. B. die getrennte Administration verschiedener Datenbanken, mit Zwischenschaltung von Datentreuhändern (Trusted Third Parties, TTP),
- Pseudonymisierung zumindest am Übergang zwischen Krankenversorgung und Forschung, oft sogar mehrstufige Pseudonymisierung

Dazu müssen organisatorische Maßnahmen kommen wie Checklisten und Vorlagen für die Aufklärung und Einwilligung der Patienten [Ha06], Policies und Verträge, die die Pflichten der Mitarbeiter und externen Auftragnehmer festlegen. Entscheidend aus informatischer Sicht ist, dass die Einhaltung der rechtlichen Anforderungen soweit möglich und sinnvoll durch technische Mittel erzwungen wird; verbleibende Restrisiken müssen transparent sein und durch rechtlich bindende organisatorische Rahmenbedingungen kontrolliert werden.

2.3 Das generische Datenschutzkonzept

Das generische Datenschutzkonzept der TMF zeigt Wege, wie man erfolgreich Forschung mit medizinischen Daten von Patienten (und natürlich auch gesunden Studienteilnehmern) betreiben und dabei den datenschutzrechtlichen Rahmenbedingungen gerecht werden kann, indem man die in 2.2 aufgezählten Methoden und Werkzeuge geeignet einsetzt.

Die erste Version des Konzepts [Re06] beschrieb zwei unterschiedliche Netztypen A und B, die auf zwei typische Netze der ersten Generation zugeschnitten waren und inzwischen von einer Reihe weiterer Projekte zum Aufbau langfristiger Datensammlungen adaptiert wurden. Die beiden Modelle unterscheiden sich in der Art der Datenspeicherung, des Zugriffs und des Pseudonymisierungsprozesses. Gespeichert wird in beiden Fällen pseudonym in einer zentralen Datenbank.

Charakteristisch für das Modell A ist, dass behandelnde Ärzte zur Dateneingabe einen – dann notwendigerweise personenbezogenen – Zugriff auf die Daten ihrer eigenen Patienten haben, wobei der Personenbezug dynamisch über eine Identitätsmanagement-Komponente namens „Patientenliste“ hergestellt wird, siehe Abbildung 1. Diese Patientenliste ist bei einer TTP angesiedelt.

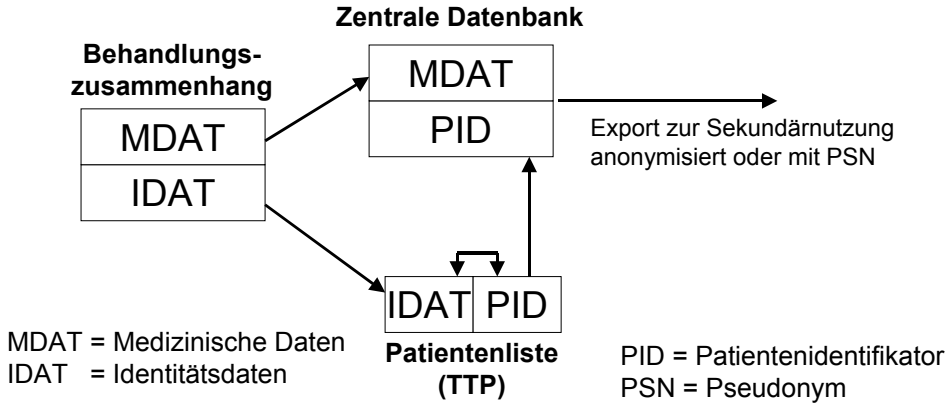

Abbildung 1: Modell A: Pseudonyme Speicherung, personenbezogener Zugriff aus dem Behandlungszusammenhang heraus.

Der „Patientenidentifikator“ PID ist ein *pseudonymes* Kennzeichen, das nur in der Patientenliste und der zentralen Datenbank bekannt ist. Für den Zugriff auf die Daten eines konkreten Patienten wird (nach positiver Rechteprüfung) von der Patientenliste ein temporäres Zugriffsticket an den Zugreifenden und die zentrale Datenbank ausgegeben.

Bei Modell B dagegen ist kein Zugriff aus dem Behandlungskontext heraus auf die Datenbank vorgesehen; eine analoge Patientenliste ist hier Teil des Qualitätsmanagements und dient der richtigen Zuordnung von Patientendaten vor einem gesonderten, unabhängigen Pseudonymisierungsschritt, der die Daten erst in die Datenbank überführt und bei einer weiteren TTP angesiedelt ist, siehe Abbildung 2. Nach der Pseudonymisierung würden Fehler in den Identitätsdaten – z. B. falsch geschriebene Namen – sonst nämlich notwendigerweise zu Fehlzuordnungen führen. Die medizinischen Daten werden dabei, damit die TTP diese nicht lesen kann, asymmetrisch verschlüsselt durch den Pseudonymisierungsprozess durchgereicht und erst in der Datenbank wieder entschlüsselt.

In beiden Modellen werden für Forschungsprojekte aus der zentralen Datenbank nur anonymisierte oder pseudonymisierte Daten bereitgestellt; um einen Abgleich von Daten aus verschiedenen Exportvorgängen zu vermeiden, wird beim Export noch ein – jeweils unterschiedlicher – weiterer Pseudonymisierungsschritt ausgeführt. Zusätzlich ist das Rückidentifizierungsrisiko jedes exportierten Datensatzes im Einzelfall zu prüfen: unter Umständen ist die Vergröberung oder gar Löschung einiger Daten notwendig [PW85].

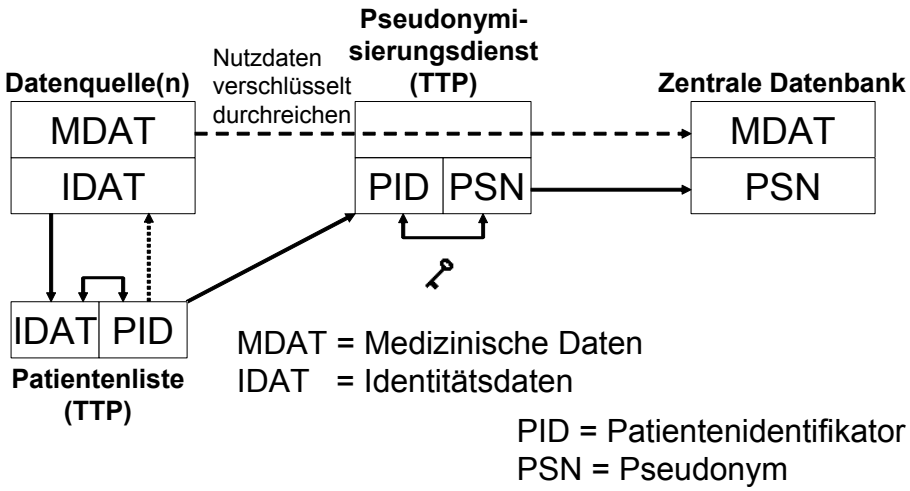

Abbildung 2: Modell B: Pseudonymisierte Übertragung an eine Forschungsdatenbank.

Dieses Konzept wurde in enger Zusammenarbeit mit dem Arbeitskreis Wissenschaft der Datenschutzbeauftragten des Bundes und der Länder entwickelt und fand dessen Konsens.

2.4 Pseudonymisierungsschritte

Die Zuordnung zwischen Klartext-Identität und Pseudonym kann grundsätzlich auf zwei Weisen geschehen: durch eine Zuordnungsliste oder durch einen kryptographischen Algorithmus. Die Patientenliste verwendet die Zuordnungsliste; dies hat bei Modell A den Sinn, den Zugriff auch für verschiedene Berechtigte zu ermöglichen. Bei Modell B dient es der korrekten Zusammenführung fehlerhafter Daten vor dem folgenden Pseudonymisierungsschritt. Dieser erfolgt dann mit einem kryptographischen Algorithmus bei einer TTP, die als zu schützendes Geheimnis nur einen kryptographischen Schlüssel hat.

Der nochmalige Pseudonymisierungsschritt beim Export von Daten zu Auswertungsprojekten dient dazu, dass nicht Daten aus verschiedenen Exportvorgängen über das gleiche Pseudonym zu einer „Schattendatenbank“ zusammengeführt werden können. Das Pseudonym der zentralen Datenbank ist also nur dort und in einer Identätsmanagement-Komponente bekannt.

2.5 Die Rolle der Kryptographie

Kryptographische Verfahren kommen in diesem Konzept an verschiedenen Stellen vor:

- Die Datenflüsse im Netz werden grundsätzlich verschlüsselt; hierbei werden Standard-Protokolle, in der Regel TLS/SSL, verwendet.

- Bei der Pseudonymisierung werden medizinische Daten asymmetrisch verschlüsselt durch den Pseudonymisierungsdienst durchgereicht.
- Der Pseudonymisierungsdienst verwendet ein starkes symmetrisches Verfahren (3DES oder AES) zur Umwandlung zwischen PID und PSN, dessen Schlüssel nur auf einer Smartcard gespeichert ist.
- Daneben ist auf Datenbankservern die verschlüsselte Speicherung als eine der Maßnahmen zur Serverhärtung vorgesehen.

Die kryptographischen Transformationen belasten die Datenprozesse im Netz nicht, wohl aber bedeutet die – durch die informationelle Gewaltenteilung erzwungene – verteilte Netzarchitektur eine erhebliche Belastung. Allerdings fällt diese mehr durch den Implementationsaufwand als durch die Bandbreite der Datenflüsse ins Gewicht.

2.6 Biobanken für die medizinische Forschung

Der Umgang mit biologischen Proben hat einige Ähnlichkeit mit der Prozessierung von medizinischen Daten; es gibt aber auch einige unterschiedliche Aspekte, insbesondere die Existenz der vollen genetischen Information in jedem noch so kleinen Teil einer Probe. Man kann daher nicht davon ausgehen, dass Proben und biologische Materialien auf Dauer anonymisierbar sind. Dies muss durch besonders strikte Nutzungsbegrenzungen ausgeglichen werden und erfordert die physische Trennung von Material und zugehörigen Daten sowie getrennte Datenbanken für aus den Proben gewonnene Analysedaten. Die TMF ergänzte das generische Datenschutzkonzept durch ein Datenschutzkonzept für Biomaterialbanken [Po09] und entwickelte dazu ein modulares und skalierbares Konzept, das viele unterschiedliche Varianten von Biobanken abdeckt und ebenfalls von den Datenschutzbeauftragten konsentiert wurde.

2.7 Die Interessen der Betroffenen

Eine direkte Kontrolle der Betroffenen, Patienten oder auch gesunden Kontrollpersonen, über die Datenverarbeitung wäre kontraproduktiv und nicht durchführbar; insbesondere ist die Verwendung von inhaberkontrollierten Pseudonymen im Sinne von Chaum [Ch85] in diesem Anwendungsfeld nicht möglich. Umso wichtiger ist eine gründliche Aufklärung, die alle geplanten oder für die Zukunft zu erwartenden Verwendungsmöglichkeiten nennt, sowie, insbesondere bei Biomaterialien, auch in der Zukunft jederzeitige Transparenz und Nachvollziehbarkeit mitsamt der Möglichkeit – sofern das noch sinnvoll ist –, die Einwilligung zurückzuziehen. In der Regel wird aber eine hohe Zustimmung der Betroffenen beobachtet [PDK08], die ja meist ein unmittelbares Interesse an den Ergebnissen der medizinischen Forschung haben.

Trotz aller Pseudonymisierungsmaßnahmen bleibt stets ein Restrisiko der Rückidentifizierung von Betroffenen. Darauf muss in der Patientenaufklärung hingewiesen werden; es ist aber auch wichtig, dass die Patienten einer transparenten Organisation mit nachvollziehbar organisierten Prozessen und klaren Verantwortlichkeiten und Ansprechpartnern gegenüber stehen, zu denen sie Vertrauen entwickeln können.

3 Die Evaluation des generischen Datenschutzkonzepts

3.1 Erfahrungen

Mehrere medizinische Forschungsnetze – bisher über 25 – haben ihr konkretes Datenschutzkonzept aus dem generischen TMF-Konzept abgeleitet. Die dabei gemachten Erfahrungen wurden von der TMF systematisch ausgewertet. Dabei wurden Informationen über die Spezifika des jeweiligen Konzepts, über die Begutachtung durch die zuständigen Datenschutzbeauftragten, über organisatorische und technische Umsetzung sowie Wünsche und Vorschläge für Verbesserungen gesammelt.

Neben vielen positiven Rückmeldungen waren die wichtigsten Erkenntnisse auf der negativen Seite: Die Ableitung eines spezifischen Datenschutzkonzepts aus dem generischen ist wegen der notwendigen Anpassung an eines der vorgegebenen Modelle oft schwierig, und die Implementation ist mühevoll. Manche konkreten Anforderungen hatten keine Entsprechung in diesen Modellen; das betraf vor allem Netze mit vielen multizentrischen klinischen Studien oder mit einer engen Verzahnung zwischen Versorgung und Forschung sowie internationale Kooperationen.

3.2 Anforderungen an eine Revision

Viele Projekte benötigen ein Zusammenspiel beider generischen Typen A und B: den Typ A etwa für die Begleitung chronisch kranker Patienten über einen langen Zeitraum oder in etwas abgewandelter Form für das zentrale Datenmanagement von klinischen Studien oder für Prozesse zur Qualitätssicherung von Daten, den Typ B zum Aufbau langfristiger Datenpools oder von epidemiologischen Registern und schließlich zusätzlich die Erweiterung für Biobanken. Daraus wurde die folgende Liste von Anforderungen für eine Revision des generischen Datenschutzkonzepts hergeleitet:

- umfangreiche Spezifikation von Anwendungsfällen (use cases) und Geschäftsprozessen,
- Ersatz der Dichotomie „A oder B“ durch einen modularen Ansatz, in dem Datenbanken im direkten Behandlungskontext von Datenbanken im Forschungskontext mit und ohne Bezug zur Behandlung unterschieden werden, aber koexistieren und kommunizieren können,
- Skalierbarkeit der Maßnahmen anhand eines Kriterienkatalogs zur Verhältnismäßigkeit, der insbesondere Anhaltspunkte zur Beurteilung des Rückidentifizierungsrisikos bietet,
- Integration von Versorgungs- und Forschungsstrukturen,
- bessere Berücksichtigung von Qualitätssicherungsprozessen,
- explizite Integration von multizentrischen klinischen Studien, Registern und Bilddatenbanken,
- Vorschläge und Muster für die Etablierung von zentralen Dienstleistungen, die die Architektur der einzelnen Netze entlasten können.

Darüber hinaus wurde eine Anzahl von offenen rechtlichen Fragen identifiziert, die noch einmal eine gründliche Analyse der rechtlichen Situation im Grenzbereich zwischen Versorgung und Forschung, insbesondere vor dem Hintergrund des Ausbaus der Gesundheitstelematik erforderlich machten.

3.3 Rechtliche Rahmenbedingungen

Für die noch offenen rechtlichen Fragen wurden Rechtsgutachten von führenden Medizinrechtlern eingeholt. Die Fragen wurden im Wesentlichen in die Komplexe

- Nutzung bereits bestehender Datenbestände für Forschungszwecke,
- Mitnutzung der „Gesundheitstelematik-Infrastruktur“ im Forschungskontext,
- Pseudonymisierungspflicht bei Vermischung von Behandlungs- und Forschungskontext, z. B. in wissenschaftsgetriebenen therapeutischen Studien,
- Anforderungen an eine (elektronische) Datentreuhänderschaft

gegliedert.

Ein wesentliches Ergebnis der Gutachten ist, dass die Bereiche Versorgung und Forschung bezüglich der Datenhaltung rechtlich zwingend abzugrenzen sind. Die Nutzung der elektronischen Gesundheitskarte (eGK) gemäß SGB V §291a für medizinische Forschungszwecke ist vom Gesetzgeber explizit ausgeschlossen – das ist auch nicht durch eine Einwilligung umgehbar. Lediglich technische Aspekte der Gesundheitstelematik-Infrastruktur werden für die Forschung nutzbar sein. Hierzu gehört z. B. die Versicherten-Nummer auf der eGK, die als Teil der identifizierenden Daten beim pseudonymen Identitätsmanagement für Patienten verwendet werden kann, sowie der Heilberufeausweis samt Verzeichnisdienst zur Zugriffssteuerung auf Forschungsdaten. Wegen unterschiedlicher rechtlicher Rahmenbedingungen sollte zweckmäßigerweise auch noch der Kontext von klinischen Studien von sonstiger Forschung abgegrenzt werden; bei letzterem Bereich sollte man außerdem zusätzlich patientennahe von patientenferner Forschung trennen, was der Einteilung in die verschiedenen Modelle des „alten“ generischen Datenschutzkonzepts entspricht. Ein Datenexport an Dritte über die Einwilligung hinaus ist auch in pseudonymisierter oder anonymisierter Form problematisch, da das Zusatzwissen der empfangenden Stelle und somit das Rückidentifizierungsrisiko kaum einzuschätzen ist. Daher sind auch „public-use“-Datenbanken nur zulässig, wenn die Daten zuvor hinreichend, und das bedeutet in der Regel sehr stark, vergröbert wurden.

In Bezug auf die Datentreuhänderschaft ergaben die Gutachten, dass die bisher oft angenommene Beschlagnahmesicherheit, wenn ein Notar diese Funktion ausübt, illusorisch ist. Es spricht also nichts dagegen, Treuhänderdienste etwa in Rechenzentren von Universitätskliniken anzusiedeln, sofern diese keine anderen Interessen im Forschungsnetz vertreten.

Projekte zur Klärung der internationalen Situation wurden gestartet, abschließende Ergebnisse liegen noch nicht vor.

3.4 Das revidierte generische Modell

Die Umsetzung in ein organisatorisches Konzept erfordert für ein „Maximalnetz“ die Abgrenzung von mindestens vier Bereichen, repräsentiert durch je (mindestens) eine Datenbank: Versorgungsdatenbank für die patientennahe Forschung, Studiendatenbank für klinische Studien, Forschungsdatenbank für die patientenferne Forschung (z. B. Register), Biomaterialbank; oft kommt noch eine Bilddatenbank hinzu. In jedem dieser Bereiche ist ein unterschiedliches Pseudonymisierungsschema zu verwenden. Dazu braucht man als zentrale Dienste im Netz u. a. das Identitätsmanagement für Patienten in Form der Patientenliste und den Pseudonymisierungsdienst, wie in den bisherigen Modellen A und B vorgesehen. Diese können treuhänderisch bei einem vom Netz beauftragten Partner, z. B. bei einer Universitätsklinik, angesiedelt werden, müssen aber disziplinarisch unabhängig von den Datenbanken betrieben werden und einer rechtlich bindenden Nutzungsordnung unterliegen.

Die zurzeit stattfindende Überarbeitung des generischen Datenschutzkonzepts wird diese einheitliche umfassende Systemarchitektur beschreiben, siehe Abbildung 3. Diese Architektur ist modular: Die verschiedenen Arten von Datenbanken werden in entsprechende Module gekapselt. Diese unterscheiden sich nach Rahmenanforderungen, Speicherart und Zugriffsart und verwenden jeweils ihre eigenen Pseudonyme.

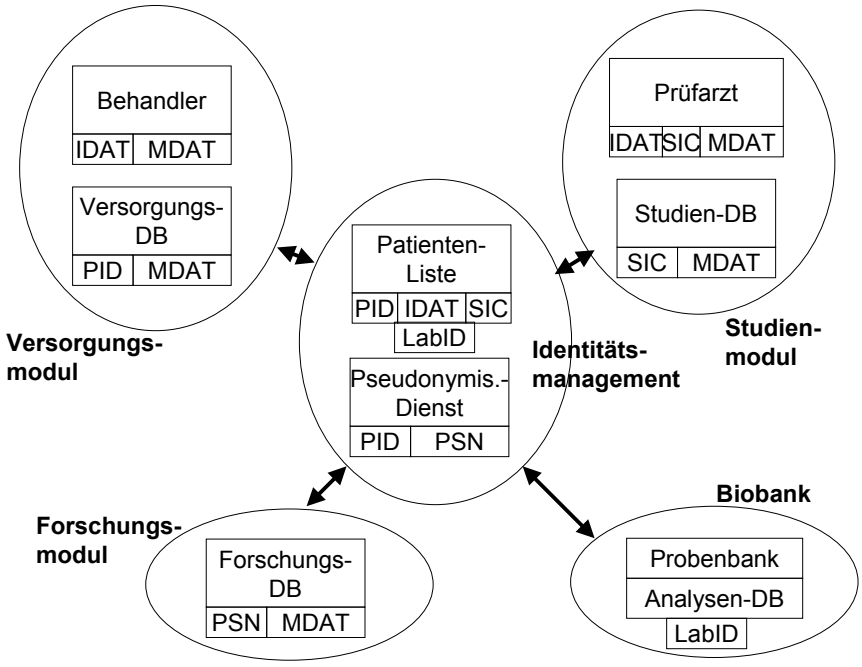

Abbildung 3: Referenzmodell im künftigen revidierten Datenschutzkonzept der TMF: Datenbanken und Daten (vereinfacht).

Zentral zwischen allen steht ein, möglicherweise auch auf mehrere TTPs verteilter, Identitätsmanagement-Modul, der bei Bedarf den Übergang zwischen diesen verschiedenen Pseudonymen und auch der wahren Identität herstellt.

Die Versorgungs-Datenbank (VDB) enthält Daten, die für die Versorgung des Patienten direkt relevant sind; sie steht im unmittelbaren Behandlungskontext, ist aber einrichtungsübergreifend und wird daher pseudonym (PID) geführt. Behandler haben einen personenbezogenen Zugriff auf die Daten ihrer Patienten (lesend und schreibend); Zugriffe auf fremde Patienten gibt es nicht. Der Zugriff geschieht mit Hilfe der Patientenliste (ID-Management für Patienten) und mit Hilfe eines Verzeichnisdienstes (ID-Management für Benutzer). Dieser Modul ist eine Weiterentwicklung aus dem TMF-Modell A. Er ist bei jeder Art von einrichtungsübergreifender Versorgung relevant; im Kontext der medizinischen Forschung wird er dort benötigt, wo langfristige Beobachtungsstudien durchgeführt werden, etwa bei seltenen oder chronischen Erkrankungen.

Die Studien-Datenbank (SDB) dient zur Durchführung klinischer Studien nach den Regularien des AMG und der guten klinischen Praxis (GCP). Sie enthält Daten zum Patienten, die für die Studie relevant sind. Die Überschneidung mit den Daten der reinen Versorgungsdokumentation ist groß. Die SDB steht im unmittelbaren Behandlungskontext, soweit es um Zugriffe durch den Prüfarzt geht; sie steht im Forschungskontext, wenn es um Zugriffe durch den „Sponsor“ oder Studienleiter geht. Sie ist, zumindest bei multizentrischen Studien, einrichtungsübergreifend. Sie wird konform zum AMG pseudonym (SIC = Subject Identification Code) geführt. Prüfärzte haben einen personenbezogenen Zugriff auf die Daten ihrer Patienten (lesend und schreibend) und kennen den SIC.

Die Forschungs-Datenbank (FDB) dient zur Langzeitspeicherung pseudonymisierter medizinischer Daten für spätere Forschungsprojekte – direkt zur epidemiologischen Forschung, indirekt zur Rekrutierung geeigneter Fälle für neue klinische oder epidemiologische Forschung. Sie bietet den nochmals pseudonymisierten Export geeigneter Daten und stellt eine Weiterentwicklung des TMF-Modells B dar.

3.5 Weiteres Vorgehen

Die Abstimmung des revidierten Konzepts mit den Datenschutzbeauftragten soll noch in diesem Jahr (2009) stattfinden, die anschließende Buchpublikation ist vorgesehen.

4 Schlussfolgerungen

Das generische Datenschutzkonzept der TMF, insbesondere in seiner revidierten Version, ist durch seine Modularität eine flexible und skalierbare Grundlage für die Vernetzung und Integration von Krankenversorgung und medizinischer Forschung, unterstützt die langfristige Datenakkumulation in beiden Bereichen und ermöglicht ein breites Spektrum der Nutzung von Patientendaten und biologischen Proben.

Ein medizinisches Netz oder Forschungsprojekt kann das generische Konzept als Vorlage benutzen, die nötigen Komponenten herauspicken und die generische Architektur nach seinen speziellen Erfordernissen modifizieren und gegebenenfalls im Hinblick auf Kriterien zur Verhältnismäßigkeit vereinfachen. Die TMF unterstützt ein solches Vorhaben bis hin zur Begutachtung durch die zuständige Datenschutzbehörde.

Die Rolle der Informatik entspricht dabei dem oft in interdisziplinären Projekten beobachteten Muster: Die Werkzeuge aus der methodischen Grundlagenwissenschaft sind unverzichtbar, decken aber die Anforderungen aus dem Anwendungsbereich nur teilweise ab; außerdem werden sie nur in recht einfacher Ausprägung gebraucht. Da die technischen Möglichkeiten zur Sicherung der Vertraulichkeit limitiert sind, müssen sie durch weitere Maßnahmen im organisatorischen Bereich ergänzt werden. Insbesondere erfordert das – auch durch noch so gute Pseudonymisierungsverfahren verbleibende – Restrisiko einer Rückidentifizierung Betroffener, dass auch der Zugang zu pseudonymisierten Daten unter Kontrolle bleibt.

Eine noch zu lösende Aufgabe für die Informatik ist der Aufbau einer Sicherheitsarchitektur, die die Einhaltung von Policies erzwingt.

Danksagung

Diese Arbeit wurde vom Bundesministerium für Bildung und Forschung als ein Projekt der Telematikplattform – Verbund zur Förderung vernetzter medizinischer Forschung (TMF) e.V. – gefördert.

Der Erstautor ist im Rahmen der TMF als Vertreter des KPOH (Kompetenznetz für die Pädiatrische Onkologie und Hämatologie) tätig. Die weiteren Autoren vertreten die Kompetenznetze „Angeborene Herzfehler“ (Fördernummer 01 GI 0201), „Leukämien“, „Hepatitis“, „Maligne Lymphome“, das KKS Leipzig, und das Netz „Epidermolysis Bullosa“.

Literaturverzeichnis

[AHF09] Kompetenznetz Angeborene Herzfehler. Online: http://www.kompetenznetz-ahf.de/ (zuletzt besucht 23. April 2009).

[Ch85] Chaum, D. Security without identification: Transaction systems to make Big Brother obsolete. Communications of the ACM 28, 1985, 1030–1045.

[Ha06] Harnischmacher, U. et al.: Checkliste und Leitfaden zur Patienteneinwilligung. MWV, München, 2006.

[PB97] Pommerening, K.; Blobel, B.: Datenschutz und Datensicherheit in Informationssystemen des Gesundheitswesens. Führen & Wirtschaften im Krankenhaus 2/1997, 133–138.

[PDK08] Pommerening, K.; Debling, D.; Kaatsch, P.; Blettner, M.: Register zu seltenen Krankheiten – Patientencompliance und Datenschutz. Bundesgesundheitsblatt 5/2008, 491–499.

[Po09] Pommerening, K. et al.: Ein generisches Datenschutzkonzept für Biomaterialbanken. MWV, München, 2009 (im Druck).

[POH09] Kompetenznetz Pädiatrische Onkologie und Hämatologie. Online: http://www.kinderkrebsinfo.de/ (zuletzt besucht 23. April 2009).

[PR04] Pommerening, K.; Reng, C.-M.: Secondary use of the Electronic Health Record via pseudonymisation. In (Bos, L. et al. Hrsg.). Medical Care Compunetics I. IOS Press, Amsterdam, 2004; S. 441–446.

[PW89] Paaß, G.; Wauschkuhn, U.: Datenzugang, Datenschutz und Anonymisierung. Oldenbourg, München, 1985.

[Re06] Reng, C.-M. et al.: Generische Lösungen zum Datenschutz für die Forschungsnetze in der Medizin. MWV, München, 2006.

[TMF09] Telematikplattform – Verbund zur Förderung vernetzter medizinischer Forschung (TMF) e. V. Online: http://www.tmf-ev.de/ (zuletzt besucht 23. April 2009).