Konzept für ein deutschlandweites Krankheitsnetz am Beispiel von mitoREGISTER

F.M. Kohlmayer [1,2], R.R. Lautenschläger [1], S.H.R Wurst [1], T. Klopstock[3], H. Prokisch[4,5], T. Meitinger[4,5], C. Eckert[2], K.A. Kuhn[1]

[1] Lehrstuhl für medizinische Informatik, Technische Universität München

[2] Lehrstuhl für IT-Sicherheit, Technische Universität München

[3] Friedrich-Baur-Institut, Ludwigs-Maximilian-Universität München

[4] Institut für Humangenetik. Helmholtz Zentrum München

[5] Lehrstuhl für Humangenetik, Technische Universität München

florian.kohlmayer@tum.de

Abstract: Diese Arbeit beschreibt die Architektur eines IT-Systems zur Erfassung und Verwaltung von Patientendaten und Bioproben in einem deutschlandweiten Netzwerk, das im Rahmen des Verbundprojektes mitoNET vom BMBF gefördert wird. Aufgrund der sensiblen Daten liegt ein besonderer Schwerpunkt auf Datenschutz und IT-Sicherheitsaspekten; u.a. werden kryptographische Methoden, Zugriffstokens und getrennte Repositories verwendet. Die Konzepte und das System sind für ähnliche Netzwerke einsetzbar.

1 Einleitung

Nach der Entschlüsselung des menschlichen Genoms hat in der medizinischen Forschung die Verknüpfung genetischer Daten (Genotyp) mit klinischem Erscheinungsbild (Phänotyp) und Verlaufsinformation massiv an Bedeutung gewonnen. Die Analyse von Phänotyp-Genotyp-Beziehungen verbessert das Verständnis von Krankheiten und unterstützt personalisierte diagnostische und therapeutische Ansätze in der Medizin. Detailliert und im Verlauf erfasste medizinische Daten, ausreichende Probanden- bzw. Fallzahlen, sowie annotierte Bioproben sind für diese Analysen entscheidend. Die Erforschung seltener Krankheiten bedarf aufgrund geringer Fallzahlen besonderer Aufmerksamkeit und einer vernetzten Zusammenarbeit, um zu gewährleisten dass möglichst viele Patienten nach definierten Ein- und Ausschlusskriterien erfasst werden. Das hier vorgestellte mitoREGISTER dient der strukturierten longitudinalen Erfassung von Patientendaten sowie zur Bioprobenverwaltung mit dem Ziel einer Erforschung mitochondrialer Erkrankungen.

Ein wesentliches Ziel dieses Projektes liegt darin, unter Wahrung aller datenschutzrechtlichen Vorgaben für die behandelnden und im Projekt mitarbeitenden Zentren (Kliniken und Praxen) ein Register [LG06] aufzubauen, das eine zentrumsübergreifende Erhebung von Patientendaten ermöglicht. Bei jedem Besuch des Patienten werden Bioproben zentral gesammelt und verwaltet. Um die Vergleichbarkeit der Untersuchungen, der Materialen und der Ergebnisse zu gewährleisten, sind standardisierte Terminologien und Standard Operating Procedures (SOPs) erforderlich. Wegen der Langzeiterfassung muss der Fall des Zentrumswechsels eines Patienten berücksichtigt werden. Aufgrund der Komplexität und Heterogenität einerseits (Einbindung existierender Terminologien und Fragebögen, verschiedene Systeme für Datenhaltung und Bioprobenverwaltung, Übernahme von Daten aus Fremdsystemen) und der Schutzwürdigkeit der Daten andererseits (um-

fassend annotierte Bioproben) gibt es derzeit keine allgemein verfügbare oder gar „off-the-shelf“ Lösung, die den Anforderungen genügt.
Diese Arbeit beschreibt ein Sicherheitskonzept und eine Architektur (mit den Schwerpunkten Geschäftsprozess und Anwendung) für eine sichere und datenschutzkonforme Sammlung von Patientendaten und Bioproben. Zwei Aspekte waren bei der Entwicklung wesentlich: die Möglichkeit der Anpassung an die Anforderungen anderer nationaler oder internationaler Forschungsnetze und die einfache Umsetzbarkeit.
In Deutschland werden derzeit 16 verschiedene Forschungsverbünde für seltene Erkrankungen vom BMBF finanziert [PD08]. Diese sind von den Kompetenznetzen in der Medizin zu unterscheiden, in denen weit verbreitete Krankheiten (wie etwa Diabetes mellitus) von ausgewählten Kompetenzzentren untersucht werden. Datenbanken zur Erfassung klinischer Phänotypen verbunden mit Bioprobensammlungen werden in zunehmender Zahl für epidemiologische Fragestellungen, in der Translationsforschung und im Zusammenhang mit klinischen Studien eingesetzt [BBMRI]. Dabei spielt auch die sichere Übernahme freigegebener Daten zwischen Systemen eine wesentliche Rolle [eSDI]. Ein generisches Datenschutzkonzept für medizinische Datensammlungen und Biobanken wurde in [PO07] beschrieben; in dieser Arbeit wird eine Weiterentwicklung vorgestellt.

2 Methoden

Die Datenerhebung erfolgt verlaufsbezogen nach Vorliegen einer Einverständniserklärung; dazu werden standardisierte Formulare und kontrollierte Vokabulare [LG06] verwendet. Zudem werden Bioproben entnommen. Zur Eingabe der medizinisch-phänotypischen Daten in mitoREGISTER werden Fragebögen eingesetzt, die von den Ärzten während oder nach einer Untersuchung ausgefüllt werden. Das System verwendet sowohl neuentwickelte als auch bereits etablierte Fragebögen aus [NMDS], [SARA] und [SF12]. Zur Sammlung von Biomaterialen werden Kits („mitoKITs“) von einem zentralen Labor zur Verfügung gestellt. Sie enthalten mit einem Barcode versehene Probenbehälter, die Einverständniserklärung zur Probenentnahme, eine eidesstattliche Versicherung und einen Dokumentationsbogen. Die Einverständniserklärung eines Patienten dient als elementarer, vertraglicher Bestandteil für die Teilnahme an diesem Forschungsvorhaben. Die eidesstattliche Versicherung dokumentiert, dass die Einverständniserklärung des Patienten korrekt ausgefüllt und archiviert wurde. Sie wird vom Arzt unterschrieben und beim Versand der Proben beigelegt. Die im mitoREGISTER umgesetzten Konzepte orientieren sich an den Vorgaben des BDSG zur Datensicherheit in Forschungsnetzen [BDSG] sowie an [PO07]. Für ein IT-System ergeben sich dabei u.a. die folgenden Anforderungen bzw. Schutzziele [EC09]. Ein Webbrowser soll als Client zum Einsatz kommen, da keine Installation zusätzlicher Software in den beteiligten Kliniken erfolgen darf. Dies gewährleistet Plattformunabhängigkeit, geringen Wartungsaufwand und breite Verfügbarkeit. Die Möglichkeit zur Langzeiterfassung von Patientendaten muss gegeben sein. Eine Anbindung eines zentralen Labors mit Lagerung, Verwaltung und Auswertung der Bioproben ist gefordert. Eine Site-based-View, d.h. nur das Behandlungsteam darf auf die direkt identifizierenden Daten wie Name und Anschrift der eigenen Patienten Zugriff haben. Ebenso ist eine Rechteverwaltung mittels RBAC (Role-based Access Control), mit den Rollen: Administrator/in, Labormitarbeiter/in, Arzt/Ärztin, Forscher/in notwendig. Die Authentifizierung am System soll keine

weitere zu installierende Hardware erfordern (Kartenlesegerät, Fingerprint etc.). Es muss eine Trennung der direkt identifizierenden Daten von den medizinischen und den Analysedaten (Pseudonymisierung) erfolgen. Die Vertraulichkeit und Integrität der Patientendaten und Nachvollziehbarkeit der Änderungen im System (Verbindlichkeit) sind sicher zu stellen.

Das System setzt sich aus verschiedenen Teilsystemen zusammen, die alle mit einem Java-Server-Faces-basierten Framework entwickelt wurden. Als Datenbank kommt MySQL mit InnoDB Engine zum Einsatz, als Web-/Applikationsserver wird Tomcat mit Apache als Frontend eingesetzt. Die einzelnen Teilkomponenten sind jeweils mittels einer Firewall geschützt. Zwecks Ausfallsicherheit werden virtuelle Server verwendet. Backups der virtuellen Server wie auch der Daten werden nur verschlüsselt übertragen und gespeichert. Verwendet wurde hierbei ausschließlich Open-Source-Software, wodurch auch in Zukunft keine Lizenzkosten oder Einschränkungen aufgrund proprietärer Technologie zu erwarten sind. Dies ist wichtig für den Fall, dass die Software anderen Arbeitsgruppen bzw. Netzen zur Verfügung gestellt wird. Durch das Model-View-Controller-Architekturmuster ist das System leicht änder- und erweiterbar. Dadurch ist eine größtmögliche Flexibilität gegenüber sich ändernden Anforderungen vorhanden und Systeme mit ähnlichen Anforderungen können mit geringem Aufwand umgesetzt werden. Das System unterstützt den gesamten Prozess der Patientenaufnahme und Patientenidentifikation, der Datenerhebung, der Probensammlung und der Datenauswertung. Jedem Benutzer werden vom Anwendungsadministrator eine Rolle und ein Zentrum zugeordnet. Wichtig ist eine strikte organisatorische und räumliche Trennung zwischen direkt identifizierenden und medizinischen Daten; sie werden auf unterschiedlichen Teilsystemen jeweils zentral gehalten. Zugriffe zwischen den Systemen werden mittels Einmal-Zugriffs-Tokens geschützt. Die direkt identifizierenden Daten der Patienten sind mittels asymmetrischer Kryptographie geschützt. Ein Audittrail auf Datenbankebene protokolliert alle Änderungen des Systems. HTTPS wird für die Kommunikation genutzt. Erste, automatisierte Penetrationstests wurden durchgeführt. Die Anwendung der o.g. Konzepte resultiert in der in Abbildung 1 dargestellten Architektur.

Das Teilsystem Patientenliste speichert die direkt identifizierenden Daten (IdentDaten) der Patienten und ordnet diesen einen eindeutigen Patientenidentifikator (PatID) zu. Dieser Patientenidentifikator ist eine zufällig generierte alphanumerische Zeichenkette und stellt ein Pseudonym dar. Die Aufgabe der ProbIDListe ist es die auf den Probenbehältern sichtbaren eindeutigen Barcode-Identifikatoren Pseudonyme zweiter Stufe zuzuordnen. Der Barcode-Identifikator wird durch den Arzt in der ProbIDListe registriert (und indirekt einem Patienten über die Med-Datenbank zugeordnet), dabei werden eigene systeminterne Probenidentifikatoren (ProbID für Med-Datenbank, ProbID‘ für Analysedatenbank) zugeordnet. Einem Patienten können mehrere Probenbehälter zugeordnet sein. Diese Probenidentifikatoren sind zufällig generierte alphanumerische Zeichenketten und stellen Pseudonyme für den physischen Probenbehälter, analog zum Patientenidentifikator (PatID), dar. Mit Hilfe der ProbIDListe ist die ProbID und ProbID‘ mit dem als Barcode-Label sichtbaren Identifikator der Probe und den Daten der Med-Datenbank und Analysedatenbank verknüpft. Die Aufbewahrung der Bioproben erfolgt im zentralen Bioprobenlager. Die Verwaltung der Bioproben erfolgt mit dem in Abbildung 1 dargestellten Labor-Informations- und Management- System (LIMS) anhand des Barcode-Identifikators auf den Probenbehältern. Lokal werden Daten der Eidesstattlichen Versicherung und Angaben zum Lagerort der Proben gespeichert. Die Med-Datenbank dient

der Speicherung medizinischer Daten. Die Analysedatenbank dient der Speicherung der Analysedaten. Diese Daten sind durch die Pseudonyme ProbID‘, ProbID und PatID den direkt identifizierenden Daten des Patienten in der Patientenliste zugeordnet.

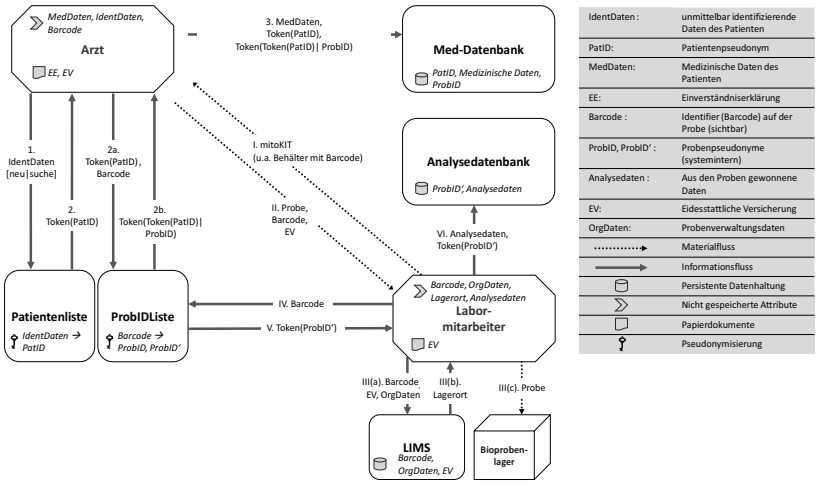

Abbildung 1: Sicherheitsarchitektur

Ein möglicher Zugriff auf das System durch eine(n) Arzt/Ärztin gestaltet sich wie folgt: Nach Authentifizierung des Arztes an der Patientenliste bekommt der Arzt die direkt identifizierenden Daten aller seiner bereits eingegebenen Patienten angezeigt. Der Arzt kann nun einen bereits vorhandenen Patienteneintrag auswählen oder einen neuen Eintrag hinzufügen. Wurden aktuell Bioproben entnommen, kann der Arzt dem Patienten diese mithilfe der ProbIDListe und Med-Datenbank zuordnen. Dies geschieht anhand des auf der Bioprobe stehenden Barcodes (Schritte 1-2b). Zum Einsehen bzw. Editieren medizinischer Daten bekommt der Arzt ein verschlüsseltes Zugriffstoken von der Patientenliste (Schritt 2) mit welchem er automatisch an der Med-Datenbank angemeldet wird (Schritt 3). Der Arzt sieht danach eine Übersicht der in der Med-Datenbank gespeicherten Daten des ausgewählten Patienten, gegliedert nach Behandlungsterminen. Die entnommenen Bioproben werden vom Arzt an das zentrale Labor geschickt (Schritt II). Das Labor kann nun die angeforderten Analysen durchführen. Anschließend werden die Proben eingelagert (Schritte IIIa-c). Die Analysedaten werden in die Analysedatenbank übertragen (Schritte IV-VI). Die einmal gültigen Zugriffstoken für den Zugriff von Patientenliste und ProbIDListe auf Med-Datenbank, sind AES-verschlüsselte und integritätsgeschützte Zeichenketten. Auf den Schlüssel haben jeweils nur die die Patientenliste, ProbIDListe und die Med-Datenbank Zugriff. Der Inhalt des Tokens setzt sich zusammen aus Benutzernamen, Passwort, fortlaufendem Zähler, Patientenpseudonym (PatID) und im Falle einer neu zugeordneten Bioprobe, dem systeminternen Probenidentifikator (ProbID), um die Zuordnung von ProbID und PatID in der Med-Datenbank zu ermöglichen. Der Zähler dient zum Schutz vor Wiedereinspielungsattacken. Durch das verschlüsselte Zugriffstoken ist keine direkte Kommunikation zwischen Patientenliste und Med-Datenbank notwendig.

3 Ergebnisse

Das mitoREGISTER-System wurde am Klinikum rechts der Isar in München entwickelt und befindet sich in einer ersten Ausbaustufe im produktiven Betrieb. Die Patientenliste und ProbIDListe ist organisatorisch wie auch räumlich von der Med-Datenbank getrennt. Die Analysedatenbank und Med-Datenbank sind momentan kombiniert (ProbID und ProbID‘ sind identisch), da nur ausgewählte Analysedaten erhoben werden. Die Patientenliste wird unter Obhut des Datenschutzbeauftragten des Klinikums betrieben, die Med-Datenbank am Institut für medizinische Statistik und Epidemiologie. Im Falle einer Kompromittierung eines der Systeme (Patientenliste, ProbIDListe oder Med-Datenbank) verhindert die pseudonymisierte Speicherung und Trennung der Systeme die Zuordnung von direkt identifizierenden zu medizinischen Daten. Die Lösung mit dem Zugriffstoken ermöglicht eine Integration der Systeme auf Präsentationsebene mithilfe von Standard-HTML-Framesets. Bei der Authentifizierung der Patientenliste an der Med-Datenbank werden in einem weiteren Frame unterhalb der direkt identifizierenden Daten die medizinischen Daten eines Patienten eingeblendet. Hierdurch ist der Zugriff auf die Systeme für den Benutzer transparent. Eine zusätzliche Sicherheitsmaßnahme ist die asymmetrische Verschlüsselung der direkt identifizierenden Patientendaten in der Patientenliste. Jedes Zentrum erhält sein eigenes Schlüsselpaar, und alle Patientendaten eines Zentrums werden mithilfe des öffentlichen Schlüssels eines Zentrums verschlüsselt und können nur durch den privaten Schlüssels des Zentrums entschlüsselt werden (site-based view). Der private Schlüssel eines Zentrums ist mit dem Passwort des Nutzers geschützt und in der Patientenliste gespeichert. Hierdurch wird das Risiko minimiert, dass Patientendaten einsehbar sind, die nicht zum Zentrum des aktuell angemeldeten Benutzers (Arzt/Ärztin) gehören. Der Einsatz asymmetrischer Kryptographie unterstützt auch den Fall eines Zentrumswechslers. In diesem Fall werden die direkt identifizierenden Daten des Patienten mit dem privaten Schlüssel des Zentrums des aktuell angemeldeten Benutzers entschlüsselt und mit dem öffentlichen Schlüssel des neuen Zentrums verschlüsselt. Somit findet eine „Übergabe“ der Daten im System an das neue Zentrum statt.

4 Diskussion und Ausblick

Das präsentierte Konzept stellt eine Erweiterung gegenüber [PO07] dar. Durch das verschlüsselte Zugriffstoken entfällt eine direkte Kommunikation zwischen Patientenliste und Med-Datenbank. Indem der Rechner des Arztes dieses Zugriffstoken weiterleitet, besteht ein logischer, aber vertraulicher und integritätsgeschützter Kanal zwischen Patientenliste und Med-Datenbank, und die Sicherheit wird dadurch gewährleistet, dass der Zugriffstoken-Inhalt auf dem Arztrechner ohne Kenntnis des geheimen Schlüssels nicht ausgelesen werden kann. Eine Veränderung würde durch den Integritätsschutz erkannt und hätte eine Ablehnung durch die Med-Datenbank zur Folge. Durch die site-based view, durch die jeder Nutzer nur Zugriff auf die Patienten seines Zentrums hat, hätte auch eine Kompromittierung eines Benutzerkontos (wie etwa durch Phishing oder Brute-Force) keinen vollständigen Verlust der Vertraulichkeit aller Daten zur Folge. Es wären nur die Daten des betreffenden Zentrums zugreifbar bzw. entschlüsselbar. Auch wird hierdurch eine Wartung durch einen Dritten i.S. des Datenschutzgesetzes (z.B. Systemadministrator) erleichtert, da dieser nur Zugriff auf verschlüsselte Patientendaten hat. Durch die Speicherung personenbezogener Daten in der Patientenliste entsteht ein erhöhter Schutzbedarf. Dem wird u.a. durch asymmetrische Verschlüsselung der direkt identi-

fizierenden Daten der Patienten Rechnung getragen. Wichtig ist, dass eine Ersetzung des Barcode-Identifikators des Probenbehälters im Labor nicht notwendig ist, da die Daten unter einem systeminternen Probenidentifikator (ProbID) in der Med-Datenbank gespeichert werden und die Einverständniserklärung mit den direkt identifizierenden Daten (wie etwa Name, Unterschrift) beim jeweiligen Arzt/Ärztin verbleibt. Damit sind im Labor keine direkt identifizierenden Daten zur Probe bekannt, wodurch ein generelles Problem des Probenversands hinsichtlich des Datenschutzes gelöst wurde. Die Teilsysteme übernehmen Aufgaben wie die Verwaltung von Benutzern, Bioproben, direkt identifizierenden Patientendaten, medizinischen Patientendaten, statistische Erhebung und grafische Aufbereitung. Der Zugriff auf das System bzw. dessen Teilsysteme erfolgt für den Benutzer transparent, da eine vollständige Integration der Präsentationebene realisiert wurde. Dezentrale Labore, analog zu dem hier vorgestellten Zentrallabor, lassen sich einfach in die Architektur integrieren. Auch ein weiterer Pseudonymisierungsschritt zwischen Patientenliste, ProbIDListe und Med-Datenbank ist möglich. Durch die symmetrische Integration von Labordaten und Eingabe der medizinischen Daten können beliebige weitere Datenquellen integriert werden, insbesondere falls eine kontinuierliche Übernahme von freigegebenen klinischen Daten in die Med-Datenbank notwendig wird. Sollten Daten exportiert werden, sind Maßnahmen zur Gewährleistung starker Anonymität (z.B. k-Anonymität) der Daten vorgesehen. Ein großes Sicherheitsrisiko im vorgestellten Konzept bleibt der Rechner des Arztes/Ärztin. Durch die dezentrale und heterogene IT-Landschaft in den beteiligten Zentren kann keine Garantie für die Schadsoftware-Freiheit der Rechner gegeben werden: hier sind die Administratoren der jeweiligen Zentren gefordert. Sicherheitsfortbildungen für Ärzte werden vorgeschlagen, um das Risiko zu minimieren. Auch weitere technische Maßnahmen wie Zwei-Faktor-Authentifizierung der Ärzte sind angedacht. Hierzu sind Machbarkeitsanalysen vorgesehen.

Danksagung Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung, und Forschung unter dem Förderkennzeichen 01GM0862 gefördert und wird unterstützt von der Graduate School of Information Science in Health (GSISH) der TU München. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

5 Literaturverzeichnis

[BBMRI] www.bbmri.eu, letzter Zugriff 30. April 2010.
[BDSG] Bundesdatenschutzgesetz, BDSG 2009.
[EC09] C. Eckert: IT-Sicherheit: Konzepte - Verfahren - Protokolle, Oldenbourg, 2009.
[eSDI] CDISC eSDI Group: Leveraging the CDISC Standards to Facilitate the use of Electronic Source Data within Clinical Trials, 2005.
[LG06] F. Leiner, et al.: Medizinische Dokumentation: Grundlagen einer qualitätsgesicherten integrierten Krankenversorgung. Lehrbuch und Leitfaden, Schattauer, 2006.
[NMDS] Schaefer et al.: Mitochondrial disease in adults: A scale to monitor progression and treatment, Neurology, 2006.
[PD08] K. Pommerening, et al.: Register zu seltenen Krankheiten, Bundesgesundheitsblatt, 51/2008, S. 491-499.
[PO07] K. Pommerening, Das Datenschutzkonzept der TMF für Biomaterialbanken, it - Information Technology, 49/2007, S. 352-359.
[SARA] Schmitz-Hübsch et al.: Scale for the assessment and rating of ataxia: Development of a new clinical scale, Neurology, 2006.
[SF12] www.sf-36.org/tools/sf12.shtml, letzter Zugriff 30. April 2010.