TECHNISCHE UNIVERSITÄT MÜNCHEN

Institut für medizinische Statistik und Epidemiologie, Lehrstuhl für medizinische Informatik

Identitätsmanagement in verteilten Umgebungen zur Unterstützung der translationalen medizinischen Forschung

Gregor Lamla

Vollständiger Abdruck der von der Fakultät für Informatik

der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Nassir Navab

Prüfer der Dissertation:

1. Univ.-Prof. Dr. Klaus A. Kuhn
2. Univ.-Prof. Dr. Alois Knoll

Die Dissertation wurde am 28.02.2013 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik

am 11.09.2013 angenommen.

Abstract

Today, personal data of patients and probands are identified and managed separately and differently and redundantly in different information systems. A specific divide exists between research and clinical environments. Both health care and research are, however, in need of complete follow-up information, regardless of its origin. This leads to complex legal and technical questions: Identities have to be managed across intra- and inter-institutional boundaries. IT solutions need to adhere to laws and regulations and, in research, they also have to safeguard intellectual property rights. The resulting requirements are complex, and to a significant degree specific for the biomedical domain.

In this dissertation, a concept for managing identities across different information systems has been developed. Identities are not only referring to actors (physicians and researchers with roles and rights) but also to subjects (patients, probands) and to biosamples. Data protection, security, and privacy, are of central importance in this context, especially when biospecimens are involved at a time, when the sequencing of a whole genome has become easy and inexpensive. A basic requirement is the reliable and regulation-compliant identification of patients and of biosamples. This task can already be non-trivial in pure research environments, but it becomes highly complex when clinical data are to be re-used. In research, access to data and biospecimens is typically based on informed consent, while in health care, access rights are depending on the context of treatment. In research environments, additional safeguarding, including pseudonymization and anonymization is needed. This work presents a software architecture which has been developed to handle identities in research environments and to access data from clinical systems if informed consent has been given. The implementation has focused on typical and frequently used clinical as well as research IT-systems.

In an initial step, wrapper components have been developed, providing search functionality on patient-identifying data in component systems. To manage the complexity of the authorization concepts, an Access Control System has been implemented. Whenever possible, the authorization check functionality of the component systems is reused. In applications where this is not possible, the authorization concepts are reproduced in the wrappers.

The module developed for distributed search of identifying data can be used to build a master patient index (MPI). This MPI is a core element of managing identities, medical data, and sample identities and sample annotations in distributed

environments. As identification errors are common, a concept for handling them (reconciliation) had to be established, building upon reconciliation concepts of the integrated component systems. In order to allow the user to view medical data in their original context, a module providing virtually seamless access to component systems has been developed, which considers the complex access permissions of the components.

The concept allows access to data that may have been collected on different legal bases and allows the reuse of this data in accordance with legal requirements and authorization concepts. A module has been developed, which is supporting regulation-compliant data transfer from clinical systems to systems for structured data entry of research data.

Zusammenfassung

In der Krankenversorgung und in der medizinischen Forschung werden personenbezogene Daten ein- und desselben Patienten oder Probanden in verschiedenen IT-Systemen unterschiedlich identifiziert und redundant verwaltet. Häufig müssen aber für Fragestellungen aus Krankenversorgung wie auch Forschung Krankheitsverläufe über Einrichtungsgrenzen hinweg charakterisiert werden. Dies stellt hohe Anforderungen an ein datenschutzkonformes Identitätsmanagement. Die in Forschung und Krankenversorgung zunehmend geforderte Einbeziehung von Bioproben, aus denen rasch und mit immer geringeren Kosten genomische Daten extrahiert werden können, erhöht die Komplexität, die mit hohen und zu einem signifikanten Grad auch domänenspezifischen Sicherheitsanforderungen einhergeht.

Im Rahmen dieser Arbeit wurde ein Konzept zur systemübergreifenden Verwaltung von Identitäten entwickelt, das Handelnde (Ärzte, Forscher) und Behandelte (Patienten, Probanden, deren Proben) einbezieht. Schwerpunkt ist die Identifikation von Patienten und Proben, daneben auch eine rechts- und ethikkonforme Weiterverwendung von Daten für Forschungszwecke. Neben den Anforderungen im Bereich Sicherheit und Datenschutz ist im Forschungskontext auch die Wahrung des geistigen Eigentums zu beachten. Im hier vor allem betrachteten Forschungsumfeld basieren die Rechte zur Erhebung von Daten und Sammlung von Bioproben auf Patienteninformation und Einverständnis, während im klinischen Bereich Behandlungsvertrag und Behandlungszusammenhang im Vordergrund stehen. Spezifika in der Forschung sind Sicherheitsmaßnahmen wie Pseudonymisierung und Anonymisierung. Eine Softwarearchitektur wurde entwickelt, die Identitäten in verschiedenen Systemen sicher und rechtskonform verwaltet und auf der Basis von Einverständniserklärungen die Übernahme von Daten ermöglicht. Die Implementierung erfolgte für typische und häufig verwendete klinische und Forschungssysteme.

Zunächst wurden Wrapper entwickelt, die in klinischen Systemen Suchfunktionalitäten auf identifizierende Daten ermöglichen. Aufgrund der Komplexität der Berechtigungskonzepte wurde ein Access Control System entwickelt. Wenn möglich wird auf die Berechtigungsprüfungen in den Komponentensystemen selbst zurückgegriffen. In Anwendungen bei denen dies nicht möglich ist, erfolgt eine Nachbildung der Berechtigungskonzepte in den Wrappern.

Das entwickelte Modul zur verteilten Suche auf identifizierende Daten bildet die Grundlage für einen Master Patient Index (MPI), der über Pseudonymisierungsmaßnahmen abgesichert werden kann. Der MPI wiederum ist die Grundlage für eine gesetzeskonforme und umfassend abgesicherte verteilte Verwaltung von Daten und Bioproben. Da typischerweise in den Quellsystemen eine auch die Identitäten betreffende Fehlerbehandlung implementiert ist, muss diese nachgeführt werden. Hierbei kann es notwendig werden, die mit den Identitäten verknüpften medizinischen Daten im Originalkontext einsehen zu können. Hierzu wurde ein Modul entwickelt, das unter Patientenbezug einen Kontextwechsel zwischen den Komponentensystemen ermöglicht. Auch hier ist die Wahrung von Rechten eine wesentliche Rahmenbedingung.

Der (pseudonymisierte) MPI ermöglicht den Zugriff auf Daten, die mit unterschiedlicher rechtlicher Basis erhoben sein können und erlaubt eine sichere, auf den jeweiligen Berechtigungskonzepten basierende Weiterverwendung. Hierzu wurde ein Modul entwickelt, welches eine (regularien) konforme Übernahme von Daten aus klinischen Systemen in Systeme für die strukturierte Erfassung von Forschungsdaten ermöglicht.

Danksagung

An dieser Stelle möchte ich mich insbesondere bei Prof. Kuhn für die Betreuung meiner Doktorarbeit bedanken. Er hat durch seine Anregung zum Thema, den fruchtbaren Diskussionen und den Erörterungen aktueller Entwicklungen in der translationalen Medizin einen wesentlichen Teil zum Entstehen dieser Arbeit beigetragen. Ebenso möchte ich mich bei Prof. Knoll für wiederholte Anregungen, kritische Nachfragen und wichtige Diskussionen bedanken. Mein Dank gilt auch meinen Kollegen am Lehrstuhl für medizinische Informatik, insbesondere Dr. rer. nat. Sebastian Wurst, Fabian Prasser und Florian Kohlmayer, mit denen ich in jeder Phase meiner Dissertation auftretende Fragen erörtern konnte. Schließlich gilt mein Dank noch der TUM Graduate School for Information Science in Health (GSISH), die meine Arbeit gefördert und einen Auslandsaufenthalt an der Vanderbilt University in Nashville, TN, ermöglicht hat.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

1 Einführung

1.1 Entwicklungen in der medizinischen Forschung

Nach der erfolgreichen ersten Sequenzierung des menschlichen Genoms stehen genomische Daten in stark wachsendem Umfang zur Verfügung. Ein Kernthema der biomedizinischen Forschung besteht nun darin, auch detaillierte klinische Verläufe zu erfassen, um die biologischen und medizinischen Abläufe vom Genom zum Protein und zur phänotypischen Ausprägung einschließlich der Interaktion mit der Umwelt besser zu verstehen. Unter „translationaler Forschung“ kann ein integrativer, in die sozioökonomische Umgebung eingebundener, Ansatz verstanden werden, der auf einem multidirektionalem Verständnis von Forschung und Medizin basiert [Sonntag2005]. Translationale Forschung wird häufig mit der Kurzformel „from bench to bedside and back“ beschrieben, wobei sich eine Einteilung in Phasen herausgebildet hat [Sung2003]: In der „T1“ Phase erfolgt die Übernahme neuer Erkenntnisse über Krankheitsmechanismen aus dem Labor für die Entwicklung neuer Methoden zur Vorsorge, Diagnose und Therapie von Krankheiten sowie deren Tests an Menschen. Die Übernahme von Ergebnissen aus klinischen Studien in die tägliche klinische Praxis erfolgt in der „T2“ Phase [Sung2003]. Der Schritt zurück, d.h. das Lernen aus klinischen Daten wird besonders in jüngerer Zeit betont [Collins2011].

Wissenschaftlicher Fortschritt in der Medizin beginnt aus Sicht der U.S. National Institutes of Health (NIH) typischerweise mit Grundlagenforschung im Labor, wo Wissenschaftler sich auf molekularer oder zellulärer Ebene mit Krankheiten befassen („bench“) [NIH2011]. Damit dort erarbeitete Erkenntnisse zu einer Verbesserung der Gesundheitsversorgung führen, müssen sie in praktische diagnostische und therapeutische Maßnahmen überführt, also auf die klinische Ebene zu realen Patienten gebracht werden („bedside“) [NIH2011].

Während Grundlagenforscher den Klinikern neue Erkenntnisse und Werkzeuge zur Behandlung ihrer Patienten sowie der Bewertung des Behandlungserfolgs zur Verfügung stellen, beeinflussen umgekehrt klinische Studien oder Beobachtungen von Krankheitsverläufen die Grundlagenforschung. Ein wesentliches Beispiel für

diesen zweiten Fall sind klinische und epidemiologische Register. Biomaterialbanken spielen hierbei eine immer wichtigere Rolle [Yuille2008]. Eine bessere Verknüpfung von Beobachtungen und Erkenntnissen aus Grundlagenforschung und klinischer Forschung kann nicht nur den Prozess der Translation beschleunigen, sondern auch helfen, die komplexen, multifaktoriellen Zusammenhänge besser zu durchschauen und zahlreiche Wissenslücken zu schließen.

Für Personen, die unmittelbar an der Behandlung von Patienten beteiligt sind, stellt translationale Forschung einen Ansatz dar, mit dessen Hilfe die Übernahme von Erkenntnissen aus der Forschung für diagnostische oder therapeutische Zwecke in die tägliche klinische Praxis beschleunigt wird. Unter industriellen Gesichtspunkten wird translationale Forschung als Prozess verstanden in dem noch während einzelner Phasen Entscheidungen über die die Weiterentwicklung oder den Abbruch von Ansätzen zu Zeitpunkten getroffen werden können, an denen die Kosten noch relativ niedrig sind [Littman2007]. Aufgrund der angestrebten engen Verknüpfung von bench und bedside wird anstelle von Translationsforschung häufig auch der Begriff „translationale Medizin“ verwendet.

Moderne Analyseverfahren, wie bspw. Transkriptom-, Proteom- oder Metabolomanalysen („omics“), produzieren große Datenmengen. Weltweit sind Ergebnisse solcher Analysen in Daten- bzw. Wissensbanken gespeichert. Gleichzeitig fallen kontinuierlich Daten bei der Behandlung von Patienten, aus Studien und in Registern an, die ebenfalls hohe Datenvolumina erreichen können. Mit dem raschen Voranschreiten der Möglichkeiten sind zahlreiche autonome Informationssysteme entstanden. Diesen Systemen liegen je nach Forschungsbereich unterschiedliche technische und organisatorische Anforderungen zugrunde, die sich in sehr heterogenen Systemdesigns niederschlagen, und die teilweise bei einer Integration nicht kompromittiert werden dürfen. Dies gilt insbesondere für rechtliche Anforderungen und den Datenschutz. Aus der Notwendigkeit, autonome und sehr heterogene Daten- und Wissensquellen für die Unterstützung der translationalen Forschung zu integrieren und aus den raschen Veränderungen in der Medizin ergeben sich komplexe Anforderungen an die IT.

1.2 Herausforderungen für die Informatik in der translationalen Medizin

Bei der Unterstützung der translationalen Medizin durch Informationstechnologie ergeben sich für die Informatik in typischer Weise Herausforderungen in den Bereichen Verteilung, Heterogenität und Autonomie [Prasser2011]. Die hohe Fragmentierung des Gesundheitswesens sowie die Trennung von Krankenversorgung, klinischer und grundlagenorientierter Forschung haben zu verteilter und semantisch heterogener Datenhaltung geführt. Die im Einsatz befindlichen Informationssysteme sind häufig nicht integrierte Legacy-Systeme und ad hoc Lösungen. Dies liegt darin begründet, dass der Entwurf, die Implementierung, die Einführung und der Betrieb großer Informationssysteme zur Unterstützung der Krankenversorgung sehr kostenintensiv und aufwendig ist. Beispielsweise können die Kosten für Beschaffung und Einführung eines Krankenhausinformationssystems im Bereich der Maximalversorgung mehrere Millionen Euro betragen, und die Dauer der Einführung kann mehrere Jahre in Anspruch nehmen. Für die Neuimplementierung eines solchen Systems ist mit einem Mehrfachen der Kosten und wesentlich mehr Zeit zu rechnen.

Bei der Informationsintegration in der translationalen Medizin gilt es Konzepte und Ansätze zu adaptieren bzw. zu erweitern. So werden im Rahmen des CTSA Programms (Clinical Translational Science Awards) des NIH in den USA 60 akademische Zentren gefördert. Der initiale 5-Jahresrahmen variiert zwischen 20 und 100 Mio. USD pro Institution [CTSA]. In den Zentren müssen heterogene Informationen (genetische / phänotypische Daten, Daten aus Krankenversorgung, Versorgungsforschung, Clinical Enterprise and Research Enterprise) integriert werden. In neueren Projekten finden u.a. ontologiebasierte Ansätze bei der Integration Verwendung z.B. bei CTSA am Health Science Center der University of Texas in Houston [Mirhaji09]. Dort wurde ein ontologiebasierter Ansatz entwickelt, in dem mit einer einzelnen Ontologie verschiedene Modelle beschrieben werden. Diese Modelle dienen u.A. zur Beschreibung von klinischen und translationalen Forschungsprozessen, medizinischen Informationen, Zugriffsberechtigungen sowie elektronischen Fragebögen.

Payne et al. sehen für die translationale Forschung in der Schaffung semantischer Interoperabilität die wesentliche Herausforderung [Payne2009]. Während sich im Laufe der Zeit innerhalb vieler Bereiche sogar konkurrierende Standards entwickelt haben, sind Standards zum Austausch von Informationen zwischen

Grundlagenforschung, klinischer Forschung und klinischer Praxis nur in Teilbereichen, wie etwa CDSIC-ODM im Rahmen von Studien, vorhanden [Embi2009]. Lediglich im Bereich der Krankenversorgung wird auf Standards wie HL7 und DICOM zum Austausch medizinischer Informationen sowie auf Terminologie- bzw. Nomenklaturstandards wie ICD und SNOMED [SNOMED] zurückgegriffen. Die während der Behandlung erfassten Daten sind häufig unterschiedlich strukturiert und werden terminologisch unzureichend kontrolliert. Daraus resultieren Herausforderungen an die Informatik hinsichtlich deren semantischer Erschließung. Im Bereich der Krankenversorgung kommen zwar Klassifikationen zur Charakterisierung von Krankheiten zum Einsatz. Sie sind für Abrechnungszwecke geeignet, aber für die Forschung nur sehr eingeschränkt verwendbar. Wesentliche Befunde mit durchaus substantiellen Inhalten liegen dagegen in Freitextform vor. Diese weisen zwar häufig eine Gliederung auf, bildet aber Strukturen und Vokabulare nur implizit ab. Kommen in der Krankenversorgung oder auch der Forschung strukturierte Formulare zum Einsatz, so sind diese oftmals zwischen Institutionen oder Projekten nicht abgeglichen. Aspekte die (Quasi-)Synonyme, Homonyme und semantische Überlappungen betreffen, werden unzureichend berücksichtigt.

Weitere Herausforderungen für die Informatik ergeben sich in der translationalen Medizin bei der Weiterverwendung klinischer Daten für die Forschung. Diese steht in Europa am Anfang. Für die Weiterverwendung müssen komplexe Herausforderungen die sich aus Zugriffen auf die Daten, aus dem Datenschutz, der Sicherheit, aus Regularien sowie neuen Technologien ergeben, gelöst werden [Deloitte2012]. Der Zugriff auf Patientendaten ist im Versorgungskontext nur gestattet, wenn ein Behandlungszusammenhang besteht. Dieser kann sich dynamisch im Behandlungskontext ändern. Werden Daten für die Forschung verwendet, müssen diese anonymisiert sein, oder es muss die Einverständniserklärung des Patienten (Informed Consent) vorliegen. Im Forschungsfall muss eine Trennung der medizinischen von den direkt identifizierenden Daten erfolgen, die zu komplexen Problemen im Fall einer Integration führen kann. Bei der Weiterverwendung von Behandlungsdaten in der Forschung muss die u.U. extrem komplexe Rechte- und Rollensituation abgebildet und zusätzlich der Schutz des geistigen Eigentums der Forscher beachtet werden. Aspekte der Sicherheit, des Datenschutzes und ethisch-rechtliche Fragen sind besonders relevant wenn Bioproben verwendet werden. Die aus ihnen leicht zu gewinnenden genetischen Daten enthalten nicht nur schützenswerte Informationen

über den Spender, sondern auch über seine Verwandten. Zahlreiche Angriffe auf die Privatsphäre sind beschrieben worden. Einen guten Überblick geben Malin et al. [Malin2010]. Lokale, nationale und internationale Gesetze und Regularien, wie Ethikrichtlinien, nationale Datenschutzgesetze und die EU-Datenschutzrichtlinie 95/46/EG müssen beachtet werden, ebenso wie Empfehlungen zu Biomaterialien [Rec2006]. Es gibt laufende Bestrebungen zur Harmonisierung, aber auch zur Anpassung an die neuen Möglichkeiten der Biotechnologie [Ethik2010].

Auch im Versorgungsbereich ergeben sich Herausforderungen bei der Integration von Informationssystemen. So werden zur Integration der vorhandenen Informationssysteme in zahlreichen Staaten derzeit nationale Infrastrukturen aufgebaut, wobei allein in Deutschland mehr als 2.000 Krankenhäuser und über 100.000 Arztpraxen zu integrieren sind. In den US sind mit dem Health Information Technology for Economic and Clinical Health (HITECH) Act [HITECH] Fördermaßnahmen zur Verbesserung der IT-Ausstattung im Versorgungsbereich in der Größenordnung von 30 Mrd. USD angelaufen. Ziel ist es den Nutzen von elektronischen Patientenakten zu erhöhen um die Krankenversorgung zu verbessern in dem z.B. Doppeluntersuchungen reduziert werden. Hierzu müssen unterschiedlichste IT-Systeme integriert werden.

Weitere Aufgaben für die Informatik ergeben sich durch die hohe Dynamik der Domäne. Medizinische Strukturen und Prozesse unterliegen ständigen Veränderungen, bspw. durch die Einführung neuer diagnostischer und therapeutischer Prozeduren [Lenz2004]. Häufig sind Experten in der medizinischen Domäne nicht in der Lage, die ihrer Arbeit zugrunde liegenden Prozesse zu beschreiben, was sich erschwerend auf die Softwareentwicklung wirkt [Wears2005], [Aarts2004], [Kuhn2001]. Zudem liegt jeder Entwicklung ein komplexes System mit dynamischen Interaktionen zwischen Technologie, Mitarbeitern und Organisationsstrukturen zugrunde. Soziale und technische Elemente sind stark voneinander abhängig und stehen in wechselseitiger Beziehung zueinander. Der IT-Einführungsprozess selbst verändert Arbeitsabläufe, wodurch sich wiederum die Anforderungen an die IT ändern können [Wears2005], [Lenz2004]. Aspekte der Nachhaltigkeit bleiben unberücksichtigt, weil häufig in Projekten eine Priorisierung von Aktivitäten vorgenommen wird, die sich durch unmittelbare Sichtbarkeit auszeichnen. Dies ist auf ein geringes Verständnis der Anwender für technische Hintergründe zurückzuführen [Killcoyne2009]. Somit müssen Methoden und Werkzeuge bereitgestellt werden, um bei häufig geänderten Anforderungen und unklaren Vorgaben ein hohes Maß an Softwarequalität zu gewährleisten.

1.3 Ziele

Für die translationale Medizin sind Daten von Interesse, die gesamte Krankheitsverläufe charakterisieren. Um diese Verläufe umfassend beschreiben zu können, ist es notwendig Daten, die während (wiederholten) Aufenthalten in unterschiedlichen Einrichtungen erfasst werden, zusammenzuführen zu können. Dies setzt voraus, dass digitale Identitäten von Patienten über Einrichtungsgrenzen hinweg verwaltet werden können.

In Einrichtungen, in denen medizinische Leistungen und Forschung auf Weltniveau erbracht wird, ergeben sich Rahmenbedingungen, die in existierenden Ansätzen zum Management von Identitäten bisher nicht bzw. zu wenig berücksichtigt werden. Existierende Infrastrukturen und Prozesse sind in diesen Einrichtungen für die Krankenversorgung und die Forschung hochkomplex, sehr divergent und über Jahre gewachsen. Häufig befinden sich mehrere 100 IT-Systeme im Einsatz, mit denen jeweils einzelne spezifische Prozessschritte unterstützt werden. So ist lediglich in führenden IT-Systemen zur Unterstützung der Krankenversorgung den verschiedenen Identitäten eines Patienten eine einheitliche ID zugeordnet. Daneben erfolgt in den Systemen für Abrechnungszwecke eine detaillierte strukturierte Erfassung von demografischen Daten zu einem Patienten.

Phänotypische Ausprägungen von Krankheiten werden während den einzelnen Patientenaufenthalten zunehmend in detaillierterer Form erfasst (Deep Phenotyping). Dabei verschwimmt die Grenze zwischen Daten, die ausschließlich für Behandlungszwecke und welche die explizit für Forschungszwecke erhoben werden, immer mehr. Die Erfassung dieser feingranularen Daten erfolgt zunehmend für alle Patienten, die ein entsprechendes Krankheitsbild aufweisen und nicht wie früher für eine kleine Anzahl an Studienteilnehmern. IT-Systeme, welche die Behandlung von Patienten unterstützen, können häufig nicht an die sich schnell ändernden Anforderungen bzgl. der Erfassung neuer phänotypischer Daten angepasst werden, zudem ist die klinische Datenverarbeitung rein rechtlich an Institutionsgrenzen gebunden. Durch die Verwendung separater Systeme für die Krankenversorgung einerseits sowie von Systemen zur verteilten Erfassung feingranularer Forschungsdaten für große Patientenkollektive andererseits, ergibt sich die Forderung nach einem verteilten Identitätsmanagement.

Ziel dieser Arbeit ist es ein Konzept und dessen Umsetzung für ein übergeordnetes Identitätsmanagement zur Unterstützung von Forschungsprozessen innerhalb sowie zwischen Einrichtungen der Hochleistungsmedizin zu entwickeln. Eine wichtige

Anforderung ist es, den Funktionsumfang des Identitätsmanagements schrittweise ausbauen zu können. Das Management von Identitäten soll in der Lage sein, sich an ändernde klinische Abläufe und Forschungsprozesse, die sich u.a. durch die Einbeziehung zusätzlicher Einrichtungen ergeben, aufgaben- und ressourcenangemessen anzupassen. Von vornherein sollen Rechtsgrundlagen und domänenspezifische Rahmenbedingungen berücksichtigt werden. Dies umfasst Einwilligungserklärungen, Zugriffsberechtigungen auf IT-Systeme, Behandlungs- vs. Forschungskontext und mehr.

Inhaltlicher Schwerpunkt der Arbeit ist die sichere Identifikation von Patienten und Biomaterialien. Die Schaffung von Suchmöglichkeiten nach Identitäten im Routineeinsatz befindlichen Anwendungen bildet dabei den Ausgangspunkt. Ausgehend von Suchergebnissen erfolgt die Übernahme der Informationen zu Identitäten in Forschungssysteme. Ziel ist es, hochstrukturierte, die Identitäten charakterisierende Information, aus der Krankenversorgung für Forschungszwecke weiterzuverwenden. Die Sicherung der Konsistenz der weiterverwendeten Informationen ist zu gewährleisten. Die verschiedenen Identifikatoren der Identitäten aus den einzelnen Systemen werden verwaltet. Die Weiterverwendung der mit den Identitäten assoziierten medizinischen Daten für Forschungsfragen wird durch diese Arbeit vorbereitet. Insbesondere wird die selektive Übernahme hochstrukturierter während der Behandlung von Patienten in IT-Systemen der Krankenversorgung erfasster Daten in Forschungsanwendungen unterstützt.

Das Strukturprojekt „Data Integration System (DIS)“ des Münchener Biotech-Spitzenclusters m4 spielte für die hier vorgelegte Arbeit eine wesentliche Rolle für die Erstellung von Use Cases für das Identitätsmanagement und der Datenübernahme einerseits sowie den Einsatz von Ergebnissen andererseits. Ziel von m4 ist es, eine „Verbesserung des Entwicklungsprozesses und die Steigerung der individuellen Wirksamkeit und Sicherheit neuer Medikamente durch die Implementierung der „personalisierten und zielgerichteten Medizin“ entlang der gesamten Wertschöpfungskette der Medikamenten- und Diagnostikaentwicklung“ zu erreichen [m4]. Hierzu wird in DIS eine Integrationsarchitektur geschaffen, die Forschungsprozesse und klinische Abläufe gemeinsam betrachtet und eine integrative und einheitliche Datenverarbeitung ermöglicht.

1.4 Begriffe

Im Folgenden werden Begriffe vorgestellt bzw. geklärt, die im weiteren Verlauf der Arbeit verwendet werden. Sie basieren auf Definitionen der ITU (International Telecommunication Union) [ITU], der ACM (Association for Computing Machinery) [ACM], und einer Arbeit aus einem ETSI (European Telecommunications Standards Institute) Security Workshop [Harrop2009]. Identität ist laut ITU „eine Repräsentation einer Entität in Form eines oder mehrere Attribute, die es erlaubt, eine Entität oder Entitäten innerhalb des Kontextes hinreichend zu unterscheiden“ [ITU1]. Identität ist „sowohl ein Konzept der realen Welt als auch ein digitales Konstrukt“ [Harrop2009]. In der digitalen Welt ist Identität “Information über eine Entität die ausreicht, diese Entität in einem bestimmten Kontext zu identifizieren” [Harrop2009]. „Eine Person kann in der digitalen Welt mehrere unterschiedliche Identitäten haben“ [Harrop2009]. Im Rahmen des Identitätsmanagements wird der Begriff Identität „als kontextabhängige Identität (Teilmenge an Attributen) verstanden, bei dem z.B. die Auswahl an Attributen durch ein Framework mit definierten Randbedingungen (dem Kontext), in dem die Entität existiert und interagiert, beschränkt ist“ [ITU1].

Unter Identifikation wird „das Verknüpfen einer unterscheidenden Kennzeichnung (Identifikator) mit etwas innerhalb einer spezifischen Gruppe oder Kontextes“ verstanden [ACM1]. Ein Identifikator (ID) besteht aus „einem oder mehreren Attributen, die benutzt werden, um eine Entität innerhalb eines Kontextes zu identifizieren“ [ITU1].

Den Entitäten der realen Welt werden in der digitalen Welt digitale Objekte mit digitaler Identität zugeordnet. Sie modellieren reale Entitäten mehr oder weniger vollständig und bilden Identitäten der realen Welt auf die digitalen Entitäten ab. Identifikatoren gibt es in der realen und in der digitalen Welt, d.h. bezogen auf reale und auf Dokumentationsobjekte.

In dieser Arbeit stehen digitale Identifikatoren im Vordergrund. Eine Reihe von Arbeiten hat Identifikatoren untersucht und verschiedene Typen beschrieben. Wichtig ist die Unterscheidung von Malin et al. [Malin2005] zwischen expliziten Identifikatoren, Quasi-Identifikatoren, und Nicht-Identifikatoren. Beispiele der zweiten Kategorie sind Datums- und Ortsangaben, die leicht zu einer Re-identifikation herangezogen werden können [Sweeney2002]. Nicht identifizierend sind bspw. Surrogatschlüssel in einer Datenbank oder Attribute, die eine sehr geringe Unterscheidbarkeit beinhalten und somit nur mit extrem großem Aufwand für eine Re-identifikation verwendet werden könnten. In der Medizin kommt der

Geheimhaltung von Identitäten eine besondere Bedeutung zu, insbesondere wenn Bioproben und damit prinzipiell auch genomische Daten verarbeitet werden.

Im Rahmen dieser Arbeit sind Patienten, Proben und Fälle die wichtigsten Entitäten der realen Welt. Den Patienten werden im Behandlungskontext stationäre Aufenthalte und ambulante Besuche als „Fälle“ zugeordnet. In der klinischen Forschung werden für Fälle dieselben oder ähnliche Begriffe verwendet: Besuche, Visits, Vorstellungen. Entnommene Bioproben werden i.a. sowohl dem Patienten als auch dem Fall (Besuch oder Behandlungsepisode) zugeordnet. In der Pathologie findet sich der Begriff „Fall“ auch für eine Probeneinsendung. Häufig sind in unterschiedlichem Kontext (Behandlung, Studien) die digitalen Identitäten der Person bzw. der zugehörigen Probe richtig zugeordnet, aber die Zuordnung der verschiedenen digitalen Identitäten ein und derselben realen Person oder auch Probe zueinander fehlt.

1.5 Aspekte im Kontext des Identitätsmanagements

Neben der immer detaillierteren Erfassung phänotypischer Daten stehen zunehmend mehr Daten, die das individuelle Genom eines Menschen beschreiben, zur Verfügung. Durch die Verknüpfungen von genetischen Daten und deep phenotyping Daten ergeben sich neue Möglichkeiten Zusammenhänge zwischen Geno- und Phänotypen zu identifizieren [Tracy2008]. Um phänotypische, genetische und Biomaterialdaten zu einem Patienten zusammenführen zu können, ist es notwendig die verschiedenen Identitäten des Patienten zwischen den IT-Systemen zur Unterstützung der Krankenversorgung, der Erfassung von Forschungsdaten, der Verwaltung von Biomaterialien und Genomdaten managen zu können. Hierfür existieren unterschiedliche Konzepte und Infrastrukturen. Eine der führenden CTSA-Infrastrukturen wurde an der Vanderbilt University (Nashville, TN, USA) entwickelt [Roden2008]. Diese unterstützt die Zusammenführung von phänotypischen Daten aus de-identifizierten elektronischen Akten, Blutproben sowie genetischen Daten. Dadurch werden neue Ansätze für die Forschung wie z.B. Phenome wide association scans (PheWAS) ermöglicht [Denny2010]. Allerdings bedarf es für die Umsetzung der Vanderbilt Infrastruktur eines Multimillionen-Investments [Mak2011].

Eine zentrale Herausforderung bei der Zusammenführung dieser Daten ist der Schutz von Persönlichkeitsrechten der Patienten. So muss sichergestellt werden, dass die Re-Identifikation eines Individuums nicht bzw. nur bei entsprechender Autorisierung möglich ist. Da bereits kurze DNA-Abschnitte (weniger als 100 von 3

Milliarden Elementen) genügen, um eine Zuordnung zu einer Vergleichsprobe bzw. einer Person herzustellen [Lin2004], müssen u.U. aufwendige Anonymisierungsschritte vorgesehen werden.

Um Daten die im Rahmen der Krankenversorgung erfasst wurden datenschutzkonform in der Forschung weiter nutzen und diese z.B. mit genetischen Daten zusammenführen zu dürfen, wurde von der Technologie- und Methodenplattform für die vernetzte medizinische Forschung (TMF) ein Konzept entwickelt. In diesen Konzepten erfolgt ein rudimentäres Management von Identitäten zwischen IT-Systemen zur Unterstützung der Forschung. Aspekte des Identitätsmanagements die sich für dessen Unterstützung an einem Standort mit unterschiedlichen Forschungsschwerpunkten ergeben, bleiben unberücksichtigt. Eine umfassende praxistaugliche Implementierung der TMF-Konzepte ist bisher nicht erfolgt

1.6 Ausgangssituation

Im Rahmen der Patientenbehandlung werden in IT-Systemen zur Unterstützung der Krankenversorgung Identitäten der Patienten neu angelegt bzw. bereits vorhandene weiterverwendet. Werden Daten für Forschungszwecke erhoben, so geschieht dies häufig in separaten IT-Systemen. Ein übergreifendes Management für Identitäten zwischen Forschungssystemen und Routinesystemen aus der Krankenversorgung existiert häufig nicht. Für ein system- und einrichtungsübergreifendes Identitätsmanagement ist es notwendig sowohl auf Identitätsinformationen aus den Routinesystemen als auch aus Forschungssystemen zugreifen zu können. Dabei müssen rechtliche Rahmenbedingungen berücksichtigt werden.

Eine Übersicht über die verschiedenen IT-Systeme, rechtlichen Rahmenbedingungen sowie Prozesse und Abläufe die in dieser Arbeit von Relevanz sind, findet sich in den folgenden Unterkapiteln.

Klinische IT-Systeme

Das Krankenhausinformationssystem (KIS) ist das zentrale Informationssystem eines Krankenhauses. Es findet in verschiedenen Bereichen eines Krankenhauses Verwendung insbesondere auf Stationen und in Ambulanzen, aber auch in der Verwaltung einschließlich Finanzbuchhaltung, Materialwirtschaft, Human Ressource Management und Controlling. Eine zentrale Funktionalität des KIS stellt die

Patientendatenverwaltung (PDV) dar. Dabei werden demografische Daten und administrative Daten der im Krankenhaus behandelten Patienten verwaltet. Jedem Patienten wird vom KIS ein krankenhausinterner eindeutiger Identifikator zugeordnet.

Eine zentrale Funktion der PDV ist die Bewegungsdatenverwaltung der Patienten, d.h. Aufnahme, Verlegung und Entlassung, verbunden mit der Dokumentation von Diagnosen und Maßnahmen als Basis für die Abrechnung. Hierzu erfolgt auch eine Datenübermittlung an Krankenkassen.

Das in das KIS integrierte Klinische Arbeitsplatzsystem (KAS) stellt die EDV-Lösungen für die eigentliche Krankenversorgung bereit. Typische Funktionen sind die Befund- und Auftragskommunikation und letztlich die Bereitstellung einer hausinternen elektronischen Krankenakte. Hierzu speichert das KAS auch Replikate aus Subsystemen, bspw. Laborbefunde. Bis zu einem gewissen Grad werden auch Abläufe unterstützt. Zur Unterstützung sog. Behandlungspfade werden Arbeitsschritte und die damit assoziierten Arbeitslisten, Dokumente, Termine und Aufgaben hinterlegt. Obwohl KAS-Systeme Replikate verwalten, dienen sie auch der direkten Eingabe klinischer Daten. Hierzu werden häufig Formulargeneratoren bereitgestellt, wobei zur Datenhaltung sowohl generische als auch konventionelle DB-Schemata verwendet werden. Ein in Deutschland aber auch weltweit sehr verbreitetes KAS ist i.s.h.med [ishmed] von Siemens. In i.s.h.med werden medizinische Daten mit sog. „Parametrisierbaren Medizinischen Dokumenten (PMDs)“ erfasst. PMDs ermöglichen es den i.s.h.med Administratoren hochstrukturierte Dokumente zu erzeugen. Jedem PMD ist ein Dokumententyp zugeordnet. Ein Dokumententyp kann mind. einer Organisationseinheit zugeordnet werden. Ebenso ist es möglich einen Dokumententyp auch einer Leistung, in deren Kontext ein Dokument erstellt wird, zuzuordnen. Ein PMD setzt sich aus Feldelementen zusammen, die sich wiederum gruppieren lassen. Als Feldelemente existieren Text-, Datums-, Zeit-, und numerische Felder. Für die Feldelemente lassen sich Wertebereichsprüfungen, Hilfefunktion und eine Werteauswahl, deren Inhalt in zu definierenden Tabellen hinterlegt wird, festlegen. Als grafische Feldelemente stehen Standardelemente wie List-, Check-, Comboboxen, Radiobuttons und Labels zur Verfügung. Änderungen am Erscheinungsbild oder die Modifikation der Funktionalität der Feldelemente erfolgt ereignisabhängig oder über ein hinterlegtes Programm. Fremddatenbausteine stellen eine Möglichkeit dar, PMDs mit Daten, die in Tabellen abgespeichert sind, vorzubefüllen. Eine Alternative sind sog. Business Add Ins (BAdI). Dabei handelt es sich um definierte

Unterprogrammaufrufe aus dem, für die Kontrolle eines PMDs verantwortlichen, Dialogprogramm. Als grafische Benutzerschnittstelle für die Visualisierung von PMDs wird die lokal auf den Clientrechnern installierte SAPGui eingesetzt.

Zusätzlich zum KIS und KAS kommen Spezialsysteme (Abteilungssysteme) zum Einsatz: Radiologieinformationssysteme (RIS) verwalten primär die Radiologie, unterstützen hierbei die Terminplanung und die Untersuchungsabläufe, die Verwaltung von Patientenstammdaten, die Anbindung von Modalitäten (CT, MRT, usw.) mit denen die Untersuchung durchgeführt wird, sowie die Dokumentation medizinischer Daten gemäß den Anforderungen der Röntgenverordnung. Zusätzliche Funktionalitäten ermöglichen die Erstellung von Untersuchungsbefunden und deren Verwaltung sowie die Dokumentation abrechenbarer Leistungen.

Für die Speicherung und die Kommunikation von Bildern werden PACS (Picture Archiving and Communication Systems) eingesetzt.

Laborinformationssysteme (LIS) dienen der Probenidentifikation, der Messwerterfassung, der Messwertauswertung, der Befunderstellung und zur Dokumentation der erbrachten Leistungen für Abrechnungszwecke. Pathologieinformationssysteme unterstützen Abläufe in der Pathologie. Dazu stellen sie Funktionalitäten zur Probenidentifikation, der Probenverfolgung, der Befunderstellung und der Leistungsabrechnung bereit.

IT-Systeme für die klinische Forschung

Im Rahmen der klinischen Forschung finden verschiedene IT-Systeme Verwendung. Ein sog. Clinical Data Management System (CDMS) dient zur Erfassung und Verarbeitung von Daten in klinischer Studien; sie weisen in ihrer Architektur Ähnlichkeiten zu KAS Systemen auf, sind aber für die Abläufe bzw. Regularien von Studien nach dem Arzneimittelgesetz [AMG2005] ausgerichtet. Wichtige Regularien sind die der US Food and Drug Administration (FDA CFR 21 Part 11 [Part11]), wobei die Systeme einen regularienkonformen Entwicklungs- und Installationsprozess gemäß ICH GCP [GCP] mit entsprechender Zertifizierung nachweisen müssen. Hierzu sind verschiedene Validierungsschritte bei der Einführung und Verwendung durchzuführen (Qualifications). Der eigentliche Betrieb eines CDMS Systems erfolgt dann auf Grundlage von standardisierten Arbeitsanweisungen (Standard Operating Procedures; SOP), die im Rahmen der Validierung eingeführt und überprüft wurden. Für die Erstellung strukturierter elektronischer Formulare (electronic Case Report Forms eCRF) zur Datenerfassung stellen CDMS-Systeme Formulargeneratoren

bereit. Um Fehleingaben zu reduzieren, ermöglichen CDMS Systeme auf einzelnen Feldern eine Überprüfung des Datentyps und des Formats der eingegebenen Daten. Zusätzlich kann eine Prüfung auf Abhängigkeiten der Eingaben zwischen Feldern unterschiedlicher Formulare (Cross Validation Checks) erfolgen. Einige CDMS Systeme unterstützen die Zuordnung von Terminologien zu den einzelnen Feldern.

Ein wichtiges Beispiel ist das Medical Dictionary for Regulatory Activities (MedDRA) [MedDRA] zur Erfassung und Codierung der Medikation sowie Logical Observation Identifiers Names and Codes (LOINC) für Laborwerte [LOINC]. CDMS Systeme ermöglichen die Gruppierung der Formulare und die Zuordnung dieser zu einem Follow-Up Termin (auch Patientenbesuch oder Visit). Weitere Funktionalitäten unterstützen das Electronic Data Capture (EDC), die doppelte Eingabemöglichkeit derselben Daten (Double Data Entry; DDE) und Abläufe zum Auflösen der sich daraus ergebenden Diskrepanzen. CDMS-Systeme setzen üblicherweise einen rollenbasierten Zugriff mit differenzierten Zugriffsberechtigungen um. Zusätzlich protokollieren CDMS gemäß FDA 21 CFR Part 11 sämtliche Zugriffe und die daraus resultierenden Änderungen in einer Protokolldatei (Audit Trail). Für statistische Analysen oder für die Übernahme der erfassten Daten in Studiendatenbanken können CDMS-Systeme umfangreiche Import/Exportfunktionalitäten in verschiedenste Formate anbieten.

Neben den streng nach AMG regulierten Studien zur Arzneimittelzulassung gibt es weitere nicht-AMG Studien, für die häufig der Begriff Register verwendet wird. Unter einem Register wird eine „standardisierte Dokumentation von Daten eines definierten Untersuchungskollektives, das Vollständigkeit innerhalb dieses Kollektives anstrebt“ verstanden [Leiner2006]. Klinische Register fokussieren darauf, Krankheitsverläufe bestimmter Erkrankungen über Einrichtungsgrenzen hinweg zu dokumentieren. Epidemiologische Register zielen darauf, alle Patienten einer bestimmten Region vollständig zu erfassen. Die Systeme ähneln den CDMS (teilweise werden die gleichen SW-Lösungen eingesetzt), allerdings sind die regulatorischen Anforderungen geringer. Generell sind die verwendeten SW-Lösungen heterogen. In den Anwendungsprojekten dieser Arbeit spielen sie eine zentrale Rolle. Das CDMS Macro [Macro2009] wird in der Münchener Forschungsumgebung häufig als Werkzeug für ein klinisches Register verwendet. Im Rahmen des Spitzenclusters m4 wurde zudem eine Lösung zur strukturieren Datenerfassung und Bioprobenverwaltung entwickelt, Macro und die m4-Lösung sind die wichtigsten im Rahmen dieser Arbeit betrachteten und in die Implementierungen eingebundenen Systeme. Sie werden im folgenden kurz als

„EDC-Systeme“ (EDC für Electronic Data Capture) oder, falls im Zusammenhang ersichtlich, als „Forschungssysteme“ bezeichnet.

Von herausragender und wachsender Bedeutung sind Biobanken [NCI2007], [Wichmann2011], [BBMRI]. Ihr Ziel ist es, Biomaterialien detailliert zu charakterisieren und in großer Stückzahl zur Verfügung zu stellen. Unter einer Biobank kann eine Sammlung von Proben menschlicher Körpersubstanzen (z.B. Gewebe, Blut, Serum, Organe) verstanden werden. Wichtig ist die Verknüpfung dieser Substanzen mit personenbezogenen Informationen der Spender. Die Besonderheit einer Biobank besteht im Doppelcharakter von Proben- und Datensammlung. Bei populationsbezogenen bzw. epidemiologischen Biobanken erfolgt die Probensammlung für bestimmte Bevölkerungskollektive. Bei krankheitsbezogenen Biobanken steht das Sammeln von Biomaterialien die mit einem bestimmten Krankheitsbild assoziiert sind, z.B. solide Tumoren, im Fokus [Ethik2004]. Für die Verwaltung der Proben und den damit assoziierten Informationen kommt verschiedenste Individualsoftware zum Einsatz. Auch Biobanken sind im Kontext der Projekte, für die diese Arbeit Konzepte und Lösungen entwickelt hat, von zentraler Bedeutung.

Für die Verwaltung von Daten für Forschungszwecke innerhalb einer Institution werden häufig Data Warehouses eingesetzt. Die in diese Warehouses replizierten Daten können zunächst aus Systemen zur Unterstützung der Krankenversorgung stammen, insbesondere KIS, KAS und LIS. Hier spielen alle an anderer Stelle genannte Aspekte von Sicherheit, Einverständnis, Identitätsmanagement, ggf. auch Anonymisierung eine Rolle. Weitere Datenquellen für Warehouses können in einem nächsten Schritt spezielle Forschungssysteme sein, bspw. für die Unterstützung von Mircoarrays, Proteomik- oder Gentypisierung. Solange Warehouses nicht anonymisierte Daten enthalten – was häufig der Fall ist – unterliegen Zugriffe auf diese Warehouses den Datenschutzgesetzen und erfordern auch positive Ethikvoten.

IT-Systeme für die Grundlagenforschung

IT-Systeme für die Grundlagenforschung finden sich etwa bei der DNA- und RNA-Sequenzierung. Analysepipelines umfassen hochspezialisierte Softwaretools, die Daten von Analyseplattformen verarbeiten. Die Softwaretools können auf verschiedenen Computern bzw. Clustern laufen, da mit ihnen häufig sehr rechenintensive Analysen durchgeführt werden. Es kann notwendig sein, auf öffentlich verfügbare Datenbanken zuzugreifen. Die Ergebnisse der Analysen werden

von den Plattformen verwaltet. Häufig sind diese Systeme bzgl. des Identitätsmanagements nicht integriert. Die Konzepte dieser Arbeit unterstützen ihre Anbindung, die jedoch sehr starke Sicherheitsmaßnahmen (doppelte Pseudonymisierung, räumlich und organisatorisch verteilte Datenhaltung) erforderlich macht. Da Biobanken das Ausgangsmaterial für Sequenzieranalysen bereitstellen, gelten die hohen Sicherheitsanforderungen auch für sie.

Standards, Vorgehensmodelle und Best Practices

In den beiden „Welten“, Forschung und Krankenversorgung, haben sich zahlreiche domänenspezifische Standards, Regularien, Vorgehensmodelle und Best Practices etabliert. Dabei wird zunehmend erkannt, dass Standards Brücken zwischen diesen Welten bieten müssen.

Im Bereich der Krankenversorgung sind HL7 (Health Level 7) [HL7] und DICOM (Digital Imaging and Communications in Medicine) [DICOM] sehr weit verbreitet.

Die HL7–Organisation hat verschiedene Standards, ursprünglich nur für den Datenaustausch zwischen IT-Systemen in der Krankenversorgung entwickelt. Heute sind die Versionen 2.x und 3 im Einsatz. In HL7 Version 2.x wird ein einfaches textbasiertes Nachrichtenformat spezifiziert. In den Textnachrichten sind u.a. Informationen zu Patienten- und Fallidentitäten enthalten. Der Versand von Nachrichten wird durch eintretende Ereignisse initiiert. Diese Ereignisse werden vom Standard vorgegeben und sind Teil der Nachricht. Mit der Clinical Document Architecture (CDA) stellt HL7 einen auf XML basierenden Standard für den Austausch und die Speicherung von klinischen Dokumenten bereit. Dabei werden drei aufeinander aufbauende Ebenen definiert, die unterschiedliche Anforderungen bzgl. des Grads an Strukturiertheit des Dokumenteninhalts vorgeben. Mit dem Clinical Context Object Workgroup (CCOW)- Standard [CCOW] unterstützt HL7 die Integration von IT-Systemen auf Präsentationsebene, in dem ein gemeinsames Kontextmanagement zwischen unabhängigen Anwendungen standardisiert wird. Ziel ist es, verschiedene unabhängige Anwendungen auf Präsentationsebene miteinander so zu synchronisieren, dass z.B. bei Auswahl eines Patienten in einer Anwendung der Kontext dieses Patienten in den integrierten Systemen aufgerufen und dem Endanwender angezeigt wird. Dabei soll der Kontextwechsel aus jedem der integrierten Systeme initiiert und in den anderen Systemen ausgeführt werden können. In dieser Arbeit wird ein Konzept für einen Kontextwechsel auf Präsentationsebene erarbeitet um den Abgleich von Informationen zu Identitäten zwischen den Systemen durch den Anwender zu erleichtern.

Die DICOM-Organisation entwickelt internationale Standards für den Austausch von digitalen Bildern und den dazugehörigen Metainformationen zwischen IT-Systemen und Geräten zur Bildgebung einschließlich der dazu benötigten Dienste und Prozesse. In der Organisation finden sich Hersteller von Software und Geräten für die Bildgebung sowie verschiedene Fachgesellschaften u.a. aus dem Bereich der Radiologie. DICOM wird im Rahmen dieser Arbeit nicht betrachtet, da der Fokus zunächst auf Daten und Bioproben liegt.

Für den Austausch von Daten im Rahmen von Studien haben sich die CDISC (Clinical Data Interchange Standards Consortium) Standards [CDISC] etabliert. Die CDISC-Organisation setzt sich aus Vertretern von Pharmaunternehmen, Contract Research Organizations (CROs), IT-Herstellern und universitären Institutionen zusammen. CDISC fokussiert dabei auf die Entwicklung von Standards um den kompletten „Lebenszyklus“ klinischer Studien von der Protokolldarstellung, der Aufzeichnung der Rohdaten, deren Analyse sowie das Abspeichern der Ergebnisse zu unterstützen. CDISC entwickelte hierfür u.a. das Operational Data Model (ODM), das Study Data Tabulation Model (SDTM), das Analysis Data Model (ADaM), sowie das Lab Data Model (LAB). SDTM und ADaM werden für die Übermittlung von Daten an Überwachungs- und Zulassungsbehörden wie der FDA eingesetzt. Dabei dient SDTM zur zusammenfassenden Beschreibung der über eCRFs erhoben Daten einzelner Studien. ADaM wird zur Dokumentation der eingesetzten statistischen Verfahren verwendet und ermöglicht die Nachvollziehbarkeit von Analysen. Für den Datenaustausch auf operativer Ebene werden LAB und ODM eingesetzt. LAB dient zur Übermittlung von Labordaten. ODM unterstützt den Austausch und die Archivierung von Daten aus klinischen Studien. Die Repräsentation von ODM erfolgt in XML. Herstellerspezifische Erweiterungen sind vorgesehen. Ein ODM-Dokument setzt sich aus vier Bereichen zusammen. Im Bereich „Study“ werden globale Variablen, Basisdefinitionen und Metadaten definiert. Globale Variablen beinhalten z.B. den Namen der Studie. Mit den Basisdefinitionen werden die in der Studie verwendeten Einheiten festgelegt. In den Metadaten werden Informationen zu den im Bereich „Clinical Data“ verwendeten Elementen abgespeichert. So enthalten die Metainformationen zu einem Element u.a. eine eindeutige ID, den Datentyp, und einen zulässigen Wertebereich. Der Bereich „AdminData“ enthält Informationen über die Anwender, die das CDMS benutzen, Informationen über die Studienzentren, die an der Studie beteiligt sind, sowie Sicherheitsinformationen. Der Bereich „ReferenceData“ enthält Informationen, die für die Interpretation der Daten nötig nicht aber studienspezifisch sind wie z.B. Werte mit denen Labordaten normalisiert

wurden. Im Bereich „Clinical Data“ werden die eigentlichen, während einer Studie erfassten Daten abgelegt [ODM]. CDMS Systeme bieten Schnittstellen für den Import/Export von CDISC-ODM. Darüber können eCRF-Spezifikationen und Daten zwischen den Systemen ausgetauscht werden.

Für die Erfassung und Weiterverwendung von Daten aus der Krankenversorgung in der Forschung insbesondere für klinische Studien (gemäß FDA 21 CFR Part 11) wurden von der CDISC eSDI-Arbeitsgruppe (eSource Data Interchange) 5 Szenarien entwickelt. Die elektronische Erfassung von Daten erfolgt im „Source at Site“-Szenario in einem CDMS durch manuelle Eingabe. Abweichend von den anderen Szenarien wird das CDMS hierbei nicht von einer dritten vertrauenswürdigen Partei betreut. Es findet keine Extraktion und Übernahme von Daten aus klinischen Informationssystemen z.B. dem KIS/KAS statt. Im „eSource System Provider (Contracted Supplier)“-Szenario wird das Forschungssystem zur Erfassung von Studiendaten von einer vertrauenswürdigen dritten Partei bereitgestellt. Die Dateneingabe erfolgt auch in diesem Szenario manuell. Eine Extraktion und Übernahme aus klinischen Informationssystemen ist ebenfalls nicht vorgesehen. Im „Direct Extraction from Electronic Health Records“-Szenario werden die Daten für Forschungszwecke im klinischen Informationssystem erfasst und dann daraus extrahiert. Im „Single Source“-Szenario erfolgt eine einmalige Eingabe der Daten. Die eingegebenen Daten werden sowohl an das CDMS als auch an die klinischen Informationssysteme weitergeleitet. Beim „Extraction and Investigator Verification“-Szenario erfolgt die Extraktion von Daten aus dem klinischen Informationssystem. Die Daten werden verifiziert und danach per EDC in ein CDMS übernommen [eSDI2005].

In der internationalen IHE-Initiative (Integrating the Healthcare Enterprise) [IHE] finden sich Vertreter aus medizinischen Fachgesellschaften, Regierungsbehörden, Verbänden und Hersteller von IT Systemen für das Gesundheitswesen. Konkret entwickelt IHE sog. IHE-Profile. Diese skizzieren Abläufe, welche mit IT-Systemen unterstützt werden sollen. Die Kommunikation zwischen den Systemen erfolgt dabei auf Grundlage existierender Protokollstandards wie HL7 und DICOM. IHE initiiert internationale Veranstaltungen auf denen die Hersteller der verschiedenen Systeme untereinander die Umsetzung der IHE-Profile testen können. Die IHE-Profile unterstützen u.a. Konzepte, die für das Management von Patientenidentitäten in verteilten Umgebungen herangezogen werden können. Ebenso stellt IHE Konzepte vor, die bei der Weiterverwendung von Daten aus der Krankenversorgung in der Forschung verwendet werden können.

Für die Übernahme von Daten, die im Rahmen der Behandlung von Patienten in den dafür vorgesehenen Anwendungen (KIS/KAS, Pathologieinformationssystem, RIS, LIS,…) erfasst werden, spezifiziert IHE das „Retrieve Form for Data Capture“ Profil (RFD) [RFD2007]. In RFD ist vorgesehen, dass Formulare, die aus einer Formularquelle entnommen werden, dem Endanwender angezeigt werden, damit dieser sie ausfüllen kann. Der Endanwender soll dazu allerdings keine spezielle Anwendung starten müssen, sondern die Eingabe der Daten soll in seiner Standardanwendung erfolgen. Zusätzlich sieht RFD vor, Daten die bereits eingegeben wurden bei erneutem Anzeigen des Formulars mit anzuzeigen. RFD beschreibt dazu folgende Komponenten. Der „Form Manager“ stellt die vom Anwender auszufüllenden Formulare bereit. Der „Form Filler“ empfängt vom „Form Manager“ die auszufüllenden Formulare. Der Form Filler hat die Möglichkeit Kontextinformationen zu einem Formular vom Form Manager anzufordern. Der „Form Filler“ kann Daten aus vollständig aber auch teilweise ausgefüllten Formularen zum Zwecke der Archivierung an den „Form Archiver“ übertragen. Der „Form Receiver“ verarbeitet Daten aus vollständig sowie aus teilweise ausgefüllten Formularen, die er vom „Form Filler“ erhält [RFD2007].

Das IHE „Cross-enterprise Document Sharing“ (XDS) dient zum Austausch von Dokumenten zwischen verschiedenen Institutionen des Gesundheitswesens. XDS beschreibt folgende Komponenten. Die „Document Source“ stellt die zu speichernden Dokumente bereit, beschreibt Dokumente mit Metadaten und übermittelt die Dokumente und Metadaten an das „Document Repository“. Dieses speichert die übermittelten Dokumente und erweitert die Metadatensätze der Dokumente um URIs für das (wieder) auffinden. Die Metadaten werden an die „Document Registry“ weitergeleitet und dort gespeichert. Die „Document Registry“ stellt Schnittstelle für Anfragen auf die Metadaten bereit. Über die „Document Consumer“ Komponente werden Anfragen gegen die Metadaten in der „Document Registry“ gestellt. Das Ergebnis der Anfrage umfasst eine Liste mit Einträgen bestehend aus Metadaten zu den Dokumenten. Nach Auswahl eines Metadatensatzes durch den Anwender erfolgt die Extraktion des Dokuments aus dem „Document Repository“ durch die „Document Consumer“ Komponente [XDS2010].

Für die Verwaltung von Patienten-IDs über Domänengrenzen hinweg stellt IHE das „Patient Identifier Cross Referencing“ (PIX) Integration Profile zur Verfügung. Der Begriff Domäne bezieht sich in diesem Zusammenhang auf den Namensraum, in dem eine Patienten-ID eindeutig ist. Damit kann eine Domäne sowohl aus einem

einzelnen System als auch aus mehreren untereinander kommunizierenden Systemen bestehen die einen einzelnen Patienten über eine gemeinsame ID identifizieren. Das Profil definiert drei Aktoren. Der „Patient Identity Source“ Aktor benachrichtigt den „Patient Identifier Cross-reference Manager“ über Ereignisse welche die Identifikation eines Patienten betreffen. Als Kommunikationsstandard wird hierzu HL7 eingesetzt. Im Profil definierte HL7 Ereignisse und damit assoziierten Nachrichten treten bei der Aufnahme, Änderung und Zusammenlegung von Patienten auf. Der „Patient Identifier Cross-reference Manager“ empfängt diese Nachrichten von Systemen aus verschiedenen Domänen, führt diese zusammen und stellt Mechanismen zur Beantwortung von Anfragen bereit. Der „Patient Identifier Cross-reference Consumer“ unterstützt Anfragen auf Basis von HL7 Query-Nachrichten [PIX2010].

Rechtliche Rahmenbedingungen und ethische Aspekte

Für die Erfassung von Daten für Forschungszwecke, der Weiterverwendung von Daten aus der Krankenversorgung in der Forschung sowie für forschungsbezogene Biomaterialsammlungen existieren verschiedene rechtliche Rahmenbedingungen die es zu berücksichtigen gilt. Einige rechtliche Aspekte lassen sich dem Ethik- und Datenschutzkonzept des Spitzenclusters m4 entnehmen [m4Ethik]:

Gemäß Art. 27 Abs. 4 Satz 1 Fall 3 BayKrG [BAYKRG2007] dürfen „Krankenhausärzte Patientendaten nutzen, soweit dies zu Forschungszwecken im Krankenhaus erforderlich ist“; allerdings gilt „dass die Patientendaten das Gewahrsam des Krankenhauses nicht verlassen dürfen“. Zu beachten ist auch: „Aus datenschutzrechtlichen, aber auch aus Akzeptanzgründen ist jedenfalls in den Fällen, in denen der Forschungszweck eine Nachbefragung erforderlich macht bzw. machen könnte, dringend die Einholung einer schriftlichen, informierten und freiwilligen Einwilligungserklärung der Patienten/Probanden anzuraten.“ (alle drei Zitate: 22. Tätigkeitsbericht des Bayerischen Landesbeauftragten für den Datenschutz, 2006 [TätigDS2006]). Es gilt generell nach §3a BDSG das Prinzip der „Datenvermeidung und Datensparsamkeit“.

Hieraus ergibt sich, dass im Falle einer standortübergreifenden Verwendung von Forschungsdaten sowie einer Erfassung von Daten, die klar über den Behandlungsfall hinausgehen, eine separate Forschungsdokumentation und eine Einwilligungserklärung erforderlich sind. Dies gilt umso mehr für forschungsbezogene Bioproben. Neben der Behandlungsdokumentation gibt es

somit auch eine Forschungsdokumentation, die auch eine Dokumentation von Patienteninformation und Einwilligung umfasst.

Ein typischer Fall ist es, dass eine standortübergreifende Daten- (und Bioproben-) Sammlung in einem Forschungsverbund eine solche getrennte Forschungsdokumentation erforderlich macht. Es ist denkbar (bzw. naheliegend), dass ein Teil der Behandlungsdaten auch für die Forschungsdokumentation benötigt wird. Eine Datenübernahme der Routinedaten („secondary use“) ist somit häufig sinnvoll bzw. anzustreben [Safran2007], [Weng2012], [Bloomrosen2008], [Powell2005], [PWC2009]. Sie benötigt einen Nachweis der Einwilligung und eine zuverlässige Identifikation des Patienten. Die resultierenden Anforderungen an das ID-Management sind komplex.

Die Komplexität wird dadurch gesteigert, dass eine Reihe gesetzlicher Vorschriften beachtet werden muss, die von der EU-Datenschutzrichtlinie über das Bundesdatenschutzgesetz (BDSG) zum Landesdatenschutzgesetz (LDSG) [LDSG] reicht, und die auch weitere Gesetze wie das Landeskrankenhausgesetz (LKHG) umfasst. Dabei besagt §40 Abs.2 BDSG, dass „personenbezogene Daten zu anonymisieren sind, sobald dies nach dem Forschungszweck möglich ist. Bis dahin sind die Merkmale gesondert zu speichern, mit denen Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren Person zugeordnet werden können. Sie dürfen mit den Einzelangaben nur zusammengeführt werden, soweit der Forschungszweck dies erfordert“ [BDSG1990].

Anonymisieren bedeutet nach BDSG das „Verändern personenbezogener Daten, so dass Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können“ [BDSG1990]. Eine Re-Identifikation solcher als „faktisch“ anonymisiert geltender Daten wird allerdings derzeit ständig einfacher, da bspw. bereits kurze DNA-Abschnitte für eine Zuordnung zu einer Vergleichsprobe und einer Person ausreichen. Generell sollte sichergestellt werden, dass eine Re-Identifikation des Patienten nicht durch die Zusammenführung von Daten oder durch einzigartige Eigenschaften der Daten erfolgen kann [Sweeney2002].

Pseudonymisieren ist das „Ersetzen des Namens und anderer Identifikationsmerkmale durch Kennzeichen zu dem Zweck, die Bestimmung des Betroffenen auszuschließen oder wesentlich zu erschweren“ [BDSG1990].

Die Empfehlungen des Nationalen Ethikrates zu Biobanken fordern, dass eine Trennung zwischen den die Betroffenen identifizierenden Daten einerseits und den Proben und übrigen Daten andererseits so früh wie möglich, spätestens aber bei Aufnahme in die Biobank erfolgen [m4Ethik].

Für die Pseudonymisierung von Daten im Rahmen der medizinischen Forschung und in Forschungsverbünden existieren Konzepte z.B. von der Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. (TMF) [Reng2006], s.a. [Kohlmayer2010]. Es wird eine zweistufige Pseudonymisierung vorgeschlagen bei der eine Trennung der Daten in identifizierende, medizinische und in Analysedaten vorgenommen werden soll. Diese drei Gruppen von Daten sind in verschiedenen Datenbanken zu speichern. Die drei Datenbanken wiederum sind drei organisatorisch getrennten Einheiten zugeordnet. Damit soll verhindert werden, dass eine einzelne Person alle Daten im Zugriff hat und diese zusammenführen könnte. Um pseudonymisierte Daten in der Forschung verwenden zu dürfen, müssen Einwilligungserklärungen der entsprechenden Patienten vorliegen, da pseudonymisierte Daten personenbeziehbar sind [Pommerening2005].

In diesen Einwilligungserklärungen kann u.a. festgelegt werden, welche Forschungsfragestellungen mit den Daten und Biomaterialien untersucht werden sollen, wer auf die Daten und Biomaterialien zugreifen darf, und an wen sie mit welcher Zielsetzung weitergegeben werden dürfen (z.B. an Kooperationspartner in Forschungsnetzen, aber auch an Externe oder sogar an Pharma- bzw. Bio-Tech Unternehmen). Eine zentrale Frage ist dabei die Zweckbestimmung der Erhebung und der Nutzung von Daten und Proben. Die Bandbreite der Möglichkeiten zu deren Definition reicht hierbei von der Festlegung eines spezifizierten Verwendungszwecks (specified Consent), über eine abgestufte Einverständniserklärung, bei der der Betroffene aus möglichen Optionen wählen kann, bis hin zu einer breiten Einwilligung (broad consent) [Lunshof2008]. Bei Biobanken werden hier i.a. geringe Einschränkungen angestrebt, denen Auflagen des Datenschutzes (keine „Vorratsdatenspeicherung“) und potentielle ethische Bedenken entgegenstehen. Eine Diskussion geht über den Rahmen dieser Arbeit hinaus, es wird exemplarisch auf das Ethik-und Datenschutzkonzept von m4 verwiesen [m4Ethik].

Vor der Erteilung der Unterschrift auf den Einwilligungserklärungen müssen Patienten / Probanden ausreichend informiert werden. Dies hat durch geeignete Personen, i.a. aus dem ärztlichen Bereich, zu erfolgen, was einen entsprechenden Personaleinsatz bedingen kann. Die unterzeichneten Einverständniserklärungen werden in Papierform archiviert und elektronisch verwaltet (u. U. sowohl im Routine-

als auch im Forschungssystem). Patienten haben das Recht, ihre Einwilligung zurückzuziehen, was ebenfalls über IT-Systeme unterstützt werden muss.

Von erheblicher Relevanz ist der Punkt, dass jede nicht anonyme Verwendung von Daten und Proben in der Forschung durch eine Einwilligung abgedeckt werden muss. Da sich Analysemethoden rasant entwickeln, stellt eine enge Zweckbindung potentiell ein Problem dar. In jedem Fall muss es aus IT-Sicht möglich sein, für die betrachteten Daten und Bioproben die zugehörige Einwilligungserklärung zu verwalten, was im Fall von Zusammenführungen ein erhebliches Problem darstellen kann [Malin2010], [Karp2008].

In m4 werden erhebliche Anstrengungen unternommen, dieses Problem durch konsentierte und flächendeckend eingesetzte Einwilligungserklärungen zu vermeiden. Zudem werden vorhandene Einwilligungserklärungen durch Statusangaben charakterisiert. Diese Arbeit verwendet diese Konzepte, wobei konform zu m4 der Schwerpunkt auf einer prospektiven Datenerfassung und Probensammlung liegt.

Behandlungsprozess

Ausgangspunkt für die Erfassung medizinischer Daten und Biomaterialien ist der Behandlungsprozess. Grundsätzlich lässt sich der Behandlungsprozess in einen ambulanten und einen stationären Teilprozess separieren. Beide Teilprozesse umfassen Prozessschritte, in denen administrative Aufgaben abgearbeitet werden, insbesondere bei Aufnahme, Entlassung und Verlegung. Die Dokumentation erfolgt hier im KIS System durch Pflegekräfte und Ärzte. Bei der Aufnahme eines Patienten im Krankenhaus wird nach dessen Identität im KIS System anhand von demografischen Daten bzw. Versicherungsnummern gesucht. Wird keine entsprechende Patientenidentität gefunden, erfolgt eine Neuanlage. Für die Patientenidentität werden demografische Daten und die für die Abrechnung benötigten Basisdaten in hochstrukturierter Form erfasst. Eine versehentliche Neuaufnahme führt zu Doppelaufnahmen aus der Probleme, insbesondere eine fehlende Vorgeschichte, resultieren. Dieser Fehler ist nicht selten; typische Ursache sind Namensänderungen und abweichende Schreibweisen. Werden solche Doppelaufnahmen entdeckt, erfolgt eine Zusammenführung der Personen-IDs und der damit assoziierten Fallidentitäten einschließlich der Daten. Weitaus gravierender wäre eine Fehlzuordnung zu einer falschen Personen-ID, was aber praktisch nicht vorkommt. „Fälle“ werden über stationäre Aufenthalte und ambulante Besuche definiert.

An die Aufnahme schließen sich Prozessschritte an, in denen die eigentliche Behandlung des Patienten vorgenommen wird. Hier entstehen Kreisläufe aus Diagnostik, Bewertung und Maßnahmen, aus denen sich erneute Bewertungen und Maßnahmen ergeben. Die elektronische Dokumentation der Behandlung erfolgt im KAS (also i.a. innerhalb des übergeordneten KIS). Befunde und Arztbriefe bilden den Kern der „elektronischen Patientenakte“, die ein Teil des KAS ist; sie wird in zunehmendem Maß erweitert um eingescannte Papierdokumente. Ein Teil der während der Behandlung eines Patienten erfassten Daten weist eine hohe Qualität bzgl. Strukturiertheit, Korrektheit und Vollständigkeit auf. Dies gilt insbesondere für die demographischen Daten und die Eckwerte des Falles (Aufnahmedatum, Entlassdatum, Datum des ambulanten Besuchs, Fallart). Medizinische Daten, die für Abrechnungszwecke benötigt werden, liegen ebenfalls in strukturierter Form vor. Sie werden in Deutschland mit Klassifikationssystemen erfasst. Primär geht es um Diagnosen gemäß der deutschen Version der „International Classification of Diseases“ ICD [ICD] sowie Maßnahmen nach dem „Operationen- und Prozedurenschlüssel“ [OPS]. Die hieraus abgeleiteten „Diagnosis Related Groups“ [DRG] bestimmen den Erlös. Die Erfassung auch dieser Daten geschieht wegen ihrer Abrechnungsrelevanz sorgfältig, zudem kann auch eine Überprüfung der Daten durch den medizinischen Dienst der Krankenkassen erfolgen. Für eine Weiterverwendung im Forschungskontext sind ICD und OPS wegen ihrer Abrechnungsorientierung und der Tatsache, dass es sich um (bewusst vergröbernde) Klassifikationen handelt, nur eingeschränkt bzw. für einen groben Überblick geeignet. Labordaten werden ebenfalls hochstrukturiert gespeichert, wobei zusätzlich auch Metadaten erfasst werden, bspw. Normbereiche, Abweichungen, Maßeinheiten und Parameter der verwendeten Laborverfahren. Auch im Rahmen der externen Qualitätssicherung werden strukturierte Daten erhoben (§ 137 SGB V), etwa im Bereich von Tumorerkrankungen (z.B. Kolonkarzinom). Neben den strukturiert während der Behandlung erhobenen Daten existieren große Datenmengen in Form von Freitexten. Dies betrifft insbesondere Befunde aus der Pathologie, Radiologie und Sonografie sowie Arztbriefe.

Die Dokumentation auf Papier spielt immer noch eine große Rolle. Dies gilt in besonderem Maß für juristisch relevante Dokumente wie Einwilligungserklärungen und unterzeichnete Aufklärungsbögen. Aber auch daneben existieren weiterhin Dokumente auf Papier, und es gibt Medienbrüche.

2 Anwendungsfälle

Grundlage der Entwicklung des im Rahmen dieser Arbeit vorgestellten Konzepts für das Identitätsmanagement und zur Übernahme von Daten in der translationalen Forschung bilden Rahmenbedingungen, Anwendungsfälle und Prozesse, die im Strukturprojekt „Data Integration System“ (DIS) des Münchener Biotech-Spitzenclusters m4 identifiziert wurden. In DIS wird eine IT-Infrastruktur zur Unterstützung translationaler Forschungsprozesse entwickelt. Ein Schwerpunkt liegt hierbei in der Erfassung von Biomaterialien. Die in diesem Umfeld ermittelten Anwendungsfälle werden in den folgenden Kapiteln erläutert.

2.1 Anwendungsfälle auf Geschäftsprozessebene

Auf Prozess- und Businessebene wurde der Anwendungsfall „Management von Daten für die translationale Forschung“ identifiziert. Dieser beschreibt das Erfassen, Speichern, Zusammenführen und Analysieren von Daten zu Biomaterialien und medizinischen Daten. Forscher, Ärzte und Dokumentationskräfte sollen beim Anlegen von Patienten- und Fallidentitäten in Forschungssystemen unterstützt werden, d.h. das Forschungssystem ist das führende System und die gesamte Datenerfassung bzw. –übernahme basiert auf der Einwilligungserklärung des Patienten. Zur Vorbereitung der Übernahme sollen Patienten- und Fallidentitäten transparent für den Anwender in verschiedenen IT- Systemen zur Unterstützung der Krankenversorgung gesucht werden können. Ausgehend von gefundenen und durch Anwender identifizierte Patienten- und Fallidentitäten, sollen die IDs des Patienten aus den klinischen Systemen, die Patientenstammdaten, die Fall-IDs und Fallinformationen automatisiert in Forschungssysteme übernommen werden. Notwendige Abgleiche und ggf. die Zuordnung zu Identitäten, die in dem führenden oder evtl. anderen zu integrierenden Forschungssystemen bereits angelegt sind, soll transparent für den Anwender erfolgen. Nach Übernahme der Identitäten in die Forschungssysteme soll die Erfassung von medizinischen Daten ermöglicht werden. Dazu sollen im Rahmen der Krankenversorgung bereits erhobene medizinische Daten zum Vorbefüllen von Formularen in Forschungssystemen genutzt werden. Hierfür sollen Formulare automatisiert in Forschungssystemen angelegt und den entsprechenden Patienten- und Fallidentitäten zugeordnet werden. Die selektive

Übernahme von Behandlungsdaten soll dann ebenfalls automatisiert in einem vordefinierten Umfang erfolgen.

Zusätzlich zur Erfassung von medizinischen Daten soll das Sammeln von Biomaterialien unterstützt werden. Hierfür sollen Identitäten von Bioproben und daraus erzeugter „Aliquots“ sowie dazugehörige Probenannotationen verwaltet werden. Aliquots sind Teilportionen der Probe, die analysiert oder weitergegeben werden können, ohne das die restliche Probe oder weitere Aliquots aufgetaut werden müssen. Das Management von Bioproben umfasst das Anlegen, Speichern sowie die konsistente Abbildung verschiedener Proben-IDs und zugehöriger Pseudonyme aufeinander, wobei Proben (und ihre IDs) wieder zu übergeordneten Fällen bzw. zu Personen (Probanden, Spendern, Patienten) gehören. Im ersten probenbezogenen Use-Case wird der Fall betrachtet, dass Proben von Anfang an als Forschungsproben genommen werden. Ein zweiter, in m4 ebenfalls relevanter Fall ist, dass erst im Laufe der Weiterverarbeitung ein Teil des Biomaterials als Forschungsmaterial ausgewiesen wird. Nicht betrachtet wird der in den USA an einigen Standorten wichtige dritte Fall, dass Material der Forschung zugeführt wird, das in klinischen Abläufen bzw. Labors nicht verbraucht worden ist („biomaterials otherwise lost“).

Für diese Anwendungsfälle lassen sich Randbedingungen identifizieren: Um den Persönlichkeitsschutz der Patienten sicherzustellen, soll der Zugriff auf identifizierende Patientendaten, medizinische Daten und Probenannotationsdaten in den Forschungssystemen nur von dazu berechtigten Personen durchgeführt werden können. Ebenso soll bei der Extraktion von Daten aus IT-Systemen zur Unterstützung der Behandlung gewährleistet werden, dass dies unter Wahrung von Berechtigungen geschieht. In Forschungssystemen erfasste Daten müssen pseudonymisiert und räumlich/organisatorisch aufgetrennt werden in medizinische Daten, Probendaten und identifizierende Daten [m4Ethik], [Ethik2010], [Reng2006]. Die organisatorische Trennung umfasst ein Treuhänderkonzept. In Deutschland hat sich bei Bioproben die Sicht durchgesetzt, dass Personen- und Proben-IDs doppelt zu pseudonymisieren sind.

2.2 Erfassungsprozess für medizinische Daten und Biomaterialien

Der zu etablierende Soll-Prozess dient zur Erfassung medizinischer Daten und Biomaterialien unter Weiterverwendung von Daten aus der Patientenbehandlung.

Das Management von Identitäten in den Teilprozessen erfolgt auf Grundlage der in dieser Arbeit entwickelten Konzepte und Implementierungen.

Der in Abb. 1 und Abb. 2 dargestellte Prozess zur Erfassung medizinischer Daten, Probendaten und Biomaterialien wurde am Institut für Medizinische Statistik und Epidemiologie in einem Team, in dem der Autor Mitglied war, entwickelt. Für die Prozessmodellierung wurde die BPMN-Notation um Elemente ergänzt, die eine kompaktere Beschreibung des Informationsflusses und eine Präsentation vor Medizinern ermöglichen. Der Schwerpunkt der Prozessbeschreibung liegt auf der Frage, welche Informationen in welcher Form zu welchen Identitäten erfasst werden, und wie die Abbildung der einzelnen IDs und Pseudonyme der Identitäten aufeinander erfolgt.

Im ersten Prozessschritt in Abb.1 erfolgt die Aufnahme des Patienten im Krankenhaus. Der bereits beschriebene Prozess der Vergabe oder Weiterverwendung von Personen- und Fall-IDs wird i.a. durch geschultes Verwaltungspersonal vorgenommen (Ausnahme: Notaufnahmen). Das KIS vergibt für die neu angelegte Patientenidentität eine systemweit eindeutige ID (IS-H-ID). Existiert bereits eine Patientenidentität im KIS wird diese weitergenutzt. Nachdem die Patientenidentität festgelegt ist erfolgt bei einer stationären Aufnahme das Anlegen eines entsprechenden Falls. Hierfür vergibt das KIS eine eindeutige Fall-ID (IS-H-Fall-ID). Bei einer ambulanten Aufnahme wird vor dem Anlegen eines neuen Falles überprüft, ob der Patient im aktuellen Quartal bereits anwesend war. In diesem Fall wird kein neuer ambulanter Fall im KIS eröffnet sondern die entsprechende Fallidentität mit der dazugehörigen Fall-ID weiterverwendet. Nach Aufnahme des Patienten im KIS werden Etiketten mit dem Namen des Patienten sowie dessen IS-H-ID bzw. IS-H-Fall-ID gedruckt.

Der darauf folgende Prozessschritt umfasst die Patientenaufklärung und Unterzeichnung oder Ablehnung der Einwilligungserklärung, typischerweise auf einer Station oder in einer Ambulanz. Aufklärung und Einwilligung werden auf einem unterzeichneten Papierformular dokumentiert. Dieses Dokument wird zur eindeutigen Identifikation mit einer IS-H-ID des Patienten versehen (Etikett mit IS-H Fall-ID, identifizierende Daten wie Name, Vorname, Geb. Datum, ggf. IS-H Personen-ID). Die Erfassung dass eine Einwilligungserklärung für den Patienten vorliegt kann optional im KAS dokumentiert werden.

Bei der Erfassung von Biomaterialien werden zwei Teilprozesse unterschieden. Sie werden im „Ethik- und Datenschutzkonzept der Biobank Alliance und des Data Integration System des m4 Spitzenclusters München“ [m4Ethik] folgendermaßen

beschrieben: „Im Teilprozess „Körperflüssigkeiten“ werden Blut, Urin und Liquor erfasst. Dazu wird die Probe entnommen und ein Probenbegleitbogen ausgefüllt. Hinsichtlich der Identifikation werden zwei Optionen unterschieden. In Option 1 wird der Probenbegleitbogen ebenso wie die Probe mit einer IS-H-ID versehen. Die Probe und der Probenbegleitbogen werden zur Sammelstelle („Hub“) transportiert. Dort erfolgt die Aliquotierung. Das Probengefäß wird verworfen. Die Aliquots tragen geheime Proben-Pseudonyme als Bezeichner (integriert in die Tubes) (geheimes Pseudonym: Zuordnung von Pseudonym zu Patienten bzw. deren Fällen, Dokumenten und Proben wird geheim gehalten; Zuordnung nur mit Hilfe des IT-Systems möglich). Sie werden eingelagert. Probenbegleitdaten, IS-H-IDs und Aliquot-IDs werden in das IT-System übertragen, die beiden letztgenannten mit Hilfe von (Rack-) Scannern. In Option 2 wird von Anfang an ein Proben-Pseudonym (mit Barcode) verwendet, sowohl als Aufkleber für das Probengefäß als auch auf dem Probenbegleitbogen, der zur verwechslungsgesicherten Handhabung auch mit der IS-H-ID versehen wird. Die Proben-Pseudonyme sind vorgeneriert (Drucker, m4-Kit). Die Probe und der Probenbegleitbogen werden zur Sammelstelle („Hub“) transportiert. Die Probe wird eingelagert. Die Probenbegleitdaten, die IS-H-IDs und Proben-Pseudonym werden in das IT-System übertragen, die beiden letztgenannten mit Hilfe von (Hand-)Scannern. Der Teilprozess „Gewebe“ beginnt im OP-Bereich, typischerweise im Krankenversorgungs-Kontext. In Option 1 werden das Gewebe und der Probenbegleitbogen mit einer IS-H-ID versehen und in die Pathologie transportiert. In der Pathologie wird entschieden, ob bzw. welches Material für die Forschung verwendet wird. Die Identifikation des für die Forschung zu verwendenden Materials erfolgt dann mittels Proben-Pseudonym. Hierzu werden lokal Barcodes generiert. Der Probenbegleitbogen wird bei diesem Teilprozess sukzessive ausgefüllt. Danach erfolgt die Einlagerung der Probe. Die Probenbegleitdaten, die IS-H-IDs und Proben-Pseudonyme werden in das IT-System übertragen, die beiden letztgenannten mit Hilfe von Scannern. In Option 2 wird von Anfang an ein Proben-Pseudonym (mit Barcode) verwendet, sowohl als Aufkleber für das Probengefäß als auch auf dem Probenbegleitbogen, der zur verwechslungsgesicherten Handhabung auch mit der IS-H-ID versehen wird. Das Proben-Pseudonym ist vorgeneriert (Drucker). Die Probe und der Probenbegleitbogen werden zur Sammelstelle („Hub“) transportiert. Die Probe wird eingelagert. Die Probenbegleitdaten, IS-H-IDs und Proben-Pseudonym werden in das IT-System übertragen, die beiden letztgenannten mit Hilfe von (Hand-) Scannern [m4Ethik]“.

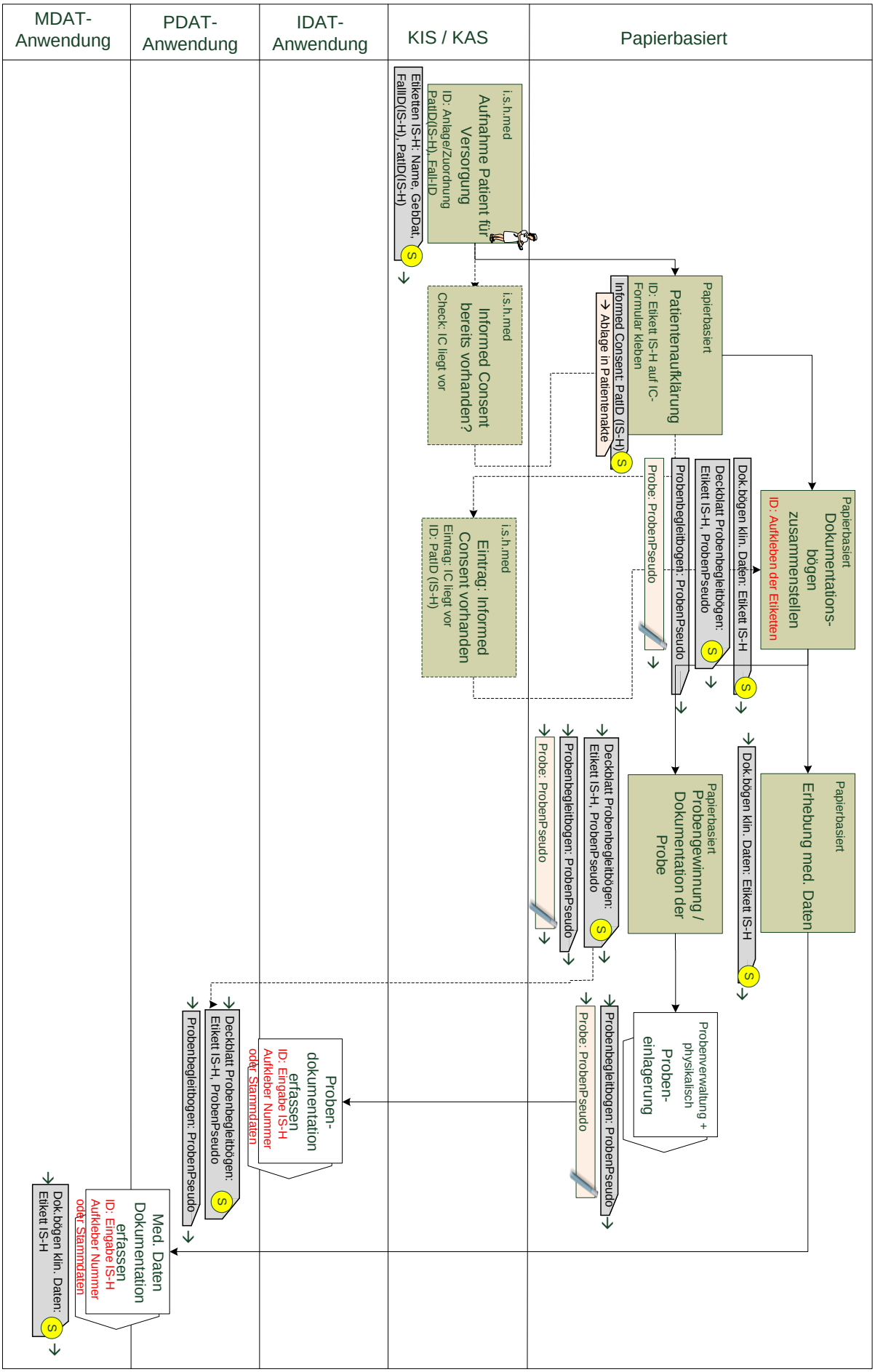
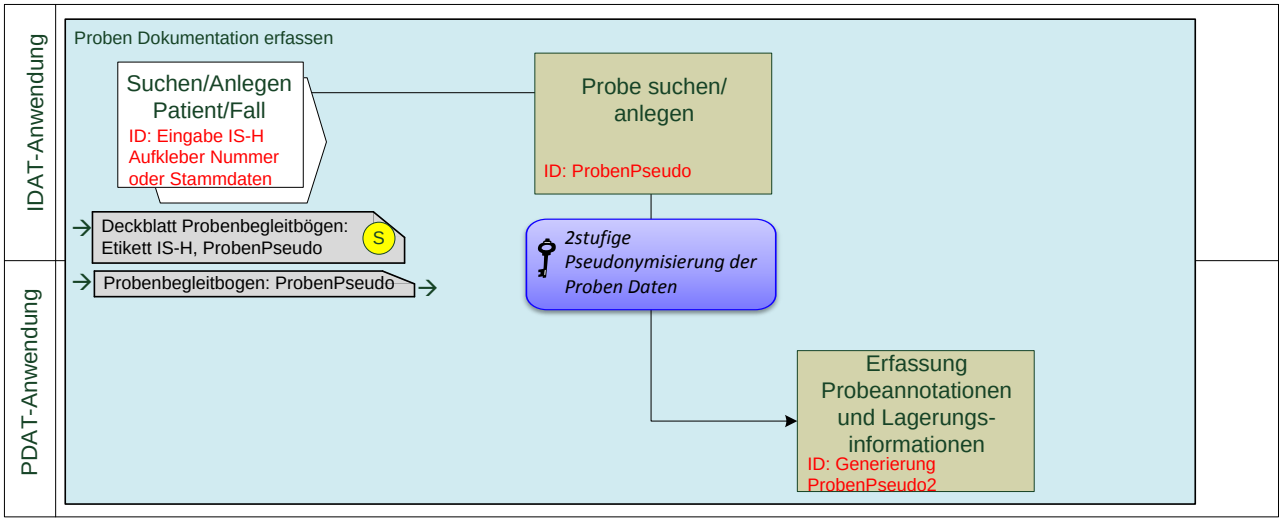
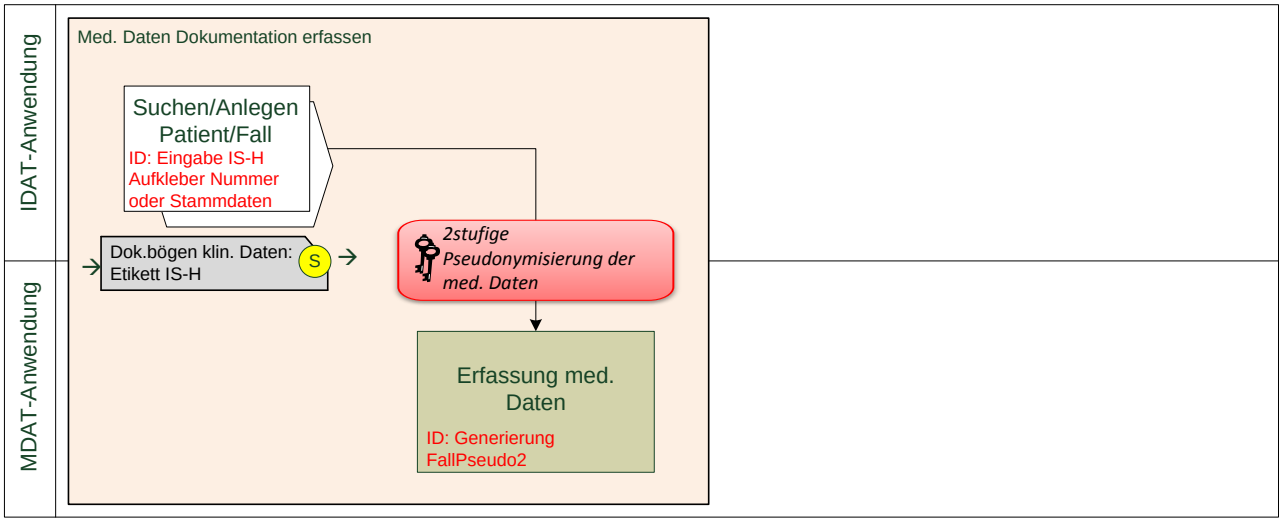
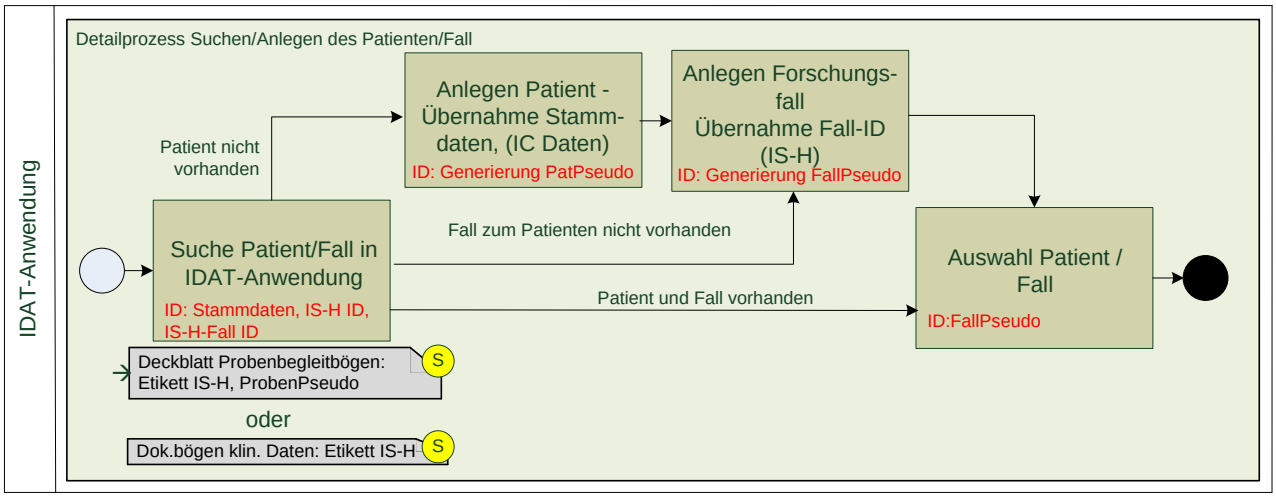


Abb. 1: Erweitertes BPMN-Model: Gesamtprozess zur Erfassung von Forschungsdaten und Biomaterialien


IDAT-Anwendung
PDAT-Anwendung
Proben Dokumentation erfassen
Suchen/Anlegen Patient/Fall
ID: Eingabe IS-H Aufkleber Nummer oder Stammdaten
Probe suchen/ anlegen
ID: ProbenPseudo
Deckblatt Probenbegleitbögen: Etikett IS-H, ProbenPseudo
S
Probenbegleitbogen: ProbenPseudo
2stufige Pseudonymisierung der Proben Daten
Erfassung Probeannotationen und Lagerungs-informationen
ID: Generierung ProbenPseudo2


IDAT-Anwendung
MDAT-Anwendung
Med. Daten Dokumentation erfassen
Suchen/Anlegen Patient/Fall
ID: Eingabe IS-H Aufkleber Nummer oder Stammdaten
Dok.bögen klin. Daten: Etikett IS-H
S
2stufige Pseudonymisierung der med. Daten
Erfassung med. Daten
ID: Generierung FallPseudo2


IDAT-Anwendung
Detailprozess Suchen/Anlegen des Patienten/Fall
Patient nicht vorhanden
Anlegen Patient - Übernahme Stamm-daten, (IC Daten)
ID: Generierung PatPseudo
Anlegen Forschungs-fall Übernahme Fall-ID (IS-H)
ID: Generierung FallPseudo
Suche Patient/Fall in IDAT-Anwendung
ID: Stammdaten, IS-H ID, IS-H-Fall ID
Fall zum Patienten nicht vorhanden
Patient und Fall vorhanden
Auswahl Patient / Fall
ID:FallPseudo
Deckblatt Probenbegleitbögen: Etikett IS-H, ProbenPseudo
S
oder
Dok.bögen klin. Daten: Etikett IS-H
S

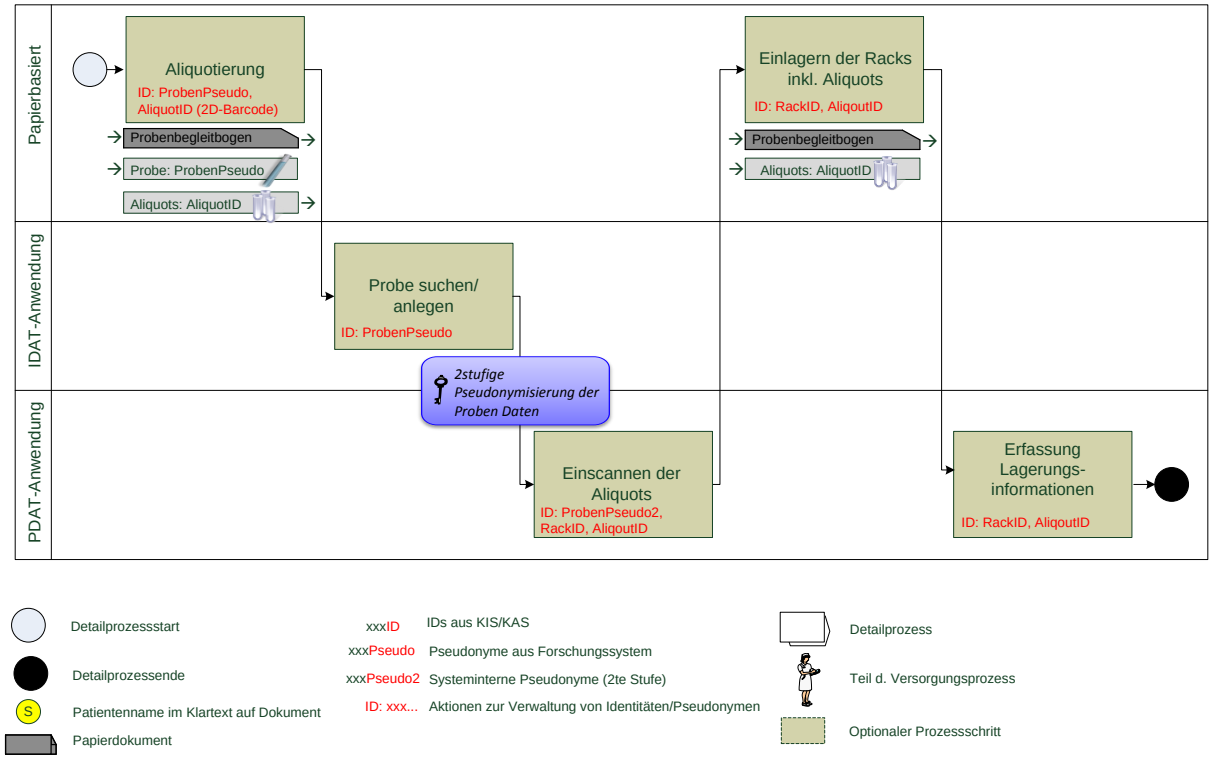


Abb. 2: Detailprozesse zur Erfassung von med. Daten, Probendaten und Biomaterialien

Für die Erfassung sucht die Dokumentationskraft den entsprechenden Patienten im Forschungssystem anhand der IS-H-Fall-ID. Existiert die Patientenidentität im Forschungssystem noch nicht, wird die Identität auf Basis der IS-H-Fall-ID im KIS transparent für die Dokumentationskraft gesucht und die vorhanden Stammdaten aus dem KIS zum Anlegen der Identität im Forschungssystem übernommen. Ist bereits die entsprechende Patienten- und Fallidentität im Forschungssystem angelegt worden, werden die Formulare zur Übernahme der auf Papier erfassten Daten instanziiert, der Fallidentität zugeordnet und ausgefüllt. Die verschiedenen Papierdokumente (Einverständniserklärung, Probenbegleitbogen, der evtl. IS-H-IDs und codierte Pseudonyme enthält, Med. Datenbögen) werden in einem abgesicherten Papierarchiv (Forschungsarchiv) verwahrt.

2.3 Systemspezifische Anwendungsfälle

Aus den Anforderungen an das Management von Identitäten unter Berücksichtigung des Datenschutzes bei der Erfassung medizinischer Daten und Biomaterialien ergeben sich systemspezifische Anwendungsfälle. Für eine Umsetzung dieser Anwendungsfälle ist es von zentraler Bedeutung auf IDs und Informationen zu einem Patienten- und deren Fallidentitäten aus im Routineeinsatz befindlichen klinischen Informationssystemen zugreifen zu können. Hierfür werden in dieser Arbeit

entsprechende Konzepte und Implementierungen erarbeitet. Dabei wird auf Arbeiten des Teams am IMSE, welches ein Pseudonymisierungskonzept und die dafür benötigten Dienste entwickelt und umgesetzt hat, aufgebaut.

Der erste in Abb. 3 dargestellte systemspezifische Anwendungsfall beschreibt das Management von Identitäten in der verteilten Forschungsumgebung unter Berücksichtigung des Datenschutzes. Der Anwender soll dazu Identitätsinformationen zu Patienten und deren Fällen aus IS-H [ISH] (KIS) und PASNet [PASNet] (Pathologieinformationssystem) über die Benutzerschnittstellen der Forschungssysteme suchen und automatisiert in diese übernehmen können. Das Suchergebnis soll nur Identitäten enthalten auf die der Anwender zugreifen darf. Es sollen maximal die Identitätsinformationen ausgelesen werden, auf die der Anwender bei direkter Nutzung der grafischen Benutzerschnittstellen von IS-H und PASNet Zugriff hat. Die Suche nach Patienten- und Fallidentitäten soll auf Basis von Stammdaten, IS-H-IDs und Fall-IDs in IS-H und PASNet ermöglicht werden. Finden sich in den durchsuchten Systemen Informationen zu Identitäten die den Zugriff weiter einschränken, so sollen diese mit in die Berechtigungsprüfung einbezogen werden. So können Einwilligungserklärungen im KIS/KAS für einen Patienten hinterlegt sein mit denen der Zugriff für Forschungszwecke auf ein definiertes Forschungsprojekt eingeschränkt wird.

Aus IS-H übernommene Informationen zu Patientenidentitäten sollen die IS-H-ID und die demografischen Daten (Vorname, Nachname, Adresse, Beruf, Konfession, Familienstand, etc.) des Patienten beinhalten. Für einen einzelnen Fall sollen die Informationen die Fall-ID, das Aufnahme- und Entlassdatum, den Typ des Falles (ambulant/stationär) und den Fallstatus umfassen.

Werden den Fallidentitäten medizinische Daten und Biomaterialien zugeordnet, so soll die Speicherung dieser Informationen im Einklang mit dem TMF-Konzept [Reng2006] bzw. dem Ethik- und Datenschutzkonzept von m4 [m4Ethik] in separaten Datenbanken erfolgen. Diese Datenbanken müssen räumlich und organisatorisch von der Anwendung zum Management von Patienten- und Fallidentitäten getrennt sein. Die Verknüpfung zwischen medizinischen Daten und Patienten-/Fallidentitäten soll nur mit Hilfe von Pseudonymen auf dem Computer des Anwenders erfolgen. Dazu muss die Anwendung zum Managen von Patienten- und Fallidentitäten Funktionalitäten zum Pseudonymisieren von Identitäten (PatPseudo, FallPseudo, ProbenPseudo) bereitstellen. Diese Pseudonyme müssen den Identitätsinformationen von Patienten und Fällen zugeordnet und persistiert werden können.

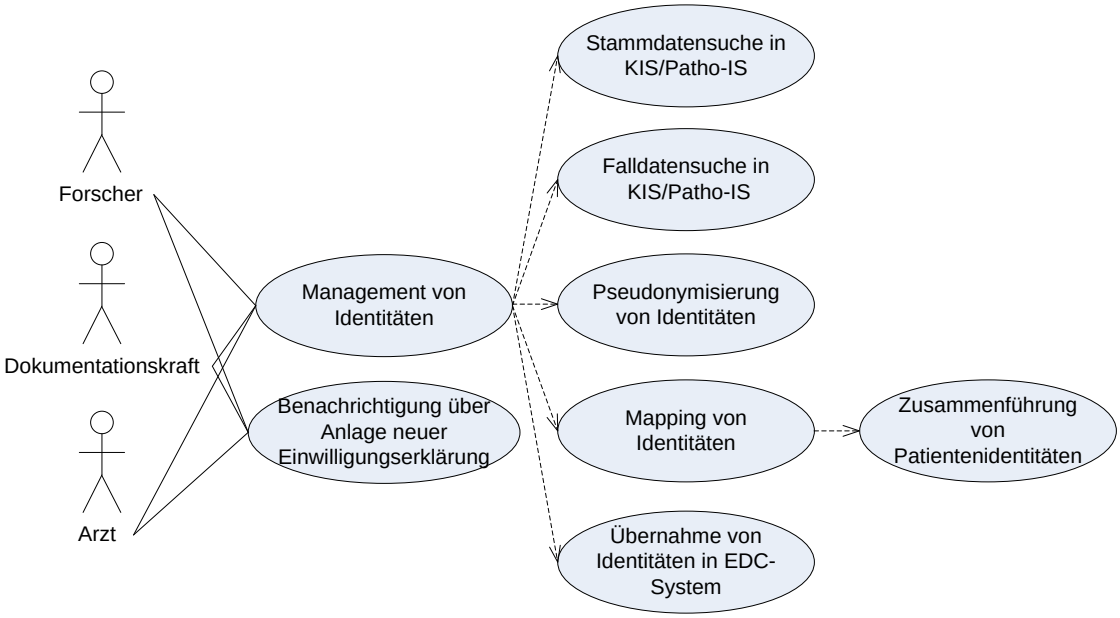


Abb. 3: UML Use Case Diagramm zur Beschreibung des Managements von Identitäten

Wird in IS-H festgestellt, dass für einen Patienten eine Doppelaufnahme durchgeführt wurde, so erfolgt die Zusammenführung der verschiedenen digitalen Identitäten dieses Patienten. Wurde der betroffene Patient ins Forschungssystem übernommen, so hat diese Zusammenführung der entsprechenden Patientenidentitäten im Forschungssystem automatisiert zu erfolgen. Hierfür haben die von IS-H versandten HL7-Nachrichten protokolliert und automatisiert ausgewertet zu werden. Zusätzlich soll es möglich sein, IS-H direkt anzufragen, ob für eine Patientenidentität eine Zusammenführung durchgeführt wurde und welche IS-H-ID als führend angegeben wurde. Dadurch sollen Zusammenführungen auch dann identifiziert werden können, wenn die Nachrichtenprotokollierung ausgefallen bzw. die Zusammenlegung zu Zeitpunkten stattgefunden hat an dem die Protokollierung nicht im Produktiveinsatz war (Altdatenübernahmen).

Der zweite systemspezifische Anwendungsfall umfasst die Benachrichtigung beim Anlegen von Einwilligungserklärungen für Patienten in IS-H/i.s.h.med [i.s.h.med]. Er wird in Abb. 3 dargestellt. Dabei hat eine Mitteilung über das Vorhandensein neu angelegter Einwilligungserklärungen an die Forschungssysteme zu erfolgen sobald ein Dokument für die Erfassung der Einwilligungserklärung in IS-H freigegeben wird. Die Mitteilung soll die IS-H-ID, den Vor- und Nachnamen sowie das Geburtsdatum des Patienten beinhalten. Weitere Informationen sollen den Typ, die Version und den Status der Einwilligungserklärung betreffen. Die Forschungssysteme sollen die Benachrichtigungen auswerten und den Inhalt in einer Arbeitsliste für den Anwender darstellen können.

Der dritte in Abb.4 modellierte systemspezifische Anwendungsfall beschreibt die Übernahme von Behandlungsdaten aus IS-H/i.s.h.med zum Vorbefüllen von Formularen in Forschungssystemen unter Berücksichtigung von domänenspezifischen Rahmenbedingungen. Dabei sollen verschiedene Szenarien unterschieden werden. Im ersten Szenario, sollen Daten aus Dokumenten zu einem Patienten übernommen zu werden, sobald die Dokumente in IS-H/i.s.h.med freigegeben werden. Der Zugriff auf die Daten für die Übernahme soll unter Berücksichtigung von Berechtigungen erfolgen. Dabei soll überprüft werden, ob der Anwender, der die Übernahme der Daten in das Forschungssystem initiiert, auch die Berechtigungen hat, die zur Übernahme anstehenden Daten im Quellsystem (KIS/KAS, Pathologieinformationssystem) einsehen zu dürfen. Ebenso hat eine Überprüfung der Berechtigungen im Forschungssystem zu erfolgen. Die Speicherung der ins Forschungssystem zu übernehmenden Daten soll nach identifizierenden-, medizinischen-, und Probendaten getrennt in verschiedenen Datenbanken erfolgen. Das zweite Szenario stellt einen Spezialfall des ersten dar. Dabei soll eine FDA 21 CFR Part 11 konforme Datenübernahme unterstützt zu werden. Hierfür hat vor der Speicherung der Daten im Forschungssystem eine Validierung der zu übernehmenden Daten durch eine autorisierte Person zu erfolgen. Die Übernahme muss im „Audit Trail“ des Forschungssystems protokolliert werden. Ebenso soll eine Protokollierung der Datenübernahme aus den klinischen Systemen in die Forschungssysteme erfolgen um deren Herkunft (Lineage) nachvollziehen zu können. Dabei soll vermerkt werden aus welchen Formularfeldern Daten aus den Quellsystemen, in welche Formularfelder der Zielsysteme übernommen werden. Mitberücksichtigt soll hierbei auch die Versionierung von Dokumenten auf Typ- und Instanzebene werden. Ebenso soll bei der Übernahme ganzer Dokumente deren Position innerhalb der Dokumentenreihenfolgen aufgezeichnet werden.

Der vierte systemspezifische Anwendungsfall beschreibt die „Konsistenzsicherung transferierter Daten“. Dabei soll zwischen Konsistenzsicherung auf Typ- und Instanzebene unterschieden werden. Auf Instanzebene sollen Änderungen an Daten im Forschungssystem welche aus IS-H/i.s.h.med und dem Pathologieinformationssystem übernommen wurden identifiziert werden. Diese Änderungen sollen in das KIS/KAS zurück propagiert werden. Änderungen an Daten in IS-H/i.s.h.med die bereits übernommene Daten betreffen, sollen im Forschungssystem angezeigt werden können. Änderungen auf Typebene (z.B. an Formularen) in i.s.h.med sollen automatisiert identifiziert werden können. Es sollen

Mechanismen bereitgestellt werden, mit denen diese Änderungen im Forschungssystem ggf. nachgezogen werden können.

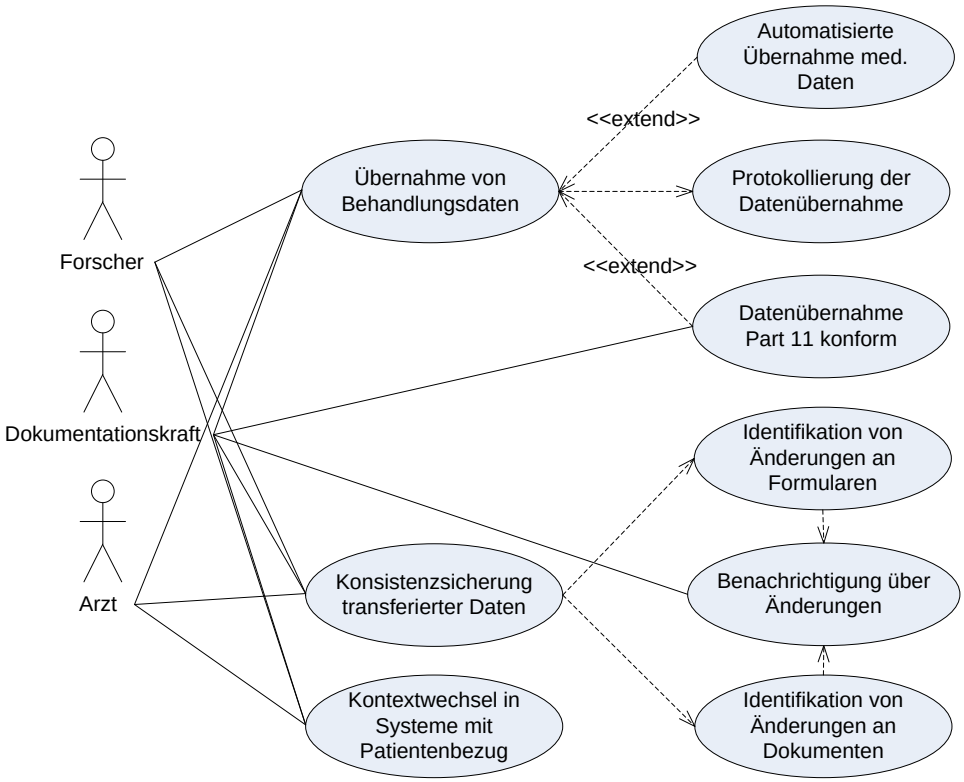


Abb. 4: UML Use Case Diagramm für die Übernahme von Behandlungsdaten und Konsistenzsicherung transferierter Daten

Der „Kontextwechsel in Systeme mit Patientenbezug“ zwischen Forschungssystemen untereinander sowie zwischen diesen und IT-Systemen aus der Krankenversorgung soll über die Clients der Systeme auf den Computern der Anwender erfolgen. Damit ist dieser Anwendungsfall einerseits zentral für die Umsetzung des TMF-Sicherheitskonzepts in dem Daten in getrennten Systemen gespeichert und erst auf den Computern der Anwender zusammengeführt werden sollen. Dazu müssen die graphischen Benutzerschnittstellen der verschiedenen IT-Systeme ggf. in einer Anwendung, die auf dem Computer des Anwenders ausgeführt wird, integriert werden. Anderseits ermöglicht der Kontextwechsel zwischen den Systemen Anwender eine einfachere Identifikation von Patientenidentitäten. Hierfür sollen, ausgehend von den Suchergebnissen der Patienten- und Fallidentitätssuche, aus den Forschungssystemen heraus die GUIs der IT-Systeme aus der Behandlung (KIS/KAS) geöffnet und die entsprechenden Patienten- und Fallidentitäten in diesen aufgerufen werden. Der Anwender kann dann Informationen zu Patienten- und Fallidentitäten im Originalkontext einsehen. Dies kann die Identifikation der gesuchten Patientenidentität durch den Anwender beschleunigen, falls das Suchergebnis nach Patienten- und Fallidentitäten mehrere Einträge umfasst. Daneben eröffnet der patientenzentrierte Kontextwechsel

zwischen Behandlungs- und Forschungssystemen die Möglichkeit, die Datenübernahme zwischen den Systemen durch Kopieren und Einfügen zu erleichtern. Dies wird dadurch erreicht, dass sichergestellt wird, dass sowohl im Quellsystem als auch im Zielsystem die korrespondierenden Patientenidentitäten zum selben Patienten ausgewählt sind. Der Kontextwechsel zwischen Systemen mit Patientenbezug erleichtert es Anwender die Plausibilität von übernommenen Daten zu überprüfen. Dazu werden ausgehend von den zu überprüfenden Daten in den Forschungssystemen die Behandlungssysteme automatisiert aufgerufen und die Patientenidentitäten von denen die Daten stammen ausgewählt. Die Zugriffe auf die Quellsysteme sollen mit den Zugangsdaten des Anwenders erfolgen welcher den Kontextwechsel initiiert hat. Damit soll sichergestellt werden, dass der Anwender nur auf die Daten im Quellsystem Zugriff erhält, auf die er auch bei direktem Aufruf des Systems zugreifen dürfte.

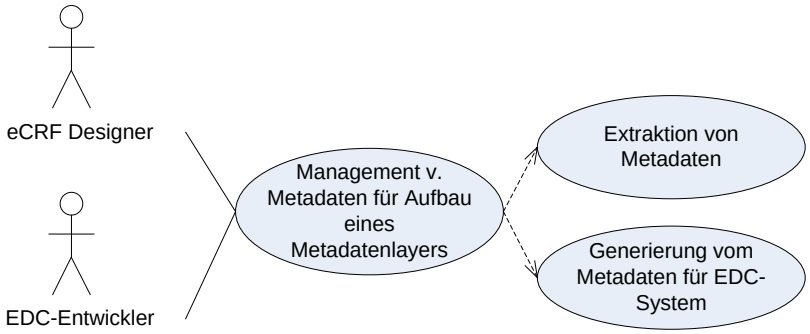


Abb. 5: UML Use Case Diagramm für das Management von Metadaten zum Aufbau eines Metadatenlayers für den Abgleich von Quell- u. Zielsystemen

Der in Abb. 5 modellierte systemspezifische Anwendungsfall beschreibt das Management von Metadaten für den Aufbau eines Metadatenlayers zum Abgleich von Quell- u. Zielsystemen (Forschungssysteme). Für die strukturierte Erfassung von Daten stellen EDC-Systeme eCRFs bereit. Die schnelle Entwicklung und Anpassung dieser eCRFs durch Anwender erfolgt mit, in EDC-Systeme integrierten, Formulargeneratoren. Beim Anlegen/Ändern eines Feldes auf einem eCRF werden u.a. der interne Name und Datentyp des Feldes, sowie bei Auswahllisten, die Vokabulare, die bei der Dateneingabe für das Feld zur Verfügung stehen sollen, angelegt. Weitere Informationen zu den Feldern betreffen die Beschriftung des Feldes im elektronischen Formular, sowie Angaben zu Wertebereiche und Formatierung der einzugebenden Daten. Aspekte, die sich aus Änderungen wie z.B. Schemaevolution, generische Schemata usw. ergeben, werden von den Formulargeneratoren transparent für die eCRF-Entwickler gekapselt. Für die strukturierte Datenerfassung in der Krankenversorgung existieren in den

Routinesystemen entsprechende Formulare. Werden die dort erfassten Daten übernommen so sollen die Entwickler der eCRFs in den EDC-Systemen Metadaten aus den klinischen Systemen extrahieren und zur Erstellung weiterverwenden können. Dazu sollen sie zu einzelnen Formularfeldern aus i.s.h.med deren externen Feldbezeichner, deren Datentyp, das Datenformat und ggf. hinterlegte Vokabulare extrahieren können. Weitere zu extrahierende Informationen sollen die Gruppierung einzelner Felder zu Gruppen und Formularen umfassen. Die aus i.s.h.med extrahierten Metadaten werden in CDISC-ODM transformiert und ins EDC-System eingelesen.

3 Konzept

3.1 Konzept zur Weiterverwendung von Daten aus der Behandlung in der medizinischen Forschung

Zentraler Bestandteil des hier vorgestellten Konzeptes ist die Weiternutzung von Daten zu Patienten- und Fallidentitäten aus Routinesystemen zur Unterstützung der Krankenversorgung insbesondere dem KIS. Dazu werden die Systeme um Such- und Extraktionsfunktionalitäten erweitert bzw. vorhandene genutzt. Die Suche nach Patientenidentitäten in den Systemen der Krankenversorgung wird initiiert wenn Forscher eine Patientenidentität im Forschungssystem anlegen. Um zu verhindern das Patientenidentitäten im Forschungssystem doppelt angelegt werden, erfolgt zunächst eine Suche nach dieser Identität im Forschungssystem. Die Suchanfrage gegen das Forschungssystem wird zusätzlich an das KIS weitergeleitet. Neben Stammdaten kann auch die innerhalb der Institution eindeutige Patienten-ID und die Fall-ID als Suchparameter übergeben werden. Diese IDs sind nicht geheim. Sie befinden sich häufig auf Aufklebern zusammen mit dem Namen des Patienten. Beim Zugriff auf das KIS im Rahmen der Suche erfolgt eine Überprüfung der Berechtigungen des Anwenders. Damit wird sichergestellt, dass nur Informationen zu Patientenidentitäten extrahiert werden auf die der Anwender auch zugreifen könnte wenn er direkt in den Routinesystemen eine Suche initiiert. Werden Patientenidentitäten entsprechend der übergebenen Suchparameter im KIS gefunden erfolgt vor der Präsentation des Suchergebnisses eine Überprüfung, ob das Vorhandensein von unterschriebenen Einwilligungserklärungen für die Weiterverwendung von Daten für Forschungszwecke im KIS bzw. KAS dokumentiert worden ist. Dazu wird eine entsprechende Suchanfrage im KIS bzw. KAS mit der einer Patientenidentität zugeordneten Patienten-ID durchgeführt. Liegt keine Einwilligungserklärung vom Patienten vor, so werden die demographischen Daten für den betroffenen Patienten aus dem Suchergebnis entfernt und durch einen entsprechenden Hinweis ersetzt. Der Anwender wählt aus dem Suchergebnis des KIS den Patienten aus. Für diesen wird dann im Forschungssystem eine Identität erzeugt und die im KIS vorhandenen demographischen Daten übernommen.

Beinhaltet die Suchanfrage eine Fall-ID, so werden zusätzlich zur Patientenidentität auch Informationen zur gesuchten Fallidentität extrahiert.

Findet eine Zusammenführung mehrerer Patientenidentitäten eines Patienten im KIS statt wird überprüft, ob davon auch Identitäten im Forschungssystem betroffen sind. Ist dies der Fall, so erfolgt auch eine Zusammenlegung der entsprechenden Identitäten im Forschungssystem. Um festzustellen, dass eine Zusammenlegung im KIS stattgefunden hat, werden vom KIS versandte HL7 Nachrichten nach Ereignissen gefiltert die dies anzeigen. Zusätzlich wird das KIS um eine Schnittstelle erweitert über die angefragt werden kann, ob eine Patientenidentität mit einer anderen Patientenidentität zusammengelegt wurde und welche der beiden die führende ist.

Neben der Weiterverwendung von Patientenidentitäten aus Routinesystemen der Krankenversorgung werden auch medizinische Daten in Forschungssysteme transferiert um dort Formulare automatisiert vorzubefüllen. Für die Initiierung des Datentransfers sind zwei Ansätze vorgesehen. Im ersten Ansatz findet die Übernahme statt nachdem eine Patienten- und falls notwendig eine Fallidentität im Forschungssystem durch den Anwender angelegt wurde. Die Initiierung im zweiten Ansatz erfolgt nachdem in einem IT-System zur Unterstützung der Krankenversorgung einem Patienten ein Dokument eines bestimmten Typs zugeordnet und freigegeben wurde. Für die Extraktion der medizinischen Daten werden die Routinesysteme um entsprechende Schnittstellen erweitert. Bei der Extraktion findet eine Überprüfung der Berechtigungen statt. Die Übernahme erfolgt nur für Anwender die auf die zu transferierenden Daten im Routinesystem zumindest lesenden Zugriff haben. Bei der Datenextraktion aus dem KAS wird der Zugriff in diesem protokolliert.

Das Forschungssystem in das Daten zu Patientenidentitäten, Probendaten und medizinische Daten aus IT-Systemen der Krankenversorgung übernommen werden, besteht aus mehreren Teilsystemen. Die Teilsysteme sind getrennten Organisations-einheiten unterstellt. Für die Speicherung von Patientenidentitäten (IDAT-Teilsystem), medizinischen Daten (MDAT-Teilsystem), Probendaten (PDAT-Teilsystem) und Pseudonymen ist jeweils ein Teilsystem vorgesehen. Mit Hilfe der Pseudonyme werden die Patientenidentitäten mit den dazugehörigen medizinischen und Probendaten verknüpft. Die eigentliche Zusammenführung der Daten aus den einzelnen Teilsystemen erfolgt erst auf dem Computer des Anwenders in einem Web-Browser. Hierfür werden die graphischen Benutzerschnittstellen der einzelnen Teilsysteme auf Präsentationsebene integriert.

Eine detailliertere Beschreibung des Prozesses der Übernahme von Patientenidentitäten, medizinische und Probendaten einschließlich des Managements der Pseudonyme findet sich in Abb. 6.

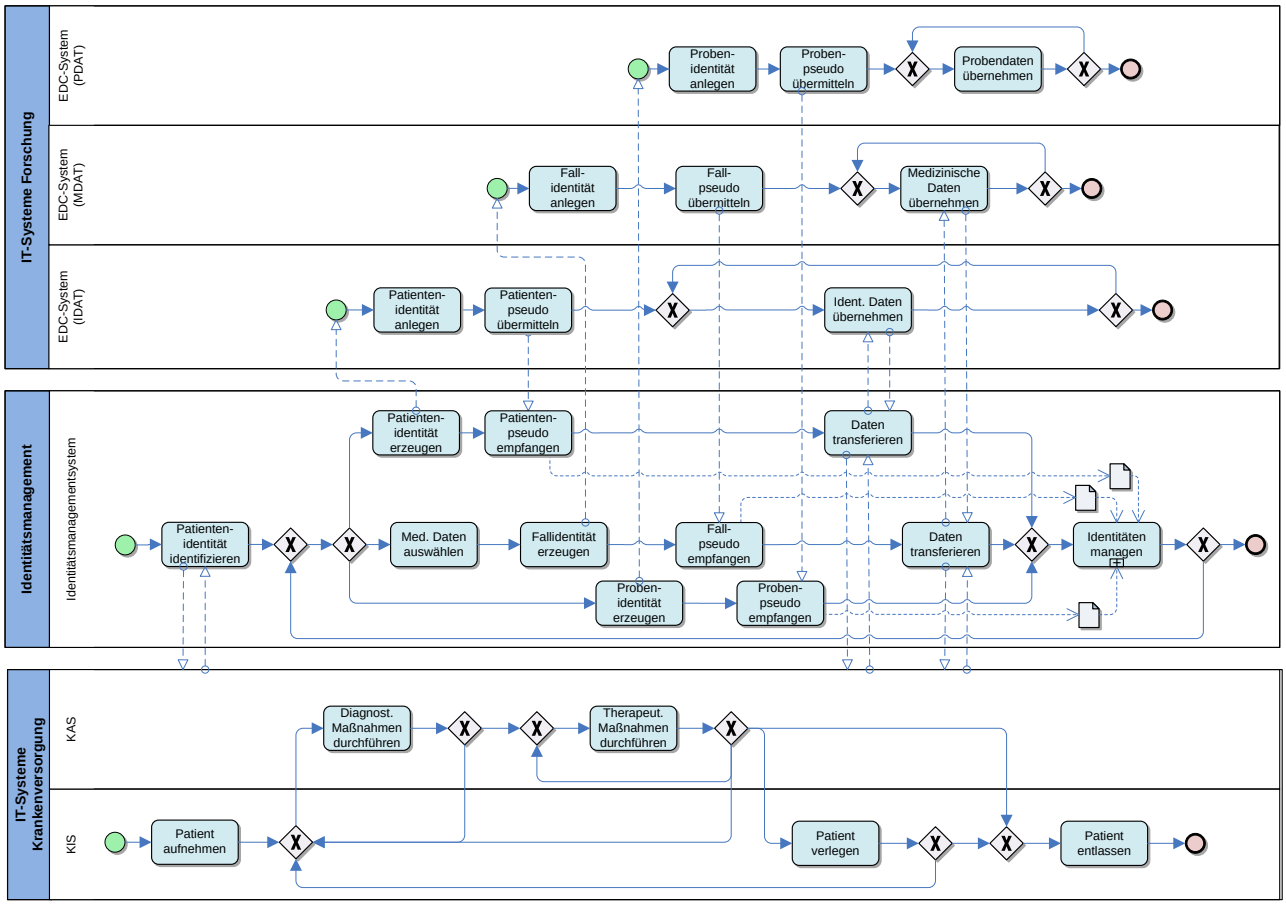


Abb. 6: BPMN-Model zur Beschreibung des Identitätsmanagementprozesses zur Weiterverwendung von Daten aus der Krankenversorgung

In den unteren beiden Lanes findet sich eine abstrakte Prozessbeschreibung des Behandlungsprozesses. Der Anwender sucht im KIS nach der Patientenidentität mit Hilfe der innerhalb eines Klinikums bekannten Patienten-ID, Fall-ID bzw. demografischen Daten. Nachdem die Patientenidentität des Patienten im KIS gefunden und festgestellt worden ist, dass das Vorhandensein einer entsprechenden Einwilligungserklärung dort dokumentiert wurde, wird eine neue Patientenidentität im Teilsystem zur Verwaltung der identifizierenden Daten (IDAT) angelegt. Im Rahmen des Identitätsmanagements wird das vom (IDAT)-Teilsystem der neu angelegten Patientenidentität zugeordnete Patientenpseudonym gespeichert. Nachdem die neue Patientenidentität im (IDAT)-Teilsystem angelegt worden ist, wird für diese eine neue Fallidentität im (MDAT)-Teilsystem erzeugt. Die neue Fallidentität im (MDAT)-Teilsystem verfügt über die aus dem KIS übernommenen Informationen. Dies umfasst ein Pseudonym der Fall-ID, den Typ

des Falles sowie Datumsangaben. Im Identitätsmanagement wird dem Pseudonym der Patientenidentität aus dem (IDAT)-Teilsystem ein Fallpseudonym zweiter Stufe der neu im (MDAT)-Teilsystem angelegten Fallidentität zugeordnet. Dazu wird das für die Fallidentität im (MDAT)-Teilsystem vergebene Fallpseudonym von einem Pseudonymisierungsdienst auf ein neues Pseudonym abgebildet. Für die Erfassung von Biomaterialien wird im (PDAT)-Teilsystem eine Einsendungsidentität angelegt. Dieser werden Informationen der korrespondierenden Einsendungsidentität aus dem Pathologieinformationssystem zugeordnet (Pseudonym der Einsendungs-ID, Datumsangaben, etc.). Das Pseudonym zweiter Stufe für die neu angelegte Einsendungsidentität wird im Identitätsmanagement dem entsprechenden Pseudonym zweiter Stufe einer Fallidentität zugeordnet; wie beschrieben erfolgen in der realen Welt Probenentnahmen bzw. Einsendungen im Rahmen von Fällen. Für einzelne Bioproben bzw. Aliquots werden Identitäten im (PDAT)-Teilsystem angelegt. Diese Identitäten sind der entsprechenden Einsendungsidentität zugeordnet. Die Proben-ID im System ist dabei identisch mit der ID auf dem Etikett der Bioprobe bzw. der Aliquots.

Werden im Rahmen einer erneuten Aufnahme eines Patienten im Krankenhaus weitere Proben gesammelt bzw. wird eine Follow-Up Dokumentation durchgeführt, so sucht der Anwender im (IDAT)-Teilsystem anhand der demographischen Daten, der Patienten- oder Fall-ID aus dem KIS nach dem Patienten. Wird keine Patientenidentität im (IDAT)-Teilsystem gefunden, da lediglich die Fall-ID eines neu im KIS angelegten Falles als Suchparameter übergeben wurde, erfolgt eine Weiterleitung der Suchanfrage an das KIS. Mit der im KIS gefundenen Patientenidentität wird die korrespondierende Identität im (IDAT)-Teilsystem identifiziert. Dazu erfolgt ein Abgleich der Patienten-ID zwischen der gefundenen Identität und denen die im Forschungssystem vorhanden sind. Patientenidentitäten werden wie oben beschrieben Identitäten von Fällen und Einsendungen in den entsprechenden Teilsystemen zugeordnet.

3.2 Verwandte Arbeiten und Abgrenzung

In der Literatur finden sich Ansätze in denen ebenfalls darauf fokussiert wird Patientenidentitäten unter Berücksichtigung domänenspezifischer Aspekte wie deren besonderer Schutzbedürftigkeit und der Einhaltung von Datenschutzvorgaben zu verwalten.

Peyton und Hu beschreiben ein Framework für ein föderiertes Identitätsmanagement für die Krankenversorgung [Peyton2010]. Mit Hilfe des Frameworks werden Patientenidentitäten aus verschiedenen Systemen verwaltet und die mit den Identitäten assoziierten Daten in einem föderierten Warehouse integriert. Hierzu werden in einem Master Patient Index für jede Patientenidentität demographische Daten und Pseudonyme gespeichert. Der MPI ist in Form eines „Liberty Alliance CoT"-Identitätsproviders realisiert. Vor der Aufnahme eines neuen Patienten im MPI wird überprüft, ob dort bereits eine Identität von ihm existiert. Hierzu kommt ein spezielles Record Linkage Verfahren (DB2 Anonymous Resolution) zum Einsatz. Existiert für den Patienten im MPI ein Eintrag, so wird diesem ein neues Pseudonym zugeordnet. Anderenfalls werden eine neue Identität und ein Pseudonym im MPI angelegt. Bevor die mit den Patientenidentitäten assoziierten medizinischen Daten aus unterschiedlichen Datenquellen ins Warehouse geladen werden findet ein Konsolidierungsprozess statt. Dabei werden die zu einem Patienten gehörenden Daten de-personalisiert, pseudonymisiert. Bei der Pseudonymisierung erfolgt die Zuordnung des im MPI hinterlegten Pseudonyms zu den Daten des Patienten [Peyton2010].

Au und Croll stellen ein Konzept vor, in dem Patienten festlegen können, ob ihre Patientenidentitäten und die damit assoziierten medizinische Daten aus unterschiedlichen Krankenhäusern zusammengeführt werden dürfen. Dazu vergibt eine vertrauenswürdige dritte Stelle eindeutige Identifikatoren, die sowohl dort als auch auf einem dem Patienten gehörenden Device gespeichert werden. Die vertrauenswürdige Stelle kann Zertifikate ausstellen mit denen die Verknüpfung zweier Identifikatoren bestätigt wird. Jeder elektronischen Akte eines Patienten in unterschiedlichen Einrichtungen ist einer dieser eindeutigen Identifikatoren zugewiesen. Sollen zwei Patientenakten miteinander verknüpft werden, so erstellt der Patient über die vertrauenswürdige Stelle ein entsprechendes Zertifikat. Danach kann die Verknüpfung der Akten erfolgen [Au2008].

Deng et al. beschreiben einen Ansatz für das Management von Identifikatoren, auf dessen Basis Identitäten eines Patienten und die damit assoziierten medizinischen Daten in unterschiedlichen Einrichtungen der Krankenversorgung zusammengeführt werden können [Deng2008], [Deng2009]. Jede Einrichtung verfügt über einen Identity Provider der eine globale ID des Patienten in einen Identifikator der lokal verwendet wird konvertieren bzw. zurückkonvertieren kann. Eine vertrauenswürdige dritte Stelle orchestriert die Konvertierung der Identifikatoren zwischen den Identity Providern der verschiedenen Einrichtungen. Dazu speichert sie Verknüpfungen

zwischen der globalen ID und den Einrichtungen in denen der Patient über eine elektronische Akte verfügt [Deng2009].

Neben Arbeiten deren Schwerpunkte auf dem Management von Identitäten in der Krankenversorgung liegen finden sich in der Literatur auch verschiedene Arbeiten, bei denen das Identitätsmanagement im Rahmen von Datenübernahmen oder der Zusammenführung von Datenbanken in der translationalen Forschung von Relevanz ist. Dabei werden medizinische Daten aus der Krankenversorgung in Forschungswarehouses übernommen. Für das Management von Patientenidentitäten werden verschiedene Konzepte umgesetzt.

Roden et al. stellen mit dem an der Vanderbilt University in Nashville TN entwickelten „Synthetic Derivative“ (SD) einen Ansatz vor, in dem Daten aus der Krankenversorgung in die Forschung übernommen und mit Biomaterialproben assoziiert werden. Die Erfassung der in das SD-System übernommenen Daten erfolgt im Rahmen der Krankenversorgung im „Electronic Medical Record“ (EMR)-System. Im Gegensatz zu der in dieser Arbeit vorgestellten selektiven Übernahme von hochstrukturierten Daten aus IT-Systemen zur Unterstützung der Krankenversorgung, werden in das SD-System ein Großteil der zu einem Patienten zur Verfügung stehenden Daten aus dem EMR übernommen. Vor der Datenübernahme in das SD-System erfolgt allerdings eine De-Identifikation der Daten. Um gemäß der „Health Insurance Portability and Accountability Act“ (HIPPA) [HIPPA] Definition de-identifizierte Patientendaten zu erhalten, werden die in den HIPPA-Datenschutzbestimmungen aufgeführten Attribute manipuliert. Dies sind u.a. die Sozialversicherungsnummer, die Namen des Patienten, Fax-/Telefonnummern, E-Mail Adressen, URLs sowie die Namen des an der Behandlung beteiligten Personals. Zusätzlich werden alle Datumsangaben innerhalb der Records zwischen einem und 365 Tagen in die Vergangenheit umdatiert. Die Verschiebung innerhalb eines Records wird für alle Datumsangaben konstant durchgeführt um zeitbezogene Analysen weiterhin zu ermöglichen [Roden2008]. Um Daten, die nach dem Übernahmezeitpunkt im EMR-System erfasst werden, dem richtigen Patienten im SD zuordnen zu können, wird ein Konzept umgesetzt das sich von dem in dieser Arbeit vorgestellten wie folgt unterscheidet. Die Verknüpfung erfolgt in dem einer Patientenidentität im SD anstelle der EMR-ID bzw. eines einfachen Pseudonyms der Wert, der sich durch Anwenden einer Einweg-Hashfunktion auf die EMR-ID errechnet (Synthetic Derivative Identifier (SD-ID)), zugeordnet wird. Dadurch können Änderungen an Datensätzen im EMR-System auch nach Übernahme und De-Identifikation an den entsprechenden Datensätzen im SD nachgezogen werden. In

dem in der Arbeit vorgestellten Konzept erfolgt die Verknüpfung von Biomaterialproben und Patientenidentität über ein zweistufiges Pseudonymisierungsverfahren. Im Gegensatz dazu werden Verknüpfungen zwischen DNA-Proben und Patientenidentität im SD-System hergestellt, in dem den Proben ebenfalls der Hashwert aus der EMR-ID des entsprechenden Patienten zugeordnet wird. Ein weiterer Unterschied zwischen der Vorgehensweise von Roden et al. und der in dieser Arbeit vorgestellten, ergibt sich aus den rechtlichen Rahmenbedingungen. So wird durch die De-Identifikation der Daten im SD erreicht, dass die Forschung mit diesen Daten und den zugeordneten Biomaterialien nicht als Forschung am Menschen klassifiziert wird [Roden2008]. Dadurch ist es möglich Proben und Daten ohne formale Einwilligungserklärungen der Patienten zu erfassen. Daraus resultieren u.a. folgende Einschränkungen: Es ist nicht erlaubt Patienten zu re-identifizieren um mit ihnen Kontakt aufzunehmen; Daten aus anderen Einrichtungen können denen im SD nicht zuordnet werden; Im De-Identifikationsschritt werden geografische Informationen gelöscht. Daher sind Forschungsfragestellungen in denen diese Informationen benötigt werden unmöglich [Langanke2011]. Mit dem „Research electronic data capture“ (REDCap)-System wird in Vanderbilt die strukturierte Erfassung von Daten zum Zwecke der Forschung unterstützt. eCRFs lassen sich sowohl über einen Formulargenerator als auch über die Angabe von Metadaten in einer Excel-Formatvorlage, welche von REDCap-Administratoren schnell in Formulare umgesetzt werden können, definieren [Harris2009]. Im Unterschied zu dem in dieser Arbeit vorgestellten Konzept wird kein Management von Identitäten zwischen IT-Systemen zur Unterstützung der Krankenversorgung und REDCap unterstützt. Eine getrennte Speicherung von identifizierenden Daten und Forschungsdaten ist in REDCap ebenfalls nicht vorgesehen.

Das am Ohio State University Medical Center entwickelte „De-identified Information Warehouse“ (DIW) dient dazu, Forschern Abfragemöglichkeiten auf Daten aus der Krankenversorgung zu ermöglichen. Die De-Identifikation der Daten findet im ETL-Prozess des DIWs statt. Die Abfrageergebnisse können direkt an die anfragenden Forscher zurückgegeben werden, ohne dass sie vorher durch eine vertrauenswürdige dritte Stelle de-identifiziert und ggf. aggregiert werden müssen. Das Schema des DIWs soll dem Schema des in der Krankenversorgung eingesetzten Warehouses ähneln, um Anwendern die mit diesem vertraut sind die Formulierung von Anfragen zu erleichtern. Die im DIW enthalten Daten sind vollständig de-identifiziert [Erdal2012]. Das Management der verschiedenen

Identifikatoren eines Patienten (EMR-ID, ambulante/stationäre Fall-IDs) aus dem Quellsystem erfolgt sowohl im ETL-Prozess als auch beim Generieren des Abfrageergebnisses. Im DIW werden verschiedene Sichten für die Abfragen bereitgestellt. Beinhalten unterschiedliche Suchergebnisse auf unterschiedliche Sichten Daten desselben Patienten, so sind die Identifikatoren des Patienten jeweils unterschiedlich. Um die Identifikation des Patienten noch weiter zu erschweren werden im Suchergebnis, das dem Anwender präsentiert wird, die Identifikatoren des Patienten von der jeweiligen Anwendersession im DIW abhängig gemacht. So sind die Identifikatoren eines Patienten bei gleicher Anfrage zwischen unterschiedlichen Logins des Anwenders am DIW unterschiedlich. Das erste Mapping der Identifikatoren des Patienten erfolgt im ETL-Prozess. Hierbei wird für die IDs eines Patienten aus dem Quellsystem ein Hashwert berechnet und diesem eine Zufallszahl zugeordnet. Dieser Zufallszahl werden für jede Sicht weitere Zufallszahlen zugewiesen. Vor der Präsentation des Abfrageergebnisses für den Anwender, wird in einem zweiten Mappingschritt jeweils eine in Abhängigkeit von der Session erzeugte Zufallszahl addiert. Datumsangaben werden im De-Identifikationsprozess umdatiert. Hierfür wird für jeden Patienten in einer Sicht ein zufälliger Offset generiert, um den anschließend alle Datumsangaben zu einem Patienten vor Präsentation der Suchergebnisse geändert werden. Informationen in Freitextform werden nicht in das DIW übernommen [Erdal2012].

In der in Havard entwickelten „Informatics for Integrating Biology & the Bedside“ (i2b2)-Plattform werden Daten aus den verschiedenen in der Krankenversorgung eingesetzten IT-Systemen repliziert. Die replizierten Daten können dann in forschungsfragestellungsspezifische Datensammlungen übernommen werden. Die Datensammlungen selbst werden wiederum in i2b2 angelegt. Im Rahmen des Übernahmeprozesses in die Forschungsdatensammlungen erfolgt die De-Identifikation der Daten. Hierfür stellt die i2b2 Plattform Werkzeuge zur Verfügung. Die in den Forschungsdatensammlungen enthaltenen Daten können dann mit den in i2b2 integrierten Datenanalysewerkzeugen ausgewertet werden. Für die Formulierung von Abfragen stellt i2b2 eine grafische Benutzerschnittstelle bereit [Murphy2010]. Das Identitätsmanagement erfolgt in einer separaten Komponente. Diese Komponente kann Kontaktdaten für einen Patienten wie dessen Name, E-Mail-Adresse oder Sozialversicherungsnummer zwischenspeichern. Daneben stellt die Komponente Schnittstellen zur Verfügung, über die Anfragen nach Kontaktdaten an EMR-Systeme weitergeleitet werden können. Ausgehend von den Abfrageergebnissen auf den Forschungsdaten sollen Studienärzte mit Hilfe der in

der Komponente gespeicherten Daten in die Lage versetzt werden, Kontakt mit Patienten aufzunehmen und diese für klinische Studien zu rekrutieren. Zusätzlich kann über die Identitätsmanagementkomponente eine Verknüpfung zwischen dem im Rahmen des De-Identifikationsprozesses einer Patientenidentität zugeordneten i2b2-internen Identifikators und der EMR-ID des Patienten im EMR-System hergestellt werden [Mendis2012].

Eine Alternative zur Extraktion von Patientenidentitäten aus Routinesystemen der Krankenversorgung auf Basis von ETL-Prozessen besteht darin, Nachrichten die zwischen den Systemen ausgetauscht werden auszuwerten. Im „Greifswald Approach to Individualized Medicine“ (GANI_MED) [Schack2010], [Langanke2011] werden dafür demographische Patientendaten aus HL7 Nachrichten extrahiert und in einer Datenbank, die einer Treuhänderstelle untersteht, gespeichert [Schack2010]. Bekanntestes Beispiel für einen solchen Treuhänderansatz ist das von deCODE entwickelte Modell [Gulcher2000].

Eine Erweiterung des Ansatzes der Extraktion von demographischen Daten aus HL7-Nachrichten, die zwischen Routinesystemen der Krankenversorgung ausgetauscht werden, beschreiben Boyd et al.. Die Autoren schlagen die Einführung eines „Honest Brokers“ vor, mit dem Nachrichten zwischen IT-Systemen der Krankenversorgung und der Forschung geroutet werden [Boyd2009]. Der „Honest Broker“ stellt zusätzlich Funktionen zur Verfügung für das Verknüpfen verschiedener IDs eines Patienten miteinander, für die De-Identifikation von medizinischen Informationen in den Nachrichten, für die Konvertierung von Nachrichten in verschiedene Formate sowie für die Verschlüsselung der Nachrichten. Ein im „Honest Broker“ integrierter MPI speichert die verschiedenen Identifikatoren des Patienten sowie die Verknüpfung zwischen den IDs. Die Entscheidung welche IDs eines Patienten miteinander verknüpft werden erfolgt durch den Anwender im „Honest Broker“.

Der Schwerpunkt des hier vorgestellten Konzepts liegt auf dem Management von Identitäten eines Patienten durch Anwender. Im Gegensatz dazu fokussieren viele Ansätze auf das Management von Identitäten des Anwenders selbst. Diese Ansätze dienen z.B. dazu, dass sich ein Anwender nur noch an einer Anwendung authentifizieren muss („Single Sign On“ (SSO)). Diesem zentralen Dienst (Identity Provider) vertrauen die anderen Service Provider (Anwendungen) und verzichten auf eine Authentifizierung des Anwenders. Föderierte Identitätsmanagementkonzepte, die auf Anwender fokussieren erweitern SSO dahin gehend, dass eine einmalige

Authentifikation nicht nur zwischen Service Providern innerhalb einer Gruppe sondern auch bei Service Providern die einer anderen Gruppe zugeordnet sind ermöglicht wird [Josang2005], [Josang2007]. Dagegen stehen im hier vorgestellten Konzept Identitäten von Patienten, den Objekten, im Vordergrund. Deren Identitäten werden von diesen nicht selbst aktiv verwaltet. Subjekte (Ärzte, Forscher, Dokumentationskräfte, etc.) legen Identitäten von Objekten in verschiedenen IT-Systemen an und führen Manipulationen darauf durch. Den Identitäten der Objekte werden von den Subjekten Identitäten von Fällen und Biomaterialien zugeordnet, die wiederum mit besonders schützenswerten Informationen assoziiert sind. Fällen sind medizinische Daten, Biomaterialien sind Analyseergebnisse zugeordnet. Im hier vorgestellten Konzept wird insbesondere berücksichtigt, dass die Objekte selbst nicht direkt festlegen können welche sie betreffenden Identitäten angelegt, mit welchen besonders sensiblen Informationen diese assoziiert und welche Subjekte darauf zugreifen können.

4 Umsetzung

4.1 Identitätsmanagement in verteilten Umgebungen

4.1.1 Komponentenarchitektur

Um auf Identitäten von Patienten, Fällen und Einsendungen unter Wahrung von Berechtigungen in den IT-Systemen aus der Krankenversorgung (KIS/KAS, Pathologieinformationssystem, etc.) zugreifen zu können, werden diese um Suchfunktionen erweitert. Diese Funktionen werden für jedes Informationssystem, welches im Rahmen des Identitätsmanagements integriert wird, bereitgestellt (s. Connector Komponenten in Abb. 7). Zusätzlich werden Zugangsdaten von Anwendern, welche die Suche initiieren, verwaltet und systemspezifisch an die Suchfunktionen mitübergeben (Abb. 7: Zugangsdatenverwaltung). Nach außen hin stellen die Suchfunktionen einheitliche Schnittstellen zur Verfügung. Ausgehend von gefundenen und durch den Anwender identifizierten Patienten erfolgt die Extraktion der assoziierten demografischen Daten und Fallinformationen. Die vom KIS vergebenen IDs für den Patienten und dessen Fälle sind Teil der extrahierten Information. Mit diesen IDs wird (mit Hilfe der in Abb.7 skizzierten Protokollierungskomponente) überprüft, ob bereits Pseudonyme für den Patienten bekannt sind. In diesem Fall werden die Pseudonyme weiterverwendet. Anderenfalls erfolgt deren Neugenerierung und Zuordnung in der Identitätsverwaltung. Anschließend wird die Übernahme der extrahierten Informationen und Pseudonyme in die EDC-Systeme durchgeführt. Für die Protokollierung von Zusammenlegungen verschiedener Patientenidentitäten eines Patienten werden entsprechende vom KIS versendete Nachrichten protokolliert (Abb.7 Komponente für Identitätszusammenlegung).

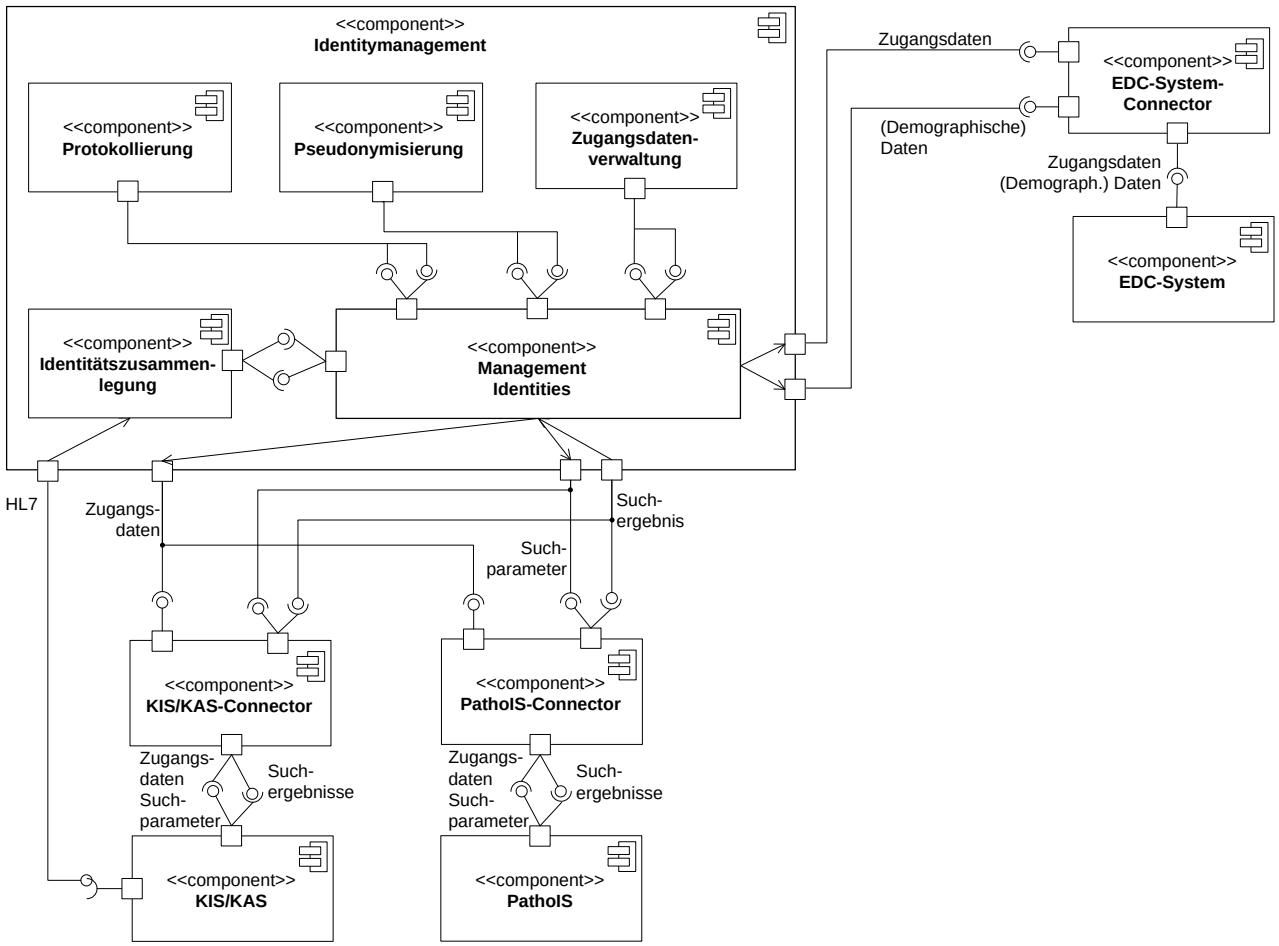


Abb. 7: UML Komponentendiagramm zur Beschreibung des Identitätsmanagements in verteilten Umgebungen

Für den lesenden Zugriff auf IS-H/i.s.h.med werden Funktionen, die die Suche nach verschiedenen Identitäten von Patienten und den damit assoziierten Fällen ermöglichen, bereitgestellt. An diese Funktionen werden Suchparameter (Zeichenketten inklusive Wildcards) für Vorname, Geburtsname, Nachname, Datumsangaben bzw. -bereiche für das Geburtsdatum übergeben. Zusätzlich ist es möglich, nach einer Patientenidentität durch Übergabe der IS-H-ID zu suchen. Ebenfalls kann eine Patientenidentität durch die Übergabe einer Fall-ID an die Suchfunktion identifiziert werden. Das Suchergebnis besteht aus einer Liste mit Patientenidentitäten. Für jeden Patienten werden dessen demografische Daten und IS-H-ID zurückgeliefert (Vorname, Nachname, Geburtsname, Geburtsdatum, Geschlecht, PLZ, Wohnort, Straße, Telefonnummer). Für die Suche nach Fällen wird die IS-H ID des Patienten an die Suchfunktion übergeben. Das Ergebnis umfasst eine Liste mit Fallidentitäten. Für jede Fallidentität wird die Fall-ID, der Typ (ambulant/stationär) des Falles, der Status, Datumsangaben über Anlage / Beendigung des Falls, sowie ggf. über Aufnahme und Entlassung des Patienten, zurückgegeben. Der lesende Zugriff auf IS-H/i.s.h.med erfolgt unter Wahrung von

Berechtigungen. Dazu werden an IS-H/i.s.h.med die Zugangsdaten des Anwenders, der die Suche initiiert, übergeben. Über diese Zugangsdaten findet eine Authentifizierung gegenüber IS-H/i.s.h.med statt. Die Zugangsdaten dienen auch zur Autorisierung beim Zugriff auf die Identitätsinformationen. Es werden nur die Informationen von der Suchfunktion zurückgeliefert, die vom Anwender durch direkte Verwendung von IS-H/i.s.h.med eingesehen werden könnten.

Die Einwilligungserklärungen für Studien werden im Spitzencluster derzeit direkt im Forschungssystem verwaltet. Da eine Dokumentation solcher Erklärungen demnächst auch in IS-H/i.s.h.med realisiert wird, wurde eine Schnittstelle für den Abgleich konzipiert. Hierfür werden beim Anlegen des Einwilligungserklärungsdokuments in i.s.h.med für einen Patienten das Datum und der Zeitpunkt protokolliert. Weitere aufgezeichnete Informationen umfassen den Typ der erfassten Einwilligungserklärung, sowie ggf. Metainformationen zur Einwilligungserklärung (Erlaubnis für Übernahme von Daten aus Krankenakte, Erhebung zusätzlicher Daten durch Analyse des Biomaterials, etc.). Das Rückgabeergebnis einer Anfrage umfasst eine Liste mit IS-H-IDs von Patienten, für die innerhalb eines übergebenen Intervalls eine Einwilligungserklärung in i.s.h.med erfasst wurde. Durch Übergabe einer IS-H-ID wird abgefragt, ob für den entsprechenden Patienten Einwilligungserklärungen vorliegen. Das Anfrageergebnis enthält die Metainformationen zu den vorliegenden Einwilligungserklärungen.

Eine weitere Funktion protokolliert die Zusammenlegung verschiedener Identitäten eines Patienten im Rahmen der Krankenversorgung. Dabei werden Nachrichten, die zwischen IS-H/i.s.h.med und den anderen im Klinikum zur Unterstützung der Krankenversorgung im Einsatz befindlichen Komponentensystemen ausgetauscht werden, belauscht und ausgewertet. Wird eine Zusammenlegung über diese Nachrichten propagiert, erfolgt die Aufzeichnung. Dabei wird erfasst, welche der verschiedenen Identitäten eines Patienten als führend angegeben wird. Es werden Anfragen auf die aufgezeichneten Informationen unterstützt. Bei Übergaben der IS-H-ID wird zurückgeliefert, ob es sich um die ID der führenden Identität handelt. Ist dies nicht der Fall, wird mitgeteilt, ob eine Zusammenlegung stattgefunden hat. Im positiven Fall werden die IS-H-ID der führenden Identität und die ihr zugeordneten Stammdaten zurückgeliefert.

Funktionen, die Zugriffe auf IS-H/i.s.h.med ermöglichen, müssen zusätzlich in der Lage sein, die von ihnen als Rückgabe gelieferten Informationen über speziell konfigurierte Netzwerke zur Verfügung stellen zu können. So gestattet das Netzwerk

im Klinikum nur externe Zugriffe von innen nach außen über einen definierten Port und über ein vorgegebenes Protokoll für den Nachrichtenaustausch.

Für den lesenden Zugriff auf Patienten-, Fall- und Probenidentitäten aus dem Pathologieinformationssystem werden ebenfalls Funktionen bereitgestellt. Es werden nur die Daten ausgelesen, auf die ein Anwender auch durch Nutzung der grafischen Benutzerschnittstelle des Systems Zugriff hat. Die Funktion stellt Suchmöglichkeiten bereit. Dazu zählen u.a. die Suche nach Patienten anhand von demografischen Daten, die Suche nach Fällen und den damit assoziierten Patienten anhand der Angabe der Einsendungsnummern (H-Nummern). Fällen wiederum sind Befundinformationen (z.B. Klassifikation eines befundeten Tumors (TNM), Klassifikation der Histologie des Tumors) zugeordnet. Da das Pathologieinformationssystem primär die Befundungen von Einsendungen sowie die Abrechnung erbrachter Leistungen unterstützt, ist das dem Pathologieinformationssystem zugrunde liegende Datenmodell „fallzentriert“. Das heißt, für eine Einsendung, die in der Pathologie bearbeitet wird, wird ein „Pathologiefall“ im Pathologieinformationssystem angelegt. Dieser kann sich von den in Klinischen Informationssystemen unterstützten Fällen (z.B. stationärer oder ambulanter) unterscheiden. Die führende Entität im Datenmodell des Pathologieinformationssystems ist der „Pathologiefall“ und nicht wie in anderen, im klinischen Umfeld im Einsatz befindlichen, Informationssystemen, der Patient. Um Suchanfragen auf Patienten- und dessen Fallidentitäten sowie die dazugehörigen Diagnosen beantworten zu können, werden entsprechende Modelltransformationen durchgeführt.

Eine weitere Funktion für das Identitätsmanagement in einer verteilten Umgebung dient dem Management von Zugangsdaten für Komponentensysteme. Dem Anwender wird mit Hilfe dieser Funktion ermöglicht seine Benutzernamen und Passwörter für die einzelnen Systeme zu hinterlegen. Diese Zugangsdaten werden dann z.B. bei einer Suchanfrage nach Patienten gegen IS-H/i.s.h.med benötigt, um sicherzustellen, dass Zugriffe des Anwenders nur auf diejenigen Daten erfolgen für die entsprechende Berechtigungen vorliegen.

Zentrale Funktionen des Identitätsmanagements dienen der Verwaltung von demografischen Daten der Patienten sowie dem Management der Abbildungen zwischen den verschiedenen Identitäten eines Patienten aus den Routine- und Forschungssystemen. Dazu werden die Abbildungen zwischen den Pseudonymen der verschiedenen Identifikatoren für einen Patienten persistiert und konsistent gehalten. Ebenso muss die Zusammenlegung von Patientenidentitäten in den

klinischen Systemen in der Identitätsverwaltung erfasst werden. Weitere zu persistierende und konsistent zu haltende Pseudonyme stammen von IDs der Fallidentitäten sowie der Identitäten von Biomaterialien. Zusätzlich gespeicherte Informationen beschreiben die Systeme, aus denen die verschiedenen IDs stammen. Daneben werden Abfragemöglichkeiten auf die persistierten Informationen bereitgestellt.

Für die Umsetzung des zweitstufigen Pseudonymisierungskonzeptes muss auf Funktionen zugegriffen werden können, welche die Pseudonymisierung vornehmen. Diese erzeugen für IDs der verschiedenen Identitäten Pseudonyme. Nach Übergabe einer ID an die Pseudonymisierungsfunktion erfolgt eine Überprüfung, ob für diese bereits ein Pseudonym existiert. Wenn dies nicht der Fall ist, erfolgt die Generierung eines neuen Pseudonyms. Die Rückgabe umfasst dann das korrespondierende Pseudonym zu der übergebenen ID.

Weitere bereitgestellte Funktionalitäten für die Umsetzung der Anforderungen für das Identitätsmanagement betreffen die Verwaltung von Benutzern. Über diese Funktionen erfolgt die Authentifizierung eines Anwenders am IT-System zum Management von Identitäten. Die Funktionen überprüfen zusätzlich, zu welchen Operationen ein Anwender im IT-System zum Management von Identitäten autorisiert ist.

4.1.2 Implementierung

Für die Umsetzung der Suche nach Patienten- und Fallidentitäten in IS-H/i.s.h.med wurde ein Wrapper entwickelt. Dieser kapselt den eigentlichen Aufruf gegen IS-H/i.s.h.med. Die vom Wrapper bereitgestellten Suchmethoden erwarten zusätzlich zu den oben beschriebenen Suchparametern auch die Zugangsdaten des Anwenders für IS-H/i.s.h.med. Auf Basis der Zugangsdaten findet eine Überprüfung der Zugriffsberechtigungen statt. Dazu setzt IS-H/i.s.h.med auf dem Berechtigungskonzept des SAP Systems auf. Im SAP-Berechtigungskonzept werden Berechtigungsobjekte definiert. Ein Berechtigungsobjekt bildet die Vorlage für eine Berechtigung. Jedes Berechtigungsobjekt kann maximal 10 Berechtigungsfelder besitzen. Von einem Berechtigungsobjekt werden Berechtigungen instanziiert. Jedem Feld, welches im Berechtigungsobjekt auf Typebene definiert wurde, wird auf Instanzebene (Berechtigung) ein individueller Wert zugeordnet. Die einzelnen Berechtigungen werden in einem Profil zusammengefasst. Dieses wird dem Benutzerstammdatensatz des Endanwenders

zugeordnet. An welchen Positionen im Programmablauf welche Berechtigungen überprüft werden, legen die Entwickler der einzelnen Programme fest. Für die Prüfung der Berechtigungen werden im auszuführenden Programm das zu prüfende Berechtigungsobjekt sowie konkrete Werte für die einzelnen Berechtigungsfelder angegeben. Das System ermittelt, über welche konkreten Berechtigungen derjenige, der das Programm ausführen möchte, verfügt und vergleicht die Werte, die in einzelnen Feldern in den Berechtigungen hinterlegt sind, mit den im Programm angegebenen Prüfwerten [Keller2006].

Die Implementierung des Wrappers für die Suche nach Patientenidentitäten erfolgt mit Java. Für den Zugriff auf IS-H/i.s.h.med wird auf Funktionalitäten der SAP Java-Connector (JCo)-Bibliothek zurückgegriffen [JCo]. Bei IS-H/i.s.h.med handelt es sich um Erweiterungen des SAP-ERP Systems. Die im IS-H/i.s.h.med umgesetzten Schnittstellen für den externen Zugriff basieren auf der, vom SAP System bereitgestellten Möglichkeit, Funktionsbausteine „remote“-fähig in Form von BAPIs bzw. RFCs zur Verfügung zu stellen. Die Implementierung der Suche nach Patienten erfolgt durch den BAPI-Aufruf der Such-Methode des „Patienten“ Business-Object in IS-H. Die vom Wrapper nach außen angebotenen Methoden werden u.a. als SOAP basierte Webservice-Schnittstellen zur Verfügung gestellt [Webservices]. Als Webservice Framework wird JAX-WS eingesetzt [Java-WS]. Zusätzlich zu den Webservice-Schnittstellen bietet der Dienst die Möglichkeit, über eine URL auf die Methoden zugreifen zu können. Dabei werden die entsprechende Methode und die für den Methodenaufruf benötigten Parameter in die URL codiert und verschlüsselt. Im Dienst erfolgen die Entschlüsselung des verschlüsselten URL-Teils, die Extraktion des Methodennamens und der Parameter, sowie die Durchführung des eigentlichen Methodenaufrufs. Für die Rückgabe der Methodenaufrufe stehen zwei unterschiedliche Ansätze zur Verfügung. Im ersten Ansatz wird das Ergebnis eines Methodenaufrufs als Post-Request an einen (Applikations-)Server gesandt. Auf diesem wird jene Anwendung ausgeführt, aus deren Weboberfläche heraus der Aufruf des Wrappers erfolgt ist. Im zweiten Ansatz stellt der Wrapper eine grafische Benutzerschnittstelle für die Visualisierung der Ergebnisse der Methodenaufrufe zur Verfügung. Dies geschieht bspw. beim Aufruf der Suchmethode als Webformular in dem die Daten zu den einzelnen gefundenen Patienten in tabellarischer Form dargestellt werden. Diese zwei Ansätze ermöglichen den Einsatz des Wrappers in Umgebungen, in denen der Zugriff auf Dienste über Netzwerkgrenzen hinweg eingeschränkt ist. So kann das Netzwerk innerhalb eines Klinikums (Intranet) vom Internet teilweise getrennt sein. Aufrufe sind lediglich aus dem Klinikumsintranet

heraus ins Internet über das Http-Protokoll und Port 80 möglich. Zugriffe aus dem Internet ins Intranet sind komplett gesperrt.

Für die Umsetzung der Erfassung von Informationen über das Vorliegen von Einwilligungserklärungen in IS-H/i.s.h.med werden parametrisierbare medizinische Dokumente (PMDs) eingesetzt. Der Anwender legt ein entsprechendes Formular an und ordnet es einem Patienten zu. Sobald das Formular zugeordnet und ausgefüllt ist, gibt der Anwender das Dokument frei. Dadurch findet eine Statusänderung des Dokuments statt. Diese Statusänderung initiiert einen „outbound“-Aufruf aus IS-H/i.s.h.med heraus. Dazu wird in IS-H/i.s.h.med intern ein RFC aufgerufen. Für diesen RFC wurde zuvor im SAP-System ein externer RFC-Server registriert. Dieser implementiert einen Handler, der den RFC-Aufruf in Empfang nimmt und auswertet. Dabei werden die mit dem RFC-Aufruf von IS-H/i.s.h.med übergebenen Parameter extrahiert. Über diese wird mitgeteilt, für welchen Patienten (durch Übergabe der IS-H ID) wann ein Dokument zur Erfassung der Einwilligungserklärung angelegt wurde. Diese Informationen werden vom RFC-Server zwischengespeichert. Im Anschluss daran ruft der RFC-Server die Identitätsverwaltungsanwendung auf und teilt dieser mit, dass für einen Patienten die Einwilligungserklärung vorliegt. Für die Umsetzung des RFC-Servers wird auf Methoden der JCo- Bibliothek zurückgegriffen. Der Aufruf der Identitätsverwaltungsanwendung erfolgt sowohl über Webservices als auch über den oben beschriebenen Methodenaufruf über URL-Parameter.

Für die Extraktion von Daten aus dem Pathologieinformationssystem wird direkt über JDBC auf die Datenbank des Systems zugegriffen. Zur Überprüfung der Berechtigungen werden die Zugangsdaten des Anwenders an den Wrapper mitübergeben. Das Pathologieinformationssystem implementiert ein einfaches Rollen und Berechtigungskonzept. Jeder Anwender, der sich gegenüber dem Pathologieinformationssystem authentifizieren kann, darf lesend auf alle Daten zugreifen. Insgesamt implementiert das System vier Rollen, die jeweils unterschiedliche Berechtigungen bezüglich dem Anlegen, Ändern und Löschen von Informationen haben. Der Wrapper, der lesenden Zugriff auf die Daten aus dem Pathologieinformationssystem bereitstellt, überprüft daher vor der Durchführung einer Anfrage, ob die Zugangsdaten, welche ihm mit der Suchanfrage übergeben wurden, mit den in der Datenbank des Pathologieinformationssystem abgelegten Informationen übereinstimmen. Suchparameter werden im Wrapper in SQL-Statements eingebettet bevor Suchanfragen durchgeführt werden. Codes in den extrahierten Daten werden im Wrapper vor der Rückgabe an den Aufrufer um sprechende Bezeichner ergänzt.

Für die Protokollierung der Zusammenlegung verschiedener Identitäten eines Patienten in Systemen zur Unterstützung der Behandlung werden die zwischen den Systemen ausgetauschten HL7-Nachrichten ausgewertet. Ein zentraler Kommunikationsserver (z.B. Cloverleaf) unterstützt die Kommunikation zwischen diesen Systemen. Die Kommunikation erfolgt asynchron, basierend auf Textnachrichten. IS-H sendet hierzu beim Eintreten von, im HL7-Standard definierten Ereignissen, Daten in Form von HCM-Nachrichten an den Kommunikationsserver. HCM-Nachrichten werden falls notwendig in HL7 V2.x Nachrichten übersetzt und an die angeschlossenen Komponentensysteme gesandt. Dazu wird der Inhalt einer HL7 Nachricht z.B. in eine Textdatei geschrieben und über FTP an einen FTP-Client des entsprechenden Komponentensystems gesandt.

Diese Textdateien werden aus dem entsprechenden Verzeichnis ausgelesen, geparst und anschließend aus dem Verzeichnis gelöscht. Für die Detektion einer Zusammenlegung zweier Patientenidentitäten eines Patienten werden Nachrichten des A40 Typs ausgewertet. Dazu werden aus dem PID-Segment der Nachricht der Identifikator der Identität, die als führend festgelegt wird, extrahiert. Im MRG Segment wird der Identifikator der Identität extrahiert, die als inkorrekt markiert wurde. Diese Informationen werden in einer Abbildungstabelle abgespeichert.

Zusätzlich lassen sich aus den HL7 Nachrichten weitere Informationen zu Patienten- und Fallidentitäten extrahieren. Damit steht zur Extraktion von Informationen zu Patienten- und Fallidentitäten eine Alternative für den direkten Zugriff auf IS-H zur Verfügung. Hierfür wird der Identifikator (IS-H ID) aus dem MSH-Segment von Nachrichten des ADT-A01 Typs extrahiert. Aus dem PID Segment werden die demografischen Daten (Vorname, Nachname, Geburtstag, Geschlecht, etc.) des Patienten extrahiert. Diese Informationen werden in einer Datenbank gespeichert. ADT-A11-Nachrichten werden ausgewertet um zu detektieren, ob eine Aufnahme abgebrochen wurde. In diesem Fall werden die Daten, die beim Parsen der zur A11 korrespondierenden A01-Nachricht erfasst wurden, wieder gelöscht.

Die Schnittstellen der Anwendung zur Protokollierung der Zusammenlegung von Patientenidentitäten werden als Webservices zur Verfügung gestellt. Die Implementierung erfolgt mit Java. Das Parsen der HL7-Nachrichten wird mit dem „HL7 application programming interface“ (HAPI) durchgeführt [HAPI]. Dabei handelt es sich um einen frei verfügbaren objektorientierten HL7 2.x Parser für Java.

Das der Anwendung zum Management von Identitäten zugrundeliegende Schema lässt sich wie folgt umreißen: Jede (digitale) Patientenentität verfügt über Attribute für die demografischen Daten (Vorname, Nachname, Geburtsname, Geschlecht,

Initialen, Titel). Zusätzlich verfügt diese Entität über Attribute, mit denen eine Zusammenlegung von Identitäten in Komponentensystemen erfasst wird. Jeder Patientenentität ist mindestens eine Organisationsentität zugeordnet. Einer Organisationsentität können mehrere Patientenentitäten zugeordnet sein. Eine Organisationsentität speichert Informationen, für welche Organisationseinheit eine Patientenentität angelegt wurde. Zusätzlich wird der Zeitpunkt erfasst, an dem die Patientenentität angelegt wurde und es wird vermerkt, von welcher Organisationseinheit zuletzt darauf zugegriffen wurde. Dadurch ist es möglich zu erfassen, welcher Patient von welcher Einrichtung erfasst wurde und ob ein Zentrumswechsel stattgefunden hat. Einer Patientenentität können eine oder mehrere (digitale) Entitäten zugeordnet werden welche die Identität eines Patienten im jeweiligen System charakterisieren. Einer solchen Identitätsentität ist genau eine Patientenentität zugeordnet. Die Identitätsentität verfügt über ein Attribut, in dem das Pseudonym eines Identifikators hinterlegt wird. Der Identifikator selbst stammt aus einem Komponentensystem und identifiziert dort den Patienten eindeutig. Weitere, den Identitätsentitäten zugeordnete Attribute, charakterisieren dieses System eindeutig. Zusätzlich kann eine Identitätsentität als „führend“ markiert werden. Einer Identitätsentität kann eine (digitale) Fallentität zugeordnet werden. Eine Fallentität ist genau einer Identitätsentität aus dem entsprechenden System zugeordnet. Fallentitäten können wiederum Referenzen auf andere Fallentitäten haben. Charakterisiert wird eine Fallentität durch eine im Quellsystem vergebene Fallnummer. In der Identitätsverwaltung selbst wird das Pseudonym der Fallnummer abgespeichert (FallPseudo). Weitere eine Fallentität beschreibende Attribute sind der Typ (ambulant/stationär/Pathologiefall (Einsendung)), das Erstellungsdatum und der Status. Jeder Fallentität können Probenentitäten zugeordnet werden. Eine Probenentität ist genau einer Fallentität zugeordnet. Die Probenentität verfügt über ein Attribut für das Proben Pseudonym (ProbenPseudo).

Die Anwendung bietet Schnittstellen für das Anlegen, Updaten und Löschen der Entitäten. Sie ermöglicht die Suche nach Patientenidentitäten anhand von Identifikatoren und demografischen Daten, auf den von ihr gespeicherten Daten. Damit lassen sich folgende Zusammenhänge ermitteln. Es kann festgestellt werden, ob für einen Patienten in einem bestimmten System Informationen vorhanden sind und in welchen weiteren Systemen für einen Patienten noch Informationen zu finden sind. Es können alle Patienten, die in einem bestimmten System vorhanden sind, identifiziert werden. Ob Identitäten eines Patienten zusammengelegt wurden lässt sich ebenfalls ermitteln. Zusätzlich zur Suche nach Identitäten von Patienten im

Datenbestand der Anwendung ermöglicht diese auch die Suche nach Identitäten eines Patienten in Komponentensystemen. Dazu bietet die Anwendung eine Benutzerschnittstelle an. Der Endanwender kann festlegen, in welchen Komponentensystemen anhand von Stammdaten und IDs gesucht werden soll. Ebenso können neue Identitäten in der Anwendung angelegt und deren Abbildungen auf Identitäten in den Komponentensystemen vom Endanwender händisch vorgenommen werden. Bevor die IDs zu den einzelnen Identitäten von der Anwendung persistiert werden, erfolgt eine Pseudonymisierung in dem entsprechenden Dienst und die Speicherung dieses Pseudonyms.

Die Identitätsmanagementanwendung ist in Java als SOA [Krafzig2005] implementiert. Die Benutzerschnittstelle wird mit Webtechnologien umgesetzt. Hierbei werden Java Server Faces [JSF] in Form der MyFaces- und Jenia-Implementierung eingesetzt [MyFaces], [Jenia]. Die Seitenstrukturen der Benutzeroberfläche werden mit Hilfe von Templates abstrahiert. Hierzu wird Tiles 2 eingesetzt [Tiles2]. Mit Hilfe von Hibernate [Hibernate] werden die von der Identitätsmanagementanwendung zu persistierenden Daten in einer relationalen Datenbank gespeichert. Für die Kommunikation mit anderen Diensten (Wrapper) wird Webservice-Technologie verwendet. Die über die Benutzerschnittstelle angebotenen Funktionalitäten werden auch als API vom Dienst bereitgestellt. Zusätzlich bietet die Anwendung eine Schnittstelle über die Post-Requests von Diensten und Wrappern entgegengenommen und verarbeitet werden können (IS-H/i.s.h.med Wrapper).

Die beschriebenen Implementierungen befinden sich in verschiedenen Projekten im Produktiveinsatz bzw. liegen als prototypische Umsetzungen vor. Der Wrapper für die Suche nach Patienten- und Fallidentitäten ist Teil des für m4 entwickelten Frameworks. Ebenso wird er bei der Weichteilsarkom-Forschungsdatensammlung eingesetzt [Lamla2010]. Dabei werden Daten, die während Tumorboardsitzungen zu Weichteilsarkomen erfasst werden, für Forschungszwecke weiter verwendet. Über den Wrapper werden Patientenidentitäten gesucht und deren Stamm- und Falldaten übernommen, für die Tumorboarddokumente in i.s.h.med vorliegen. Ebenso wird der Wrapper in der Biomaterialverwaltungsanwendung der Pathologie eingesetzt. Dabei wird die Suche nach und die Übernahme von Stammdaten in die Anwendung unterstützt. Der Wrapper wird auch dazu verwendet, um Patienten in bestehenden Forschungsdatensammlungen deren IS-H-ID zuzuordnen. Dadurch eröffnet sich die

Möglichkeit, Daten, die zu einem einzelnen Patienten in verschiedenen Forschungsdatensammlungen erfasst wurden, nachträglich zusammenzuführen.

Die Anwendung, mit der die Zusammenlegung verschiedener Identitäten eines einzelnen Patienten protokolliert wird, ist derzeit als prototypische Implementierung vorliegend. Sie wurde ausführlich mit Echtdaten getestet. Die Anwendung wird im Rahmen des Rollouts der Softwarekomponenten des Spitzenclusterprojekts m4 in den Einsatz gebracht.

Die Identitätsmanagementanwendung ist ein wichtiger Bestandteil des m4-Frameworks. Mit ihr werden die Patientenidentitäten und zugeordneten, öffentlich bekannten IDs (Patienten- und Fall-ID) verwaltet. Die Identitätsmanagementanwendung entspricht der im Konzept genannten IDAT-Anwendung. Daneben ist die Identitätsmanagementanwendung Teil eines prototypisch implementierten Portals mit dem Routine- und Forschungssysteme auf Präsentationsebene integriert werden können. Dabei werden Identitäten aus IS-H/i.s.h.med, der Biomaterialverwaltungsanwendung der Pathologie, der Forschungsdatensammlung für kolonrektale Tumoren, und der Spezialdokumentation für Schilddrüsenerkrankungen der Chirurgie verwaltet.

4.2 Übernahme von Daten aus Routine- in Forschungssysteme

4.2.1 Komponentenarchitektur

Für die Übernahme von Daten aus der Krankenversorgung in Forschungssysteme werden Routinesysteme um Schnittstellen für die Extraktion von klinischen Daten erweitert. Beim Zugriff über diese Schnittstellen findet eine Überprüfung der Berechtigungen statt. Zusätzliche Schnittstellen ermöglichen die Abfragen von Statusinformationen zu Dokumenten sowie das Auslesen von Formularmetadaten. Diese Metadaten (z.B. Feldname, Datentyp, Auswahllisten) bilden die Grundlage für die Identifikation von Änderungen an Formularen.

Forschungssysteme werden um Schnittstellen für den schreibenden Zugriff unter Wahrung von Berechtigungen erweitert (s. Abb. 8: UML Komponentendiagramm: EDC-System-Connectoren). Zugriffe über die Schnittstelle werden im „Audit Trail“ des Forschungssystems aufgezeichnet. Auch aus Forschungssystemen können Formularmetadaten ausgelesen werden. Für die Zugriffe unter Wahrung von

Berechtigungen auf Routine- und Forschungssysteme werden Zugangsdaten der Anwender verwaltet (Abb. 8: Zugangsdatenmanagementkomponente).

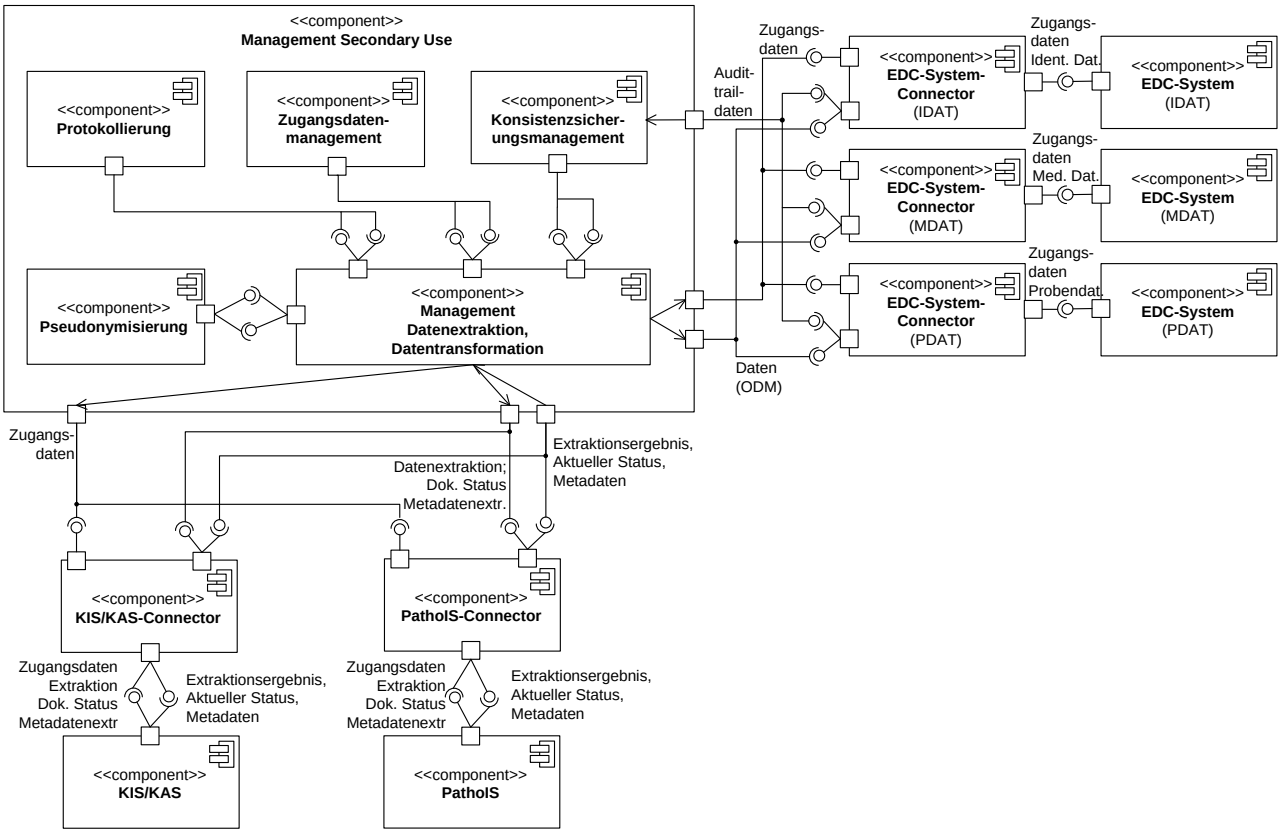


Abb. 8: UML Komponentendiagramm zur Beschreibung der Übernahme von Daten aus Routine- in Forschungssysteme

Der Datenübernahmeprozess wird initiiert sobald einem Dokument im Routinesystem ein vorher festgelegter Status zugeordnet wird bzw. der Anwender den Übernahmeprozess startet. Im Übernahmeprozess erfolgt eine Transformation der Dokumente in die Zielformate der Forschungssysteme. Neben der Möglichkeit den Inhalt gesamter Dokumente übernehmen zu können ist es ebenso möglich selektiv einzelne Daten zu extrahieren. Damit wird die getrennte Speicherung von Daten aus den Routinesystemen in unterschiedlichen Forschungssystemen ermöglicht, z.B. in medizinische Daten (MDAT) einerseits und in Biomaterialannotationsdaten (PDAT) anderseits. Des Weiteren erfolgt eine Dokumentation des Übernahmeprozesses (Abb.8: Protokollierungskomponente).

Zur Konsistenzsicherung auf Typebene werden extrahierte Formularmetadaten mit Vorgängerversionen abgeglichen (Abb.8: Konsistenzsicherungskomponente). Auf Instanzebene werden die „Audit Trails“ der Forschungssysteme auf Einträge hin überprüft, die spezifizierte Änderungen an Daten anzeigen. Identifizierte Änderungen führen zur Benachrichtigung entsprechender Anwender.

Basis für die Umsetzung bilden Funktionen, die lesende Zugriffe auf klinische Informationssysteme und schreibende Zugriffe auf Forschungssysteme ermöglichen.

Für den lesenden Zugriff auf i.s.h.med wird eine Funktion bereitgestellt, mit der alle Dokumente eines bestimmten Typs, die in Form von sog. „Parametrisierbaren Medizinischen Dokumenten“ (PMDs) vorliegen, für einen Patienten extrahiert werden können. Für die Extraktion der in PMDs erfassten Daten zu einem Patienten werden an die Funktion Zugangsdaten des Anwenders, der die Daten extrahieren möchte, der Dokumententyp und die -version, die IS-H ID des Patienten und eine Zeichenkette, mit dem Begründungstext für die Extraktion, übergeben. Die Begründung, der Zeitpunkt an dem der Zugriff erfolgt ist und wer diesen veranlasst hat, wird protokolliert. Das Protokoll wird vom Datenschutzbeauftragten ausgewertet. Durch die Übergabe der Zugangsdaten des Anwenders an die Funktion wird sichergestellt, dass nur die Daten aus i.s.h.med extrahiert werden, auf die der Anwender auch Zugriff hatte, wenn er direkt über die grafische Benutzerschnittstelle des Systems interagieren würde. Änderungen an den Berechtigungen im IT-System (falls dem Anwender z.B. der Zugriff auf einen Dokumententyp entzogen wird) werden damit unmittelbar bei der Datenextraktion für Forschungszwecke mit berücksichtigt.

Für die Abfrage des Dokumentenstatus von PMDs in i.s.h.med wird eine weitere Funktion entwickelt. Damit können alle Patientenidentitäten identifiziert werden, denen ein Dokument zugeordnet ist, welches über einen definierten Status verfügt. Um das Suchergebnis einschränken zu können, wird eine Datumsangabe an die Funktion übergeben. Damit können alle Patientenidentitäten identifiziert werden an deren Dokumente zwischen übergebenem Zeitpunkt und dem Anfragezeitpunkt Statusänderungen sattgefunden haben.

Auch diese Funktionen müssen ihre Ergebnisse aus dem Klinikumsintranet an eine Anwendung im Internet übermitteln zu können.

Für das im Konzept aufgeführte MDAT-Teilsystem kann ein CDMS System eingesetzt werden. Die im klinischen Informationssystem erfassten hochstrukturierte Daten werden zum Vorbefüllen der Formulare in diesen CDMS-Systemen weiterverwendet. Im Rahmen dieser Arbeit wird das CDMS-System Macro von Infermed [Macro2009] verwendet. Für die Anbindung von Macro werden hierfür Funktionen für den schreibenden Zugriff bereitgestellt. Der Zugriff erfolgt ebenfalls unter Wahrung von Berechtigungen. Damit wird sichergestellt, dass nur berechtigte Anwender Übernahmen der Daten veranlassen können. Dazu werden die Zugangsdaten des Anwenders, der Name der Studie, der Name der Datenbank in

der die Studie abgespeichert ist und der Name des Studienzentrums übergeben. Der Rückgabewert ist die von Macro für die neu angelegte Patientenidentität vergebene ID. Zusätzlich ist es möglich, für eine bereits angelegte Patientenidentität neue Formulare (eCRFs) zu erzeugen bzw. Änderungen an den erfassten Daten vorzunehmen. Um einen FDA 21 CFR Part 11 konformen Datenübernahmeprozess zu unterstützen, stellt Macro eine Programmierschnittstelle zur Verfügung. Im Rahmen der Arbeit wurden Funktionen konzipiert, mit denen die zu übernehmenden Daten an diese Schnittstelle im CDISC-ODM-Format übergeben werden können. Die finale Übernahme in Macro erfolgt nachdem ein bzw. im Falle von „Double Data Entry“ zwei Anwender mit entsprechenden Berechtigungen, die Daten überprüft und freigegeben haben.

Für die Koordinierung und Protokollierung der Datenübernahme zwischen den klinischen Informationssystemen und den Forschungssystemen wurde im Rahmen dieser Arbeit eine zusätzliche Funktionalität entwickelt. Hierbei wird erfasst, von welcher Patienten- und Fallidentität aus dem Quellsystem Daten extrahiert und welcher Patienten- und Fallidentität diese im Zielsystem zugeordnet werden. Zusätzlich kann es notwendig werden, dass Daten für den Austausch zwischen den Systemen transformiert und dabei auch um zusätzliche Daten ergänzt werden müssen. So gilt es verschiedene IDs, die die verschiedenen Identitäten von Patienten/Fällen in Quell- und Zielsystemen eindeutig identifizieren, durch Pseudonyme zu ersetzen.

Die zwischen den klinischen Informationssystemen (z.B. i.s.h.med) und Forschungssystemen (z.B. Macro) replizierten Daten müssen konsistent gehalten werden. Hierfür werden entsprechende Funktionen bereitgestellt. Es wird zwischen der Konsistenzsicherung auf Typ- und Instanzebene unterschieden.

Konsistenzsicherung auf Typebene stellt dabei sicher, dass Änderungen z.B. an elektronischen Formularen (PMDs, eCRFs) identifiziert werden. Änderungen an Formularen führen zu Änderungen in den Schemata der Komponentensysteme bzw. zu Änderungen in den die Formulare beschreibenden Daten (Formularmetadaten). Gründe für Änderungen sind u.a. das Löschen, das Hinzufügen und Ändern von Feldern auf Formularen. Änderungen betreffen u.a. die Zuordnung neuer Datentypen und Wertebereiche zu Feldern, Änderungen an den Bezeichnern der Felder, Änderungen an hinterlegten Auswahlmöglichkeiten für die Belegung der Felder, Änderungen an der Zuordnung von Feldern zu Formularen bzw. zu Unterstrukturen in den Formularen. Weitere bei Formularen auftretende Änderungen entstehen durch die Versionierung ganzer Formulare. Für die Identifikation von Änderungen an

Formularen (Typebene) werden Funktionen für die Extraktion von Schemainformationen bzw. von Formularmetadaten bereitgestellt. Änderungen werden durch den Vergleich von aktuellen Schemainformationen mit Schemainformationen, die zu einem früheren Zeitpunkt ermittelt wurden, identifiziert. Finden sich in den zu unterschiedlichen Zeitpunkten extrahierten Informationen Unterschiede, so werden verschiedene Eskalationsstrategien zur Behandlung angewendet. Handelt es sich bei dem zugrunde liegenden System um ein Quellsystem und betreffen die Unterschiede Attribute, die nicht in Zielsysteme übernommen werden, können folgende Fälle unterschieden werden.

Wurde ein Attribut gelöscht ist nichts weiter zu tun. Bei Änderungen an einem Attribut bzw. beim Neuanlegen eines Attributes sind Hinweise zu generieren. Ziel dabei ist es, auf semantische Überlappungen hinzuweisen. Diese können u.a. auftreten, wenn das Quellsystem erweitert wird, um zusätzliche Daten zu erfassen, ohne dabei semantische Aspekte der mit den Zielsystemen erfassten Daten zu berücksichtigen. Semantische Aspekte betreffen hierbei (Halb-/Quasi)Synonyme. Selbige Vorgehensweisen werden durchgeführt, wenn die Unterschiede in Schemainformationen von Zielsystemen identifiziert werden und Attribute betroffen sind, für die auf Instanzebene keine automatisierte Datenübernahme zwischen den Systemen erfolgt. Betreffen die identifizierten Unterschiede Attribute, die in Quell- und Zielsystemen vorkommen und für die Daten ausgetauscht werden, so treten folgende Fälle ein. Werden Attribute im Quellsystem gelöscht, sind die Administratoren der Zielsysteme mit dem Hinweis darüber zu informieren, dass entsprechende Daten nicht mehr erfasst werden. Erfolgt die Löschung in einem Zielsystem, so werden die Administratoren der Zielsysteme informiert, dass bei zukünftigen Schreibzugriffen im Rahmen der automatisierten Datenübernahme Fehlermeldungen auftreten werden und ggf. die Transformationsregeln im Transferprozess anzupassen sind. Werden Änderungen an den entsprechenden Attributen im Quellsystem vorgenommen, müssen diese Änderungen in den Transformationsregeln der Datentransferprozesse berücksichtigt werden. Selbiges gilt auch für Änderungen an Attributen in den Zielsystemen.

Mit der Sicherung der Konsistenz auf Instanzebene wird erreicht, dass Änderungen an Daten, die nach dem Datentransfer erfolgen, in betroffenen Systemen nachgezogen werden können. Es lassen sich, abhängig von den Systemen in denen die Änderungen zuerst stattgefunden haben, verschiedene Szenarien identifizieren, in denen unterschiedliche Vorgehensweisen für die Sicherung der Konsistenz der Daten durchgeführt werden. Ein wichtiger Aspekt, der Einfluss auf die

Vorgehensweise hat, betrifft den Status der elektronischen Dokumente in den Quellsystemen zum Zeitpunkt der Initiierung der Datenübernahme. Dokumente, in denen während der Behandlung die Eingabe von Daten erfolgt, können verschiedene Status zugeordnet werden. Abhängig vom Status erlauben die IT-Systeme unterschiedlichen Benutzern unterschiedliche Operationen auf diesen Dokumenten. So können Dokumente z.B. nach Freigabe nicht mehr bearbeitet und keine Änderungen daran vorgenommen werden. Eine Strategie zur Sicherung der Konsistenz von Daten, die aus Systemen zur Unterstützung der Behandlung in Forschungssysteme übernommen werden, besteht darin, die Datenübernahmeprozesse erst zu starten wenn die entsprechenden Dokumente freigegeben sind. Dies würde bei der Dokumentation medizinischer Daten auf Papier dem Fall ähneln, Daten aus medizinischen Dokumenten erst in Forschungsdokumentationen zu übernehmen, wenn diese von einer autorisierten Person unterschrieben und freigegeben sind. Erfolgt die Datenübernahme aus Routinesystemen nach Freigabe der Dokumente, müssen Änderungen an diesen Daten, die in einem Forschungssystem vorgenommen werden, an die Routinesysteme zurückpropagiert werden. Dazu werden für die Forschungssysteme Funktionen bereitgestellt, die Änderungen an bereits erfassten Daten erkennen können und in denen hinterlegt werden kann, bei welchen Änderungsoperationen wer zu benachrichtigen ist. Müssen Daten in freigegebenen Dokumenten der Routinesysteme geändert werden, so geschieht dies in dem eine neue Version des entsprechenden Dokuments im Routinesystem angelegt werden muss. In dieses werden üblicherweise die neuen und die nicht geänderten Daten aus der Vorgängerversion übernommen. Nach Freigabe der neuen Version des Dokuments kann ein erneuter Datenübernahmeprozess initiiert werden. In diesem sind dann Handlungsweisen zu hinterlegen, wie mit Daten aus Dokumenten aus höheren Versionen umzugehen ist.

4.2.2 Implementierung

Für die Extraktion von Daten aus i.s.h.med, die mit PMDs erfasst werden, wurde der bestehende Wrapper für lesende Zugriffe erweitert. Die Extraktion aller Dokumente eines bestimmten Typs für einen Patienten erfolgt durch Aufruf eines RFCs. Dabei erfolgt zuerst die Überprüfung, ob derjenige Person (identifiziert anhand der übergebenen Zugangsdaten), die die Extraktion durchführen will, die Berechtigungen zugeordnet sind, um auf die Daten des Patienten (der durch die

übergebene IS-H-ID identifiziert wird), zugreifen zu dürfen. Daran schließt sich eine Überprüfung an mit der sichergestellt wird, dass der übergebene Dokumententyp auch für den Zugriff mit dem Extraktions-RFC freigegeben ist. Danach wird überprüft, ob für den übergebenen Dokumententyp auch Dokumente hinterlegt sind, die vom Typ „PMD“ sind. Durch die Überprüfung der Zugriffsberechtigung des Aufrufers auf den Dokumententyp wird sichergestellt, dass über den i.s.h.med-Dienst für den lesenden Zugriff nur auf die Daten aus den Dokumenten Zugriff gewährt wird, auf die auch zugegriffen werden könnte, wenn die Zugangsdaten für den Login in der GUI des IS-H / i.s.h.med System benutzt würden. Nach Abschluss der Überprüfungen, die bei negativem Ausgang jederzeit zum Abbruch der eigentlichen Extraktion und zur Generierung einer entsprechenden Fehlermeldung führen, wird eine Liste erzeugt. Diese enthält alle Dokumente des angefragten Dokumententyps für einen Patienten. Bei der Erstellung der Liste wird darauf geachtet, ob nur die Dokumente des aktuellen Falls berücksichtigt werden müssen. Danach erfolgt die Serialisierung der Objektrepräsentation der Dokumente. Das Ergebnis der Serialisierung ist ein XML-Dokument, welches in seinem Wurzelknoten u.a. die Patienten-ID enthält. Jeder Kindknoten des Wurzelknotens repräsentiert ein extrahiertes Dokument. Die Kindknoten der „Dokumentenknoten“ enthalten Beschreibungen der entsprechenden PMD Felder und die zugeordneten Werte, die an den entsprechenden Positionen in den PMDs enthalten sind. Das XML-Dokument wird als Zeichenkette an den Aufrufer zurückgegeben. Zuvor wird allerdings die erstellte Dokumentenliste noch an die Datenschutzkomponente übergeben. Diese iteriert über die Liste und trägt für jedes enthaltene Dokument in einer Tabelle ein, welcher Aufrufer des RFC-Bausteins, wann mit welcher Begründung zugegriffen hat. Die Implementierung des RFC-Bausteins wurde von der Firma GSD/Siemens durchgeführt.

Für die Abfrage des Dokumentenstatus in IS-H steht eine weitere Funktion zur Verfügung. Diese setzt ebenfalls einen RFC-Aufruf eines Funktionsmoduls um. In diesem ist eine Open-SQL-Anfrage hinterlegt. Damit erfolgt ein „Join“ der Tabelle, welche die Kerndaten zum angefragten Dokumententyp enthält, mit der Tabelle, in welcher der aktuelle Status eines Dokuments abgelegt ist, und der Tabelle, welche die Verknüpfungsinformationen zwischen Patient, Fall und Dokument speichert. Die an die Funktion übergebenen Selektionssbedingungen sind der Dokumentenstatus, der Dokumententyp und die Angabe des Datums, an dem die Statusänderung durchgeführt wurde.

Für die Übernahme der Daten in das Forschungssystem, bietet das CDMS-System Macro eine API an. Die API ermöglicht die Überprüfung von Berechtigungen eines Anwenders. Nach erfolgreicher Überprüfung wird von der Schnittstelle ein Token zurückgeliefert. Über dieses Token, welches beim Folgeaufruf einer Methode der API mitübergeben werden muss, erfolgt die Authentifizierung und Autorisierung. Die API stellt eine Methode bereit, über die eine neue Patientenidentität in Macro angelegt werden kann. Als Rückgabe erhält man die ID mit der die Patientenidentität in Macro identifiziert werden kann. Für die Übergabe von Daten zum Vorbefüllen der eCRFs in Macro steht eine eigene Methode zur Verfügung. An diese muss neben den Informationen zur Identifikation der entsprechenden Studie, des Studienzentrums, und der Datenbank auch ein XML-Dokument mit den zu übernehmenden Daten übergeben werden. Das XML-Dokument selbst ist studienspezifisch. Es enthält eine ID der Studie, um diese eindeutig zu identifizieren. Zusätzlich repräsentiert die Struktur des XML-Dokuments die Struktur der zu erfassenden Daten in den eCRFs, sowie deren Zuordnung zu Patientenbesuchen. Jedes Feld im XML-Dokument muss über den Macro internen Feldnamen adressiert werden und einen Wert beinhalten. Bei Auswahlfeldern muss der Macro-interne Wert für das Feld vorbelegt werden. Daneben muss das XML-Dokument auch die ID der Patientenidentität enthalten, deren eCRFs vorbefüllt werden. Die von Macro bereitgestellte API für den schreibenden Zugriff liegt als „Dynamische Linked Library“ (DLL) vor. Der Zugriff über Java auf die DLL erfolgt über Com4J [Com4J]. Dieses Framework kapselt JNI-Zugriffe auf DLLs.

Für die Protokollierung des Datentransfers zwischen klinischen Informationssystemen und Forschungssystemen wurde folgendes Schema umgesetzt. Die Entität „MasterPatient“ stellt die führende Entität dar. Ihr sind Attribute zugeordnet, in denen vermerkt wird, dass eine Zusammenlegung verschiedener Identitäten eines Patienten in den Routinesystemen erfolgt ist. Findet eine Zusammenlegung statt, so wird eine Verknüpfung zwischen den korrespondierenden „MasterPatient“-Entitäten erstellt. Jeder dieser Entitäten können beliebig viele Entitäten zugeordnet werden („Pat_Identity“). Sie beschreiben die Identität eines Patienten in einem Komponentensystem, für den ein Datentransfer zwischen den Systemen stattgefunden hat. Jede Entität welche die Identität eines Patienten charakterisiert, ist genau einer „MasterPatient“ Entität zugeordnet. Um die Komponentensysteme, aus denen die einzelnen Identitäten zu einem Patienten stammen, näher zu beschreiben, wird der „Pat_Identity“-Entität eine entsprechende Entität („System“) zugeordnet.

Jeder Entität, die eine Patientenidentität charakterisiert, werden Entitäten zugeordnet, die Informationen zu Fällen beinhalten. Jedem Fall können beliebig viele Dokumente zugeordnet werden. Für ein Dokument werden dessen Nummer, der Typ, die Version und der Status des Dokuments gespeichert. Ein Dokument kann sich aus Gruppen von Feldern zusammensetzen, denen wiederum einzelne Felder zugeordnet sind. Ein einzelnes Feld wird durch folgende Attribute charakterisiert. Einem Feld wird eine eindeutige Nummer zugeordnet. Ebenso wird der Name des Feldes, dessen Version und der Typ des Feldes erfasst. Im Attribut „Position“ wird bei sich wiederholenden Feldern festgelegt, an welcher Stelle sich ein Feld befindet. Dies umfasst die Positionen in den Quell- und Zielformularen. Zusätzlich wird beim Datentransfer für ein Feld vermerkt aus welchem Quellsystem die Daten zum Feld stammen und in welches Zielsystem die Daten übernommen wurden. Daten aus einem Feld können in beliebig viele Zielsysteme übernommen werden. Für jeden Transfer wird der genaue Zeitpunkt vermerkt. Wird im Rahmen des Datentransfers der Inhalt eines kompletten Dokuments aus dem Quellsystem in das Zielsystem übernommen, so wird dies protokolliert, in dem für ein Dokument genau eine Gruppe von Feldern, der wiederum genau ein Feld zugeordnet ist, gespeichert wird. In diesem Fall sind die Dokumentennummer, die Nummer der Gruppe der Felder und die des einzelnen Feldes gleich. Für Abfragen auf den protokollierten Informationen werden folgende Funktionen bereitgestellt.

Beschreibung	Parameter
Es wurde ein Dokument (Gruppe von Feldern, einzelnes Feld) eines bestimmten Typs, mit gegebener ID und Version transferiert	• Dokumentennummer (Feldgruppennummer, Feldnummer) • Version • Typ
Es wurde ein Dokument (Gruppe von Feldern, einzelnes Feld) eines bestimmten Typs, mit gegebener ID und Version aus einem bestimmten Quellsystem in ein bestimmtes Zielsystem übernommen und umgekehrt	• Dokumentennummer (Feldgruppennummer, Feldnummer) • Version • Typ • Quellsystem / Zielsystem
Rückgabe einer Liste aller Dokumente (Gruppe von Feldern, einzelnes Feld) die aus einem bestimmten Quellsystem / Zielsystem übernommen wurden	• Quellsystem / Zielsystem

Rückgabe aller zu einem Patienten verfügbaren Informationen der über ein Identifikator identifiziert wird	• Identifikator • Quellsystem / Zielsystem
Rückgabe aller Patienten für die Daten aus bestimmten Quellsystem / Zielsystem übernommen wurden	• Quellsystem / Zielsystem
Rückgabe einzelner Dokumente (Gruppen von Feldern, Felder) abhängig von Quellsystem, Zielsystem und Datumsbereich	• Quellsystem / Zielsystem • Datumsbereich
Anlegen eines neuen „Master_Patient" Datensatzes im Transferprotokollierungs-dienst.	• Name und Beschreibung des Quellsystems • Fallnummer Quellsystem • Dokumentennummern, -namen, -typen, -version, Status für Quellsystem • Nummer, Name und Version für Gruppe von Feldern für Quellsystem • Nummer, Name, Typ, Version und Position für Frage für Quellsystem • (Rückgabe: ID des „Master_Patient“)
Anlegen eines neuen Transferdatensatzes (Quellsystem / Zielsystem)	• (ID des „Master_Patient“) • Name und Beschreibung des Quell-/ Zielsystems • Identifikator des Patienten in Quell-/Zielsystem • Dokumentennummer (Feldgruppennummer, Feldnummer) • Version, Typ, Status, Position • Verknüpfung zwischen Quell-/Zielsystem
Aktualisieren eines Transferdatensatzes (Quellsystem / Zielsystem)	• (ID des „Master_Patient“) • Name und Beschreibung des Quell-/ Zielsystems • Identifikator des Patienten in Quell-/Zielsystem • Dokumentennummer

	(Feldgruppennummer, Feldnummer) • Version, Typ, Status, Position • Verknüpfung zwischen Quell-/Zielsystem
Löschen eines Transferdatensatzes (Quellsystem / Zielsystem)	• (ID des „Master_Patient“) • Name und Beschreibung des Quell-/Zielsystems • Identifikator des Patienten in Quell-/Zielsystem • Dokumentennummer (Feldgruppennummer, Feldnummer)
Liste aller bekannten Quellsysteme	• --
Liste aller bekannten Zielsysteme	• --

Tabelle 1: Methoden: Protokollierung Datenübernahme

Die Sicherung der Konsistenz von transferierten Daten auf Typ und Instanzebene wurde wie folgt gelöst: Um Änderungen auf Typebene an Formularen aus dem KAS und eCRFs aus dem EDC-System identifizieren zu können, wurden im Rahmen dieser Arbeit Funktionen für die Abfrage von Informationen zu den Schemas der, den Formularen und eCRFs zugrunde liegenden Datenmodelle entwickelt. Zur Beschreibung wird das in [Wurst2010] vorgestellte Modell verwendet. Dieses sieht vor, Schemainformationen in Form eines RDF-Graphen zu repräsentieren [RDF]. Das Datenmodell setzt sich aus Attribut-Wert Tupeln zusammen. Untereinander stehen die Tupeln in einer Graphstruktur miteinander in Beziehung. Die Position eines Tupels in diesem Graph entspricht dem korrespondierenden Schemaelements aus dem Schema des Komponentensystems. Dabei werden Beziehungen von Entitäten (z.B. Relationen, Assoziationen, Hierarchien) aus den Datenmodellen der Komponentensysteme über die Verknüpfung von RDF Ressourcen durch Prädikate beschrieben. Alle Ressourcen, die als Objekt unterhalb derselben Subjekt Ressource eingeordnet sind, befinden sich im Datenmodell des Komponentensystems im selben Kontext. Die Repräsentation von Beziehungen (Relationen, Assoziationen, Hierarchien) zwischen Entitäten, wird durch das Verschachteln von Ressourcen dargestellt. Ein Attribut aus dem Komponentensystem wird im RDF-Modell dadurch abgebildet, dass einer Ressource nur noch Literale zugeordnet sind. Die Namen der Literale beinhalten datentypspezifische Metadaten. Ein komplexer Datentyp eines Attributs aus dem Datenmodel des Komponentensystems wird durch die Zuordnung mehrerer, diesen

Datentyp charakterisierenden Literale, zu dem korrespondierenden Knoten im RDF-Graph repräsentiert. Literale selbst verwenden primitive Datentypen aus XML-Schema. Die Datentypen werden in folgende Klassen eingeteilt. Die Klasse „Primitiver Datentyp“ umfasst Text, Ganzzahl, Gleitkommazahl und Boolean. In der Klasse „Wert mit Einheit“ werden Ganzzahlen und Gleitkommazahlen zusätzlich mit Metainformationen, die eine Einheit darstellen versehen. In der Klasse „Datum und Uhrzeit“ werden neben dem Textwert auch Metadaten zugeordnet, die das Format spezifizieren. In der Klasse „Terminologisch kontrollierte Datentypen“ werden die einem Datentyp zugeordnete Terminologie und deren Version spezifiziert [Wurst2010]. Beispiele für die Darstellung von Attributen aus den Schemas der Komponentensysteme in RDF sehen wie folgt aus.

```
 <ishmed:lastname rdf:parseType="Resource">
    <attr:type rdf:datatype="http://www.w3.org/2001/XMLSchema#string">java.lang.String</attr:type>
 </ishmed:lastname>

<ishmed:birthday rdf:parseType="Resource">
    <attr:type rdf:datatype="http://www.w3.org/2001/XMLSchema#string">java.util.Date</attr:type>
    <attr:format rdf:datatype="http://www.w3.org/2001/XMLSchema#string">dd.MM.yyyy</attr:format>
</ishmed:birthday>

<ishmed:icd rdf:parseType="Resource">
    <attr:type rdf:datatype="http://www.w3.org/2001/XMLSchema#string">java.lang.String</attr:type>
    <attr:terminology rdf:datatype="http://www.w3.org/2001/XMLSchema#string">ICD</attr:terminology>
    <attr:version rdf:datatype="http://www.w3.org/2001/XMLSchema#string">10</attr:version>
</ishmed:icd>

<ishmed:gender rdf:parseType="Resource">
    <attr:type rdf:datatype="http://www.w3.org/2001/XMLSchema#string">java.lang.String</attr:type>
    <attr:controlled rdf:datatype="http://www.w3.org/2001/XMLSchema#boolean">true</attr:controlled>
    <attr:option rdf:datatype="http://www.w3.org/2001/XMLSchema#string">M</attr:option>
    <attr:option rdf:datatype="http://www.w3.org/2001/XMLSchema#string">W</attr:option>
    <attr:option rdf:datatype="http://www.w3.org/2001/XMLSchema#string">?</attr:option>
</ishmed:gender>
```

Abb. 9: Bsp.: Attributdarstellung in RDF-Metadatenmodel

Beispielhaft illustriert das letzte Element <ishmed:gender> wie ein Attribut („gender“), dem im Komponentensystem ein komplexer Datentyp zugeordnet ist, im generischen Metadatenmodell abgebildet wird.

Änderungen an PMDs und eCRFs auf Typebene werden identifiziert, in dem ein Metadatenmodell (Graphstruktur) generiert und mit Metadatenmodellen, die zu einem früheren Zeitpunkt erzeugt wurden, auf Isomorphie hin überprüft werden. Die Implementierung der Funktionen erfolgt mit Java. Für die Verarbeitung von RDF Daten wird das Jena-Framework eingesetzt [Jena]. Dieses stellt u.a. Funktionalitäten für die Überprüfung von RDF-Graphen auf Isomorphie bereit. Für den Zugriff auf i.s.h.med werden die oben beschriebenen Frameworks eingesetzt. Der Zugriff auf

Macro wird über eine JDBC-Verbindung auf das Macro zugrundeliegende DBMS-System realisiert.

Für die Sicherung der Konsistenz der transferierten Daten auf Instanzebene werden die „Audit Trails“ der Forschungssysteme auf Änderungen hin überwacht. Dies betrifft Änderungsoperationen (Anlegen, Ändern, Löschen) an Werten von Attributen, die aus einem klinischen Informationssystem übernommen wurden. Ein Attribut steht hier stellvertretend für ein Feld auf einem eCRF (der wiederum einer Patientenidentität zugeordnet ist), in das Werte eingetragen bzw. zugeordnet werden. Für die Identifikation von Änderungen wird dazu festgelegt, bei welchen Änderungsoperationen, an welchen Attributen wer zu benachrichtigen ist.

Einer Studie ist hierzu eine Menge an Attributen zugeordnet. Ebenso auch der Name der Datenbank, in der diese gespeichert sind. Die Identifikation eines Attributes erfolgt über die Studie. Jedem Attribut ist ein Listener zugeordnet. Dieser wird aufgerufen, wenn eine vorgegebene Operation auf dem Attribut durchgeführt wird. Der Endpunkt des Listeners wird durch eine URL beschrieben. Einem Attribut können verschiedene Operationen zugeordnet werden. Wird eine der zugeordneten Operationen (Anlegen, Ändern, Löschen) auf dem Attribut ausgeführt, so erfolgt die Notifikation. Auf Instanzebene kann ein Attribut über verschiedene Ausprägungen verfügen. Daher ist jedes Attribut einer Patientenidentität zugeordnet. Eine Patientenidentität kann mehrere Attribute haben. Zusätzlich persistiert der Dienst den von Macro vergebenen Zeitpunkt, an dem der Eintrag in den „Macro-Audit Trail“ erfolgt ist. Der „Audit Trail“ von Macro findet sich in einer Tabelle im Macro zugrunde liegenden DBMS. Jede Zeile in der Tabelle beschreibt eine Änderung an einem Attribut. Dabei wird neben dem Attributnamen der Identifikator der Studie, in der das Attribut verwendet wird, eingetragen. Weitere Einträge pro Zeile umfassen den zugeordneten Wert des Attributs, den Zeitpunkt, an dem diese Zuordnung erfolgte und die ID mit der die Patientenidentität eindeutig identifiziert wird. Die Überprüfung auf Änderungen an Attributen erfolgt, in dem alle Einträge zu dem zu überprüfenden Attribut aus der Tabelle des Macro-Audit Trails extrahiert werden. Nach der Extraktion wird die Gruppierung der Einträge nach Patientenidentitäten vorgenommen. Innerhalb jeder Gruppe wird dann nach den gespeicherten Zeitpunkten sortiert. Diese sortierten Gruppen werden ausgewertet. Dabei wird identifiziert, welche Änderungsoperationen auf einem Attribut stattgefunden haben. Sind die für das Attribut hinterlegten Operationen darunter, erfolgt eine temporäre Speicherung der ID der Patientenidentität, das betroffene Attribut und die Operation. Nach Abschluss der Überprüfung werden die temporär gespeicherten Informationen

ausgewertet und die registrierten Listener aufgerufen. Dabei wird eine Liste von IDs der betroffenen Patientenidentitäten und den diesen zugeordneten Attributen mit den Änderungsoperationen und Werten sowie Informationen zur Studie übermittelt. Die Implementierung dieser Funktionen erfolgt mit Java. Bei den Listenern handelt es sich um Webservices, die eine definierte Schnittstelle bereitstellen müssen. Die wsdl Locations der Listener Webservices werden in einer Datenbank persistiert. Der Zugriff auf die Macro Datenbank für das Auslesen der „Audit Trail"-Tabelle erfolgt über JDBC.

Die beschriebenen Komponenten für die Übernahme von Daten aus Systemen der Krankenversorgung in Forschungssysteme befinden sich im Produktiveinsatz. Dabei werden medizinische Daten, die im Rahmen von interdisziplinären Tumorboardsitzung in i.s.h.med für Weichteilsarkomerkrankungen erfasst und aufbereitet werden, in Macro übernommen [Lamla2010]. (Stand Juli 2011: für 252 Patienten wurden 308 Tumorboarddokumente übernommen). Dort werden diese um forschungsfragestellungsspezifische Daten erweitert. Die in der Tumorboardsitzung zu einem Patienten erfassten Daten beinhalten neben den demografischen Informationen, Daten zur Anamnese, der Tumorklassifikation, der Befundlokalisation und der histologischen Sicherung. Weitere Daten betreffen die am Patienten durchgeführten Therapien, wie Operationen, Strahlen- und Chemotherapie. Der Prozess zur Vorbereitung und Durchführung der Tumorboardsitzung in i.s.h.med gestaltet sich wie folgt. Der in der Tumorboardsitzung zu besprechende Patient wird im i.s.h.med-System von einem Arzt gesucht. Der Arzt navigiert zum aktuellen Fall des Patienten und ordnet diesem Fall, ein im System hinterlegtes Formular, mit dem die oben genannten Daten erfasst werden können, zu. Die Felder des Formulars, in denen die Stammdaten des Patienten erfasst werden, werden automatisch vom System vorbelegt. Die Daten zur Tumorklassifikation, histologischen Sicherung usw. erhält der Arzt aus den jeweiligen im i.s.h.med abgespeicherten Freitextbefunden des Patienten. Diese Daten „übernimmt" er in das Tumorboardformular für den entsprechenden Patienten, in dem er die entsprechend zur Verfügung stehenden Parameter durch die Auswahl von Checkboxen, Dropdown-Menüs und Radio-Buttons festlegt. Zu Beginn der Tumorboardsitzung werden alle Patienten im i.s.h.med System gesucht, denen ein noch nicht abgeschlossenes Weichteilsarkom-Tumorboardformular zugeordnet ist. Diese Patienten werden nun in der Sitzung besprochen. Eine Dokumentationskraft dokumentiert in den Tumorboardformularen die Beschlüsse, die in der Tumorboardsitzung für den entsprechenden Patienten

getroffen wurden. Dabei wird festgehalten, welche weiteren Therapien durchgeführt werden sollen, ob die Tumorklassifikation und die Metastasenlokalisation aktualisiert werden müssen, und welche Entscheidungsträger dies veranlasst haben. Nach der Tumorboardsitzung überprüft ein Oberarzt die ausgefüllten Tumorboarddokumente und setzt den Status der Dokumente auf „freigegeben“.

Einmal täglich wird überprüft, für welche Patienten neue Tumorboarddokumente freigegeben wurden. Für diese erfolgt dann die Extraktion aus i.s.h.med. Danach wird ermittelt, welcher Patienten- und Fallidentität in Macro die Daten aus den Dokumenten zugeordnet werden sollen. Dazu werden die Protokolle vorausgegangener Übernahmen überprüft. Werden keine Einträge gefunden, wird eine neue Patienten- und Fallidentität in Macro angelegt. Die IDs von Patienten- und Fallidentitäten sowie die extrahierten Daten werden in das von Macro benötigte Eingabeformat transformiert und an Macro zum Abspeichern übergeben.

Die Konsistenzsicherung auf Typ- und Instanzebene wurde prototypisch implementiert. Ein Generator zur Erzeugung der Metadatenmodelle zur Beschreibung der Schemas der den PMDs und eCRFs zugrunde liegenden Datenmodelle wurde entwickelt. Allerdings muss für die Generierung jedem Attribut in einem PMD ein Datentyp zugeordnet werden, der in der Definition des Metadatenmodells spezifiziert worden ist. Um diese Zuordnung bei PMDs, die in der Routine eingesetzt werden und über mehrere Hundert Attribute verfügen, nicht händisch durchführen zu müssen, ist es notwendig spezielle Werkzeuge dafür zu entwickeln. Für den Produktiveinsatz selbst, müssten zusätzlich noch Funktionalitäten für die Verwaltung der erzeugten Metamodelle umgesetzt werden, um Vergleiche der Modelle über Zeiträume auf Änderungen hin durchführen zu können. Zusätzlich müssten Prozesse festgelegt werden, in denen das weitere Vorgehen nach der Detektion von Änderungen an den Schemas beschrieben wird.

Für Konsistenzsicherung der Daten auf Instanzebene wurden Anfragen für die Extraktion der benötigten Informationen aus dem EAV-Schema des Macro-Audit Trails entwickelt. Zusätzlich wurde die Auswertung der Abfrageergebnisse für die Identifikation der durchgeführten Operation auf den Werten der Attribute entwickelt und getestet. Für die Verwendung der entwickelten Komponenten im Produktiveinsatz müssen Prozesse definiert werden. Dabei muss festgelegt werden, welche Person bei welchen Änderungsoperationen auf welchen Attributen basierend auf welchem Medium benachrichtigt werden sollte. Die Konsistenzsicherung von i.s.h.med Daten auf Instanzebene wird in der bisherigen Lösung dadurch sichergestellt, dass nur freigegebene Dokumente an denen keine Änderungen mehr

durchgeführt werden können, übernommen werden. Änderungen in i.s.h.med führen bei diesen Dokumenten zur Erzeugung eines neuen Dokuments mit identischer ID aber höherer Versionsnummer. Für den Produktiveinsatz müssen ebenfalls Prozesse festgelegt werden, in denen spezifiziert wird wer, wie benachrichtigt werden soll bzw. wie vorgegangen werden soll, wenn Daten aus einem Dokument mit höherer Versionsnummer zur Übernahme anstehen. Für die Konsistenzsicherung von Daten, die aus nicht freigegebenen i.s.h.med PMDs übernommen werden, gilt es für den Routineansatz taugliche Lösungen zu erarbeiten. So könnte nach Freigabe der entsprechenden Dokumente eine erneute Extraktion der Daten und ein Abgleich mit den bereits übernommenen, vorgenommen werden.

Die im Rahmen dieser Arbeit entwickelte Komponente zur Datenübernahme aus Routine- in Forschungssysteme ist Teil der m4 Software und befindet sich im Einsatz. Es wurden sowohl Labordaten als auch ICD und OPS für über 500 Patienten (Stand: Dezember 2012) aus IS-H/i.s.h.med in das Forschungssystem übernommen. In der m4 Software wurde das CDMS-System Macro durch eine selbst entwickelte EDC-Komponente für die Erfassung von Forschungsdaten abgelöst. Angedacht ist die Importfunktionalität der EDC-Komponente für Daten im CDISC-ODM-Format weiterzuentwickeln. Dafür werden die hier vorgestellten Funktionen, welche die aus i.s.h.med extrahierten Daten in das für Macro bzw. m4 spezifische Übernahmeformat transformieren, angepasst. Diese müssen dann die extrahierten Daten in CDISC-ODM umsetzen und dabei Informationen bzgl. der Abbildung der verschiedenen IDs von Patienten- und Fallidentitäten aus den Quell- und Zielsystemen mit verarbeiten. Auch die in der Arbeit entwickelten Funktionen, welche den Datentransfer protokollieren und dabei die Abbildung der verschieden IDs von Patienten- und Fallidentitäten aus den verschiedenen Systemen aufeinander persistieren, sind Teil der m4 Software. Allerdings wurden im Rahmen des in m4 vorgesehenen IT-Sicherheitskonzeptes Anpassungen vorgenommen. So dürfen die öffentlichen IDs der verschiedenen Identitäten zu einem Patienten nicht wie in der ursprünglichen Implementierung vorgesehen, in einer Datenbank abgespeichert werden. Es war notwendig, Abbildungen der originalen IDs auf Pseudonyme vorzunehmen und diese dann abzuspeichern.

4.3 Integration von Routine- u. Forschungssystemen auf Präsentationsebene

4.3.1 Komponentenarchitektur

Werden wie im Konzept vorgesehen identifizierende, medizinische sowie Biomaterialdaten in getrennten Teilsystemen gespeichert und erfolgt die Zusammenführung der Daten erst auf dem Computer des Anwenders so kann es notwendig sein, die Benutzerschnittstellen der einzelnen Subsysteme zu integrieren. Dies ist insbesondere dann relevant, wenn bei der Umsetzung des in Kapitel 3.1 beschriebenen Konzepts Funktionalitäten eines der Teilsysteme von einem „Fremdsystem“ zur Verfügung gestellt werden. So kann für das MDAT-Teilsystem ein separates CDMS zur Erfassung medizinischer Daten eingesetzt werden. Während die Darstellung der einzelnen Formulare, die Benutzerinteraktion bei der Eingabe der medizinischen Daten sowie deren Überprüfung auf Plausibilität über die GUI des CDMS erfolgt, ist diese wiederum Teil der GUI des Forschungssystems. Die erfassten medizinischen Daten werden im CDMS System abgespeichert. Die Zusammenführung der identifizierenden Daten aus dem IDAT-Teilsystem und der medizinischen Daten aus dem CDMS erfolgt dann in der Benutzeroberfläche. Hierfür zeigen beide Systeme die für eine selektierte Patientenidentität gespeicherten Daten in einer gemeinsamen GUI an. Auf diese Weise kann der Implementierungsaufwand für ein einzelnes Teilsystem reduziert werden. Neben der Möglichkeit den Implementierungsaufwand für die Umsetzung des Konzepts zu reduzieren, kann durch die Integration eines Forschungssystems mit Routinesystemen auf Präsentationsebene die gezielte Datenübernahme für einen Patienten durch den Anwender erleichtert werden. Hierbei werden für eine, im Forschungssystem durch den Anwender selektierte, Patientenidentität die dazugehörigen Daten im Routinesystem angezeigt. Der Anwender hat dann die Möglichkeit durch „Kopieren und Einfügen“ Daten aus den Routinesystemen ins Forschungssystem übernehmen.

Für die Integration der verschiedenen (Teil-)Systeme auf Präsentationsebene werden die in Abb. 10 beschriebenen Komponenten bereitgestellt. Die Komponenten kapseln die grafischen Benutzerschnittstellen der einzelnen (Teil-)Systeme. Die einzelnen Komponenten (Wrapper) stellen einheitliche Schnittstellen nach außen zur Verfügung (Wrapper Komponenten in Abb. 10). Darüber werden Daten, welche die

Identität im gekapselten System identifizieren sowie die Zugangsdaten des Anwenders, für den der Zugriff auf das System erfolgt, mit übergeben.

Nach seinem Start generiert der Wrapper ein Fenster. In dieses wird die grafische Benutzerschnittstelle des aufgerufenen (Teil)Systems gerendert. Der Wrapper übt über das neu erzeugte Fenster die Kontrolle aus. Die kontrollierten Parameter sind die Größe und Position (einschließlich der Z-Order) des Fensters, die Weitergabe von Ereignissen an Subfenster und die Möglichkeit das Fenster zu schließen. Der Wrapper übernimmt die „Navigation“ (Kontextwechsel) in der grafischen Benutzerschnittstelle, zur aufzurufenden Identität.

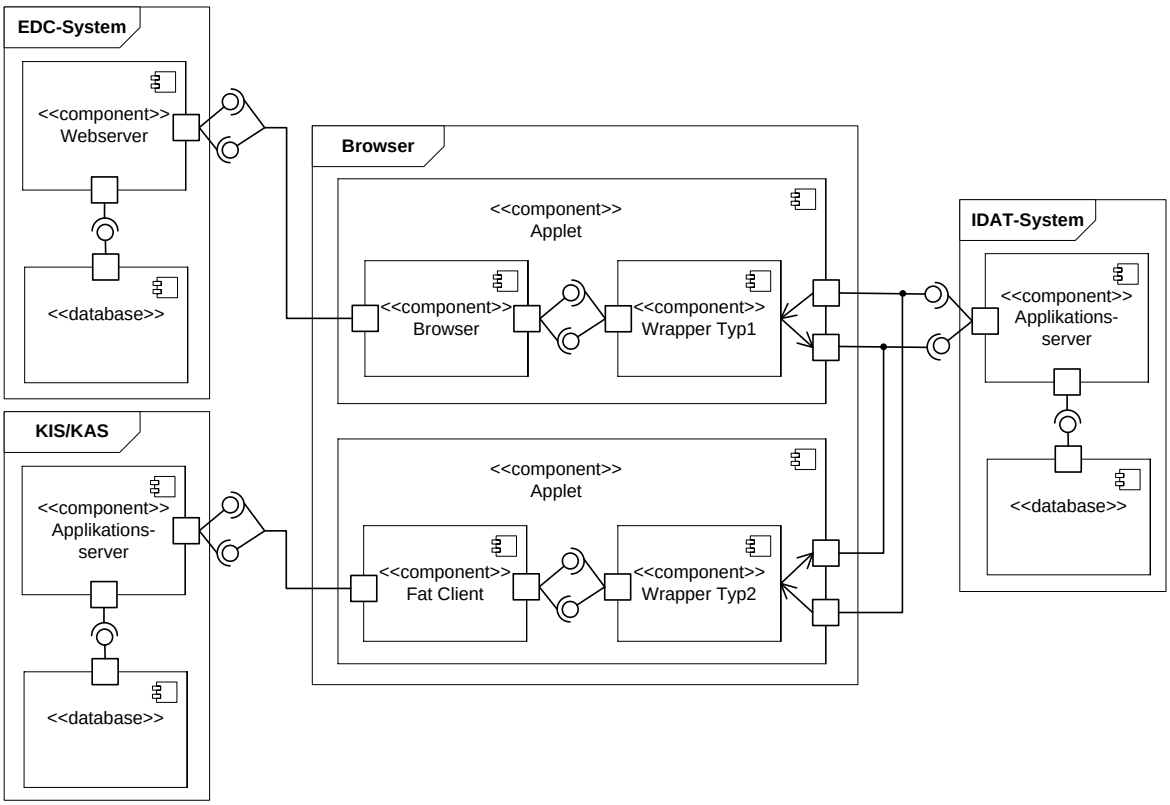


Abb. 10: UML Komponentendiagramm zur Beschreibung der Integration von Routine- u. Forschungssystemen auf Präsentationsebene

Bei der Umsetzung wird zwischen zwei Typen von Wrappern unterschieden. Typ 1-Wrapper integrieren grafische Benutzerschnittstellen von Systemen, die auf Basis von Webtechnologien realisiert sind und mit Hilfe von Web-Browsern visualisiert werden. Auf Basis dieses Wrappertyps können das CDMS-System Macro sowie die oben beschriebenen Teilsysteme zur Erfassung von identifizierenden Daten und Probendaten integriert werden.

Typ 2-Wrapper kapseln grafische Benutzerschnittstellen von Anwendungen, deren Clients lokal auf den Rechnern der Endanwender installiert sind. Mit einem Wrapper dieses Typs erfolgt die Integration der Benutzerschnittstelle von IS-H/i.s.h.med (SAP GUI).

4.3.2 Implementierung

Typ 1 Wrapper nutzen für die Darstellung der Benutzerschnittstelle, Funktionalitäten des lokal auf dem Computer des Endanwenders installierten Browsers. Dies umfasst die Kommunikation zwischen dem Browser und dem Server, der die Webseiten für die Realisierung der Benutzerschnittstelle des zu integrierenden Komponentensystems bereitstellt. Außerdem übernimmt er das Rendern der entsprechenden HTML/XHTML-Seiten, einschließlich des Ausführens von Java-Script, das Session Management (Session IDs oder Cookies) sowie die Visualisierung von Grafiken. Die vom lokal installierten Browser gerenderten Webseiten werden auf einer, vom Wrapper bereitgestellten Zeichenfläche, visualisiert. Der Wrapper reagiert auf Ereignisse die vom lokal installierten Browser ausgelöst werden. Dies geschieht, wenn das Laden einer Webseite begonnen bzw. abgeschlossen ist oder wenn der Browser beginnt Java-Script Funktionsaufrufe auszuführen bzw. damit fertig ist. Daneben kann der Wrapper auf Ereignisse reagieren, die von externen Anwendungen ausgelöst werden. Der Wrapper kann Aktionen im Browser antriggern. Diese Aktionen umfassen das Laden von Websites sowie das Ausführen von Java-Script Code.

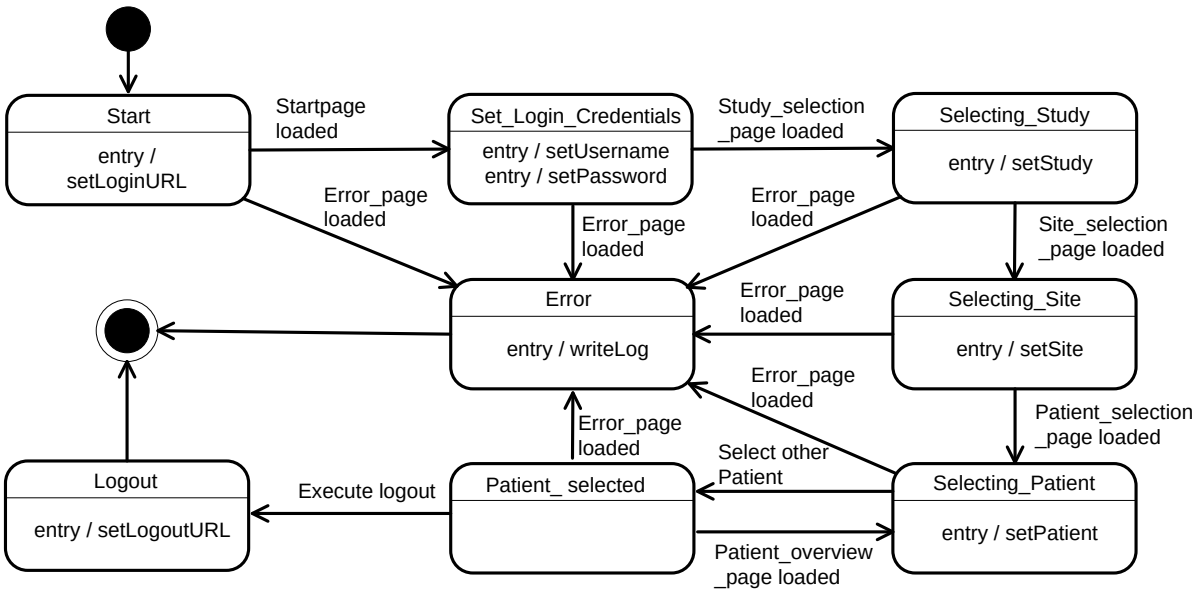


Abb. 11: UML Zustandsdiagramm zur Beschreibung eines kontextbezogenen Patientenwechsel

Die Integration auf Präsentationsebene von Komponentensystemen mit grafischer auf Webtechnologie basierender Benutzerschnittstelle erfolgt durch die Darstellung der vom Browser gerenderten Webseiten auf der Zeichenfläche des Wrappers. Die „Navigation“ in der integrierten Benutzerschnittstelle erfolgt mit Hilfe eines Zustandsautomaten. Ereignisse wie z.B. das vollständige Laden einer Webseite, führen zu Zustandsübergängen im Zustandsautomaten. Beim Erreichen eines Zustands wird eine Eingangsaktion im Zustandsautomaten ausgeführt, die mit einer Aktion im Browser assoziiert ist. Auf diese Weise ist es möglich, Benutzerinteraktionen im Browser durch den Wrapper auszulösen und deren Ausführung zu überwachen.

Im dargestellten Zustandsautomaten (Abb.: 11: UML Zustandsdiagramm) wird nach dessen Start eine Aktion ausgelöst, bei dem eine URL an den Browser übergeben wird (setLoginURL). Nachdem der Browser die entsprechende Webseite geladen hat (z.B. Loginseite), löst dieser ein Ereignis aus. Dieses führt zu einem Zustandsübergang. In Abb. 11 vom Zustand „Start“ in den Zustand „Set_Login_Credentials“. In diesem Zustand werden Aktionen durchgeführt bei denen durch die Aktion „setUsername“ JavaScript vom Zustandsautomaten an den Browser übergeben wird. Dieses wird vom Browser ausgeführt. Das in diesem Zustand übergebene JavaScript veranlasst den Browser z.B. in ein Textfeld auf der dargestellten Webseite Werte (Username) einzutragen.

Für die Beschreibung des Zustandsautomaten wird bei der Umsetzung die XML-basierte Markup-Sprache „State Chart XML“: State Machine Notation for Control Abstraction (SCXML) [SCXML] verwendet. Die im Wrapper implementierte Ausführungsumgebung für die Ausführung von mit SCXML beschriebenen Zustandsautomaten steuert durch das Auslösen von Aktionen (URLs bzw. JavaScript werden an den Browser übergeben) und durch das Reagieren auf Ereignisse (z.B. das Laden einer Webseite ist abgeschlossen) den lokal auf dem Computer des Endanwenders installierten Browser fern und führt den patientenzentrierten Kontextwechsel in der zu integrierenden Anwendung durch. Beendet der Endanwender die Wrapperanwendung, so führt dies im Zustandsautomaten zu einem Ereignis, in dem Aktionen für das Ausloggen aus der integrierten Anwendung veranlasst werden.

Die Implementierung des Wrappers erfolgt mit Java u. a. in Form eines Applets. Damit in diesem Fall die Java-Runtime-Environment Zugriffe auf die Schnittstellen des lokal installierten Browsers gestattet, wird der Wrapper signiert. Der Zugriff auf den Browser erfolgt über dessen „Component Object Model“ (COM) Schnittstellen.

Für den Zugriff auf die COM-Schnittstellen werden Komponenten aus dem „Standard Widget Toolkit“ (SWT) [SWT] eingesetzt. Diese verwenden das „Java Native Interface“ (JNI), um den eigentlichen Zugriff aus Java heraus zu realisieren.

Typ2-Wrapper kapseln Benutzerschnittstellen von Anwendungen, welche als Fat-Client Anwendungen auf den Computern der Endanwender installiert sind. Bei der Beschreibung der Umsetzung wird davon ausgegangen, dass auf dem Computer des Endanwenders Microsoft Windows installiert ist.

Der Wrapper ermittelt, ob und in welchem Verzeichnis die zu integrierende Fat-Client-Anwendung installiert ist, in dem er das Softwareregistrierungsverzeichnis (Windows Registry) durchsucht. Ist die Anwendung vorhanden, startet der Wrapper diese und bestimmt die Prozess-ID unter welcher die Anwendung vom Betriebssystem verwaltet wird. Davon ausgehend, ermittelt er das Hauptfenster der Anwendung und bestimmt dessen Window-Handler. Dabei handelt es sich um einen Identifikator, mit dem das Betriebssystem ein Fenster Objekt eindeutig identifiziert. Ausgehend vom Hauptfenster werden diejenigen Kindfenster einschließlich deren Window-Handler ermittelt, die für den patientenzentrierten Kontextwechsel benötigt werden und bereits mit dem Aufruf des Hauptfensters zur Verfügung stehen. Danach erzeugt der Wrapper eine Zeichenfläche mit den gewünschten Dimensionen, auf der das Hauptfenster der zu integrierenden Anwendung gezeichnet werden soll. Für diese Zeichenfläche bestimmt der Wrapper den assoziierten Window-Handler. Im Anschluss daran, wird die Zeichenfläche des Wrappers zum Elternfenster des Hauptfensters der Anwendung. Das Hauptfenster wird zum Maximieren und Neuzeichnen aufgefordert. Das Ändern der Eltern-Kind Beziehung zwischen Hauptfenster der Anwendung und dem Desktop geschieht dadurch, dass der Handler des Elternfensters des Hauptfensters der Anwendung (Desktop Fenster des Betriebssystems) durch den Handler der Wrapper Zeichenfläche ersetzt wird. Die Manipulationen der Handler und das Maximieren mit dem damit verbundenen Neuzeichnen des Hauptfensters wird durch den Aufruf entsprechender Methoden des Betriebssystemkernels erreicht (user32.dll).

Der Wrapper ist mit Java implementiert. Der Zugriff auf Methoden des Betriebssystemkernels erfolgt über eine in C++ entwickelte Komponente, die über JNI mit dem in Java implementierten Wrapper kommuniziert.

Für die „Installation“ der Typ1- und Typ2-Wrapper auf den lokalen Computern der Endanwender sind verschiedene Szenarien vorgesehen. Ist der Wrapper Teil einer Web-Anwendung, erfolgt die „Installation“ temporär. Dazu wird der Wrapper beim Aufruf der entsprechenden Seite in der Webanwendung lokal auf den Rechner des

Endanwenders heruntergeladen und dort in einer Laufzeitumgebung ausgeführt. Der Wrapper kann den Server, von dem er heruntergeladen wird, identifizieren und mit diesem kommunizieren. Nach Beendigung der Webanwendung wird der Wrapper gelöscht. Um die Downloadzeiten für die Wrapper im Produktiveinsatz zu verbessern und Netzwerkressourcen zu schonen, ist auch die Möglichkeit vorgesehen den Wrapper lokal in einem entsprechenden Cache vorzuhalten. Der Wrapper kann auch Bestandteil einer Fat-Client Anwendung sein. Die Installation erfolgt dann entsprechend.

Bei der Kommunikation zwischen den auf die Client Rechner der Endanwender heruntergeladenen, Wrappern und dem Applikationsserver von dem sie heruntergeladen werden, kommen zwei verschieden Kommunikationsmodelle zum Einsatz. Bei der Kommunikation zwischen den Wrappern und dem Applikationsserver werden Informationen ausgetauscht, die der Wrapper für den patientenzentrierten Kontextwechsel in der vom ihm gekapselten Anwendung benötigt. Dies umfasst die Zugangsdaten des Anwenders, der die Daten des entsprechenden Patienten in der über den Wrapper integrierten Anwendung einsehen möchte. Zusätzlich benötigt der Wrapper Daten, die den gewünschten Patienten eindeutig identifizieren. Im Push-Modell lässt sich der Wrapper vom Applikationsserver mitteilen, wenn der Endanwender innerhalb einer Anwendung einen Kontextwechsel zu einem Patienten durchführt. Dadurch kann der Wrapper, in der von ihm integrierten Anwendung, den Kontextwechsel zum selben Zielpatienten durchführen. In einem ersten Schritt ermittelt der Wrapper dazu die IP-Adresse und den Port des Applikationsservers von dem er heruntergeladen wurde. Zusätzlich speichert der Wrapper, die vom Server vergebene ID für die vom Endanwender gestartete Session. Mit dieser SessionID identifiziert sich der Wrapper gegenüber dem Server. Damit der Wrapper vom Server über durchzuführende Kontextänderungen informiert werden kann, ermittelt der Wrapper die IP-Adresse des Computers, auf dem er ausgeführt wird und legt einen Port fest an dem er auf Nachrichten vom Server lauscht. Diese Informationen teilt er dem Server unter Angabe der SessionID mit. Im Pull-Modell verzichtet der Wrapper auf die Ermittlung der lokalen IP-Adresse und der Festlegung eines Ports. Der Wrapper beschränkt sich darauf unter Verwendung der SessionID periodisch beim Server nach durchgeführten Kontextwechseln anzufragen.

Für die beschriebenen Wrapper liegen prototypische Implementierungen vor. Im Rahmen dieser Implementierungen wurde ein Identitätsmanagementdienst erweitert,

so dass dieser die Wrapper integriert und mit diesen kommunizieren kann. Zusätzlich wurde der Identitätsmanagementdienst noch um Funktionalitäten für die Verwaltung von Zugangsdaten der Anwender in Anwendungen ergänzt. Damit können den Wrappern Zugangsdaten der Anwender für die integrierten Anwendungen bereitgestellt werden. Der Endanwender kann über den Identitätsmanagementdienst die Abbildungen der verschiedenen IDs der Patientenidentitäten in den verschiedenen, über die Wrapper integrierten, Anwendungen managen. Durch Auswahl eines Patienten im Identitätsmanagementdienst werden die Anwendungen mit Hilfe der Wrapper gestartet, deren Benutzerschnittstellen integriert und die entsprechende Patientenidentität in den Anwendungen aufgerufen. Die Benutzerschnittstellen der aufgerufenen Anwendungen werden in die Benutzerschnittstelle des Identitätsverwaltungsdienstes integriert. Prototypisch ist die Integration der Benutzerschnittstellen für die SAP-GUI (Fat-Client: implementiert die grafische Benutzerschnittstelle für IS-H/i.s.h.med) für die EDC-Benutzerschnittstelle von Macro, die Benutzerschnittstellen der Biomaterialverwaltung der Pathologie und der Pankreas-Forschungsdatensammlung (beides Web-Technologie) umgesetzt.

Für den Produktiveinsatz müssten die Wrapper um Fehlerbehandlungsfunktionalitäten erweitert werden, um die Robustheit zu erhöhen. Dazu gehören Maßnahmen mit denen bereits laufende Instanzen der Wrapper und Anwendungen identifiziert werden. Weitere Maßnahmen umfassen ein ordnungsgemäßes Abmelden der Anwender in den aufgerufenen Anwendungen, einschließlich der Freigabe der gekapselten GUIs beim Schließen des Browsers und der daraus resultierenden Terminierung der Wrapper. Zusätzliche, vor dem Produktiveinsatz zu untersuchende Rahmenbedingungen, betreffen das Antwortzeitverhalten der zu integrierenden Anwendungen. So kann es in der Praxis vorkommen, dass der Aufruf von Informationen zu einem Patienten über die grafische Benutzerschnittstelle in den Produktivsystemen in ungünstigen Fällen über eine Minute dauert.

4.4 Aufbau eines Metadatenlayers zum Abgleich von Quell- u. Zielsystemen

4.4.1 Komponentenarchitektur

Zum Abgleich von Quell- und Zielsystemen wird ein Layer aufgebaut, das die Weiterverwendung von Metadaten elektronischer Formulare zur Erfassung von Behandlungsdaten ermöglicht. Dies soll die Erstellung von elektronischen Formularen in Forschungssysteme vereinfachen. Ziel ist die Definition entsprechender Felder auf eCRFs die durch die Übernahme von Daten aus IT-Systemen der Krankenversorgung (Quellsysteme) in Forschungssysteme (Zielsysteme) befüllt werden und dort eine Teilmenge der insgesamt zu erhebenden Forschungsdaten bilden.

Von besonderer Bedeutung sind hierbei die dafür vom KIS/KAS bereitgestellten Formulare. Diese unterliegen, verglichen mit den in den (Sub)-Systemen verwendeten Formularen, häufigeren Änderungen. Dies liegt u.a. an der hohen Dynamik in der Domäne. Medizinische Strukturen und Prozesse unterliegen ständigen Veränderungen, bspw. durch die Einführung neuer diagnostischer und therapeutischer Prozeduren oder sich ändernden Rahmenbedingungen (z.B. neue Dokumentationsvorschriften für Qualitätsberichte) [Lenz2004].

Für den Aufbau des Metadatenlayers werden in einem ersten Schritt Metadaten der Formulare aus den Routinesystemen extrahiert (BPMN Diagramm Abb. 12: erste Aufgabe in untersten Lane). Dazu werden die in Abb. 8 beschrieben KIS/KAS Konnektorkomponenten erweitert. Der Formularname, die Beschriftungen, die Längen, die Datentypen sowie im Falle von Auswahllisten die dafür hinterlegten Vokabulare sind die für die Felder auszulesenden Metadaten. Nach der Extraktion werden die Daten in einem zweiten Schritt in ein „Zwischenformat“ übertragen (mittlere Lane). Im Rahmen dieser Arbeit wird CDISC-ODM als „Zwischenformat“ eingesetzt. CDISC-ODM ist ein Format für den Austausch und die Archivierung von Daten aus klinischen Studien. Im Metadatenteil von CDISC-ODM werden die einzelnen Felder der eCRFs spezifiziert. Die eCRFs selbst werden wiederum Fällen und Studien zugeordnet. Für die Übertragung in das „Zwischenformat“ wird eine Komponente bereitgestellt mit der Felder aus den Formularen der Routinesysteme ausgewählt und die dazugehörigen Metadaten in CDISC-ODM übernommen werden. Nach der Übertragung der Metadaten aus den Routinesystemen in CDISC-ODM erfolgt in einem dritten Schritt das Einlesen der eCRF-Spezifikation in das

EDC-System (s. Prozessschritt in oberste Lane). Im vierten Schritt werden die im EDC-System neuangelegten Formulare um zusätzliche Formulare und Felder für Inhalte, die nicht aus den Routinesystemen übernommen werden, ergänzt.

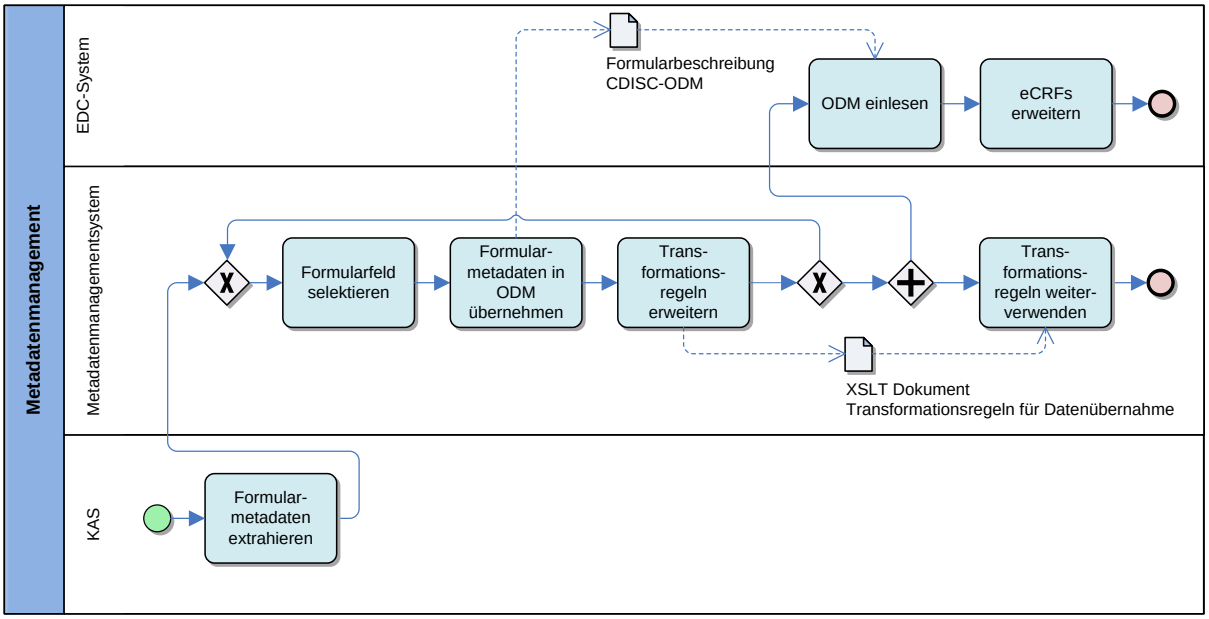


Abb. 12: BPMN-Modell zur Beschreibung des Prozesses für den Aufbau eines Metadatenlayers zum Abgleich von Quell- u. Zielsysteme

Die während des Übertragungsprozesses der Metadaten in CDISC-ODM festgelegten Abbildungsregeln werden für die Datenübernahme zwischen Routine- und EDC-System weitergenutzt. Im Rahmen des Übertragungsprozesses werden Mappings definiert die bestimmen welches Feld aus den Formularen des Quellsystems auf welches Feld in welchem Formular des Zielsystems abgebildet wird. Falls notwendig, werden die Mappings um zusätzliche Transformationsregeln erweitert. Die Transformationsregeln selbst werden dann wiederum in dem in Kapitel 4.2 beschriebenen Datenübernahmeprozess zwischen IT-Systemen in der Krankenversorgung und Forschungssystemen weiterverwendet.

4.4.2 Implementierung

Für die Umsetzung werden Funktionen für die Extraktion von Metadaten der PMDs aus i.s.h.med bereitgestellt. Die Metadaten für ein PMD-Feld beinhalten die Beschriftung des Formularfelds auf dem Formular, den internen Feldbezeichner, den Datentypen und bei Auswahllisten, die dafür hinterlegten Vokabulare, einschließlich der Positionen der einzelnen Terme in den Auswahllisten. Für die Extraktion von Metadaten wird die in Kapitel 4.2.2 beschriebene Funktion für den Zugriff auf PMDs erweitert. Diese Funktion liefert ein XML-Dokument zurück. In diesem XML-

Dokument finden sich alle, über PMDs erfasste, Dokumente eines bestimmten Typs zu einem Patienten. Jeder Kindknoten des Wurzelknotens des XML-Dokuments repräsentiert ein extrahiertes Dokument. Die Kindknoten der „Dokumentenknoten“ enthalten Beschreibungen der entsprechenden PMD Felder und die zugeordneten Werte, die an den entsprechenden Positionen in den PMDs enthalten sind. Ist einem Feld im PMD kein Wert zugeordnet, so findet sich an der entsprechenden Position im XML-Dokument ein XML-Element ohne zugewiesenem Wert. Die Attribute des XML-Elements verweisen auf die Tabelle und die Spalte in der der Wert aus dem PMD gespeichert ist. Der Namen der Tabelle und der Spalte werden benutzt, um Metadatentabellen in i.s.h.med zu durchsuchen. In diesen werden die in den Spalten einer Tabelle erfassten Daten näher charakterisiert. Dazu existiert in i.s.h.med für eine einzelne Spalte einer Datentabelle ein entsprechendes Datenelement. Dieses wird mit der ersten Anfrage ermittelt. In dem Datenelement sind dessen interner Name, der dazugehörige Domänenname, der Datentyp und die Länge des Datenfeldes sowie die Beschriftung des Formularfeldes hinterlegt. Ist für ein PMD-Feld ein Vokabular hinterlegt, so findet sich ein Verweis auf dessen Terme in den korrespondierenden Datenelementen. Diese Informationen werden mit Hilfe einer zweiten Anfrage extrahiert. Die Extraktion der Metadaten beruht auf dem Aufruf eines SAP-RFC. Die Rückgabe der Anfragen enthält folgende Metainformationen zu einem PMD-Feld: Den internen Feldbezeichner, die Beschriftung des Feldes auf dem PMD, den Datentyp des Wertes, der über das Feld erfasst werden kann sowie bei Auswahlfeldern eine Liste des Vokabulars, welches dem Endanwender zur Auswahl steht. Für jeden Term des Vokabulars stehen der Name, der im PMD angezeigt wird, der Code, der bei Auswahl des Terms in die Tabelle geschrieben wird sowie die Position des Terms, falls dieser z.B. in einer Drop-Down Liste visualisiert wird zur Verfügung.

Diese Informationen werden in den CDISC-ODM Metadatenteil übernommen. Dabei erfolgt eine Abbildung: Der interne Feldbezeichner für ein Feld auf dem PMD wird auf eine „Eintragsdefinition (ItemDef)“ im ODM-Format abgebildet. Der Feldname wird auf das „OID-Attribut“ im „ItemDef Element“ gemappt. Die Beschriftung des Feldes auf dem PMD wird auf das Attribut „Name“ des „ItemDef Elements“ abgebildet. Handelt es sich bei dem Datentypen des PMD-Feldes um einen elementaren Datentypen, findet eine einfache Umsetzung satt. Handelt es sich um eine Auswahlliste, für die mehrere vordefinierte Werte existieren, werden die Terme des hinterlegten Vokabulars in „CodeListItems-Elemente“ des ODM-Formats transformiert. Die einzelnen „CodeListItem-Elemente“ werden „CodeList-Elementen“

zugeordnet, welche aus den „ItemDef-Elementen“ heraus referenziert werden. Die einzelnen „ItemDef-Elemente“ werden in Gruppen zusammengefasst („ItemGroupDef“). Diese Gruppen werden von den Elementen, welche Metainformationen zu einzelnen Formularen („FormDef“) enthalten, referenziert. Einzelne Formulardefinitionen werden wiederum Elementen („StudyEventDef“) zugeordnet, über die die Reihenfolgen festgelegt werden, in der die Formulare angeordnet werden sollen.

Der Prozess zur Erstellung des „Zwischenformats“ für die Spezifikation von eCRFs gestaltet sich wie folgt. Im ersten Prozessschritt erfolgt die Extraktion eines PMDs mit dem in i.s.h.med Daten zu Patienten erfasst und die für Forschungszwecke in EDC-Systeme übernommen werden sollen. Im zweiten Prozessschritt werden die benötigten Felder im extrahierten PMD (XML-Dokument) ausgewählt. Durch die Auswahl eines Feldes im PMD erfolgt die Extraktion der Metadaten, in dem die Namen der Tabelle und der Spalte des Felds an die Extraktionsfunktion übergeben werden. Die zurückgegeben Metadaten werden wie oben beschrieben auf die einzelnen ODM-Metadaten-Elemente abgebildet. Als Ergebnis des Prozessschritts steht ein CDISC-ODM Dokument mit Metadaten zur Verfügung. Im dritten Prozessschritt werden Transformationsvorschriften als Nebenprodukt der Weiterverwendung von Metadaten für die Datenübernahme abgeleitet. Diese Transformationsvorschriften werden mit XSLT beschrieben. Durch die Auswahl eines PMD-Feldes im extrahierten XML-Dokument, kann dessen Position innerhalb des Dokuments (Baumstruktur) bestimmt werden. Die Position wird als XPath-Ausdruck beschrieben. Dieser Ausdruck wird erweitert, so dass er auf den Wert des ausgewählten XML-Elements (PMD-Feld) verweist. Aus den spezifizierten ODM-Metadaten wird ein Dummy für ODM-Instanzdaten abgeleitet, der in ein XSLT-Dokument überführt wird. Durch die Abbildung des PMD-Feldes auf das korrespondierende Feld im ODM-Metadatenteil ist die Zuordnung bekannt. Dadurch kann auf Instanzebene festgelegt werden an welcher Stelle der Wert aus dem PMD-Feld in den ODM-Instanzdatensatz geschrieben werden muss. Dies wird dadurch erreicht, dass im mit XSLT beschriebenen ODM-Instanzdatensatz für den Wert des Value-Attributs des ODM-„ItemData-Elements“ der entsprechende XPath-Ausdruck eingesetzt wird. Damit steht ein XSLT-Dokument zur Verfügung, dass aus einem PMD-Dokument, welches in XML-Repräsentation vorliegt, ein CDISC-ODM Dokument erzeugt.

Das Ergebnis des Prozesses zur Erstellung des „Zwischenformats“ umfasst somit zwei XML Dokumente. Das erste Dokument beschreibt eine „Studie“ mit einem

eCRF und dessen Felder im CDISC-ODM Format. Unter Verwendung dieser Beschreibung kann im EDC-System der eCRF automatisiert angelegt werden. Das zweite Dokument ist ein XSLT-Dokument auf dessen Basis XML-PMD-Dokumente in CDISC-ODM-Dokumente transformiert und damit medizinische Daten im Rahmen des Datenübernahmeprozesses ins EDC-System eingelesen werden können.

Die im Rahmen der vorgelegten Arbeit entwickelte Extraktionskomponente für PMDs befindet sich im Einsatz. Mit ihr werden die in Tumorboardsitzungen für Weichteilsarkome erfassten Daten aus i.s.h.med in das Forschungssystem Macro übernommen (s. Kapitel 4.2). Die Erweiterung der Komponente zur Extraktion von Metadaten für PMDs ist prototypisch umgesetzt. Damit lassen sich für jedes Feld aus einem PMD die in i.s.h.med hinterlegten Metadaten (Name des Feldes auf dem Formular, Datentyp, Länge des Feldes, Liste an Auswahlmöglichkeiten falls vorhanden) extrahieren. Die Übernahme der Metadaten in das CDISC-ODM Zwischenformat und die Erstellung der XSLT-Beschreibung der Transformationsvorschriften wurden prototypisch umgesetzt. Hierfür wurde ein Prototyp für ein graphisches Werkzeug implementiert. Dieses stellt die XML-Repräsentation des PMDs und des CDISC-ODM-Zwischenformats in Baumform dar. Per „Drag and Drop“ können Felder aus dem PMD in den Metadatenteil von CDISC-ODM übernommen bzw. bei Fehlern Abbildungen verworfen werden. Im Rahmen des Erstellungsprozesses werden die Metadaten zu dem im PMD selektierten Feld aus i.s.h.med extrahiert und als Knoten im „CDISC-ODM-Baum“ eingefügt.

5 Diskussion

5.1 Diskussion der entwickelten Methoden

Die Verbundforschung und die Translationsforschung haben in der Medizin in den letzten Jahren national und international neue Anforderungen an die Informatik gestellt und innovative Konzepte und Lösungen initiiert. Im Folgenden soll die Positionierung der vorgelegten Arbeit in diesem Umfeld erörtert werden.

5.1.1 Identitätsmanagement in verteilten Umgebungen

Die Anwendungsdomäne ist charakterisiert durch heterogene verteilte Umgebungen, die sich dynamisch verändern. Für das Zusammenführen von Daten oder auch Bioproben im zeitlichen Verlauf ist das Identitätsmanagement ein zentrales Element. Die Grundidee des vorgestellten Ansatzes ist es, auf identifizierende und weitere Daten aus verschiedenen klinischen (oder auch Forschungs-) Systemen so zuzugreifen, dass ein übergeordnetes abgesichertes und regularienkonformes Identitätsmanagement und eine Übernahme von Daten im Rahmen von Berechtigungen möglich werden.

- Identifizierende Daten und die zugehörigen Falldaten bilden in der Krankenversorgung zusammen mit der Dokumentation von Diagnosen und Maßnahmen die Grundlage der Abrechnung. Sie werden deswegen mit hoher Sorgfalt durch ausgebildetes Personal strukturiert erfasst. Demografische Daten werden aus den Krankenversicherungskarten automatisiert eingelesen, und Fehler werden zeitnah korrigiert. Die Datenqualität unterliegt zudem bis zur Abrechnung einer Reihe von Prüfungen. In Forschungssystemen werden identifizierende Daten häufig separat eingegeben, verwaltet und geprüft. Ihre Qualität kann variieren, ist aber generell ebenfalls als hoch einzustufen. Eine Verwendung klinischer Identifikatoren im Forschungskontext ist v.a. aus zwei Gründen sinnvoll: Die Übernahme der in hoher Qualität vorhandenen demographischen Daten hilft beim Anlegen einer Forschungsdokumentation und ermöglicht es, verschiedene Dokumentationen konsistent zu halten. Das

Mitführen klinischer Identifikatoren unterstützt zudem die Übernahme klinischer Daten. Entscheidend ist dabei, dass die rechtliche Grundlage, im Rahmen dieser Arbeit zumeist die explizite Einwilligung, dafür vorhanden ist.

- Im klinischen Bereich werden Korrekturen von identifizierenden Daten und von Falldaten im führenden Patientendatenmanagementsystem (PDMS) vorgenommen (z.B. Patientenzusammenführung, verschiedene Fälle von „Reconciliation“ bei fehlerhaften Stammdaten einschl. versehentlicher Neuaufnahme). Weitere klinische Systeme werden durch das führende Patientendatenmanagementsystem benachrichtigt. Die hierfür übermittelten Nachrichten können ausgewertet werden, um die Änderungen auch im Identitätsmanagement der Forschungssysteme zu berücksichtigen. Das Nachführen solcher Änderungen in der Forschungsumgebung ist ein wesentliches Element des vorgelegten Konzepts.
- Typischerweise werden in klinischen Systemen verschiedene Identifikatoren verwendet, insbesondere die Personen-ID und die Fall-ID des Patientendatenmanagementsystems, zusätzlich etwa die Fall-IDs von Pathologiesystemen. Das hier vorgestellte Konzept umfasst die Verwendung dieser Identifikatoren. Die Arbeit schafft eine Grundlage für ihre sicherheitstechnische bzw. kryptographische Handhabung, behandelt diese aber nicht. Im Allgemeinen sind klinische Identifikationsnummern keineswegs geheim, und sie müssen deswegen wie Name und Geburtsdatum gesichert werden.
- Soweit Übernahmen von identifizierenden und begleitenden Daten durch die Rahmenvorgaben (das Zugriffsrecht basiert auf einer Einwilligungserklärung, der Zugriff erfolgt dementsprechend rollenbasiert) zulässig sind, können sie gemäß dem hier vorgelegten Konzept automatisiert vorgenommen werden. Dies hat gegenüber getrennt verwalteten und bei Bedarf manuell abgeglichenen Identitäten naturgemäß erhebliche Vorteile.
- Maßgeblich für die Rechte, im Forschungskontext in einem klinischen System nach Patienten oder Fällen zu suchen, sind entweder gesetzliche Grundlagen (laut BayKhG ist Forschung mit Daten der Klinik erlaubt, solange diese das Haus nicht verlassen) oder das Vorliegen einer Einwilligungserklärung. Diese Einwilligung kann im Forschungssystem oder im klinischen System dokumentiert sein. Alle (drei) Optionen werden in Projekten verfolgt und im Rahmen der vorgelegten Arbeit unterstützt.

- Die im Rahmen der Arbeit konzipierte und aufgebaute Komponente zum Identitätsmanagement ist komplex. Sie berücksichtigt Identitäten von Personen, Fällen, Proben, verschiedene Fallstrukturen sowie verschiedene Einwilligungserklärungen.

Im Rahmen der Arbeit wurden für den eigentlichen Zugriff zwei Varianten betrachtet.

- Option 1: Direktzugriff auf das führende PDMS und weitere klinische Systeme. Bevorzugt eingesetzt wird der Zugriff auf ein Patientendatenmanagementsystem über eine Programmierschnittelle, der sofort die gewünschten Daten liefert. Er ist allerdings systemspezifisch, d.h. er muss ggf. an Systeme anderer Hersteller angepasst werden. Dies kann aufwendig sein, da auch Berechtigungsprüfungen mit berücksichtigt werden müssen.
- Option 2: Zugriff auf den HL7-Strom. Als gegenüber Option 1 einfacher zu handhabende Alternative wurde eine Lösung realisiert, die die Nachrichtenströme zwischen den klinischen Systemen überwacht. Ihr Vorteil liegt darin, dass sie auf dem weit verbreiteten Nachrichtenstandard HL7 aufsetzt, also weniger herstellerspezifische Modifikationen erfordert. Die Einschränkung liegt allerdings im Inhalt des Datenstroms selbst: Es ist nicht generell sichergestellt, dass alle interessierenden Daten im HL7-Strom übertragen werden. Wichtig ist auch, dass nicht davon ausgegangen werden kann, dass das Feld „Einwilligung für Studie xy vorhanden“ mit übertragen wird. Das zugreifende System muss aber sicherstellen, dass es nur Daten zu solchen Patienten filtert, für die eine Einwilligung vorliegt. Dies wird ermöglicht über eine Verwaltung klinischer Fall-IDs im Forschungssystem. Um sie noch vor der Übernahme der demographischen Daten fehlerfrei einlesen zu können, werden Barcode-Scanner eingesetzt, die klinische IDs nach Vorliegen einer Einwilligung direkt übernehmen können.

Eine wesentliche Erweiterung, die angedacht ist, aber über den Rahmen dieser Arbeit hinausgeht, liegt in der Einbindung von Record Linkage Ansätzen, z.B. [Schnell2009]. Bei einer Privacy Preserving Record Linkage besteht die Grundidee darin, einen identifizierenden Kerndatensatz in einem System zu suchen (bzw. eine linkage herzustellen), auf das nur eingeschränkte Zugriffsrechte bestehen.

Die im Rahmen dieser Arbeit entwickelten Konzepte und Lösungen stellen sich im Vergleich zu wichtigen anderen Systemen folgendermaßen dar:

- Vergleich mit dem SD-System der Vanderbilt University:

  Im „SD-Ansatz“ von Vanderbilt wird einem Patientenidentifikator im de-identifizierten Forschungsdatenbestand ein Hashwert zugeordnet, der sich aus der Anwendung einer Einweg-Hashfunktion auf die vom KIS vergebene PatientenID errechnet. Damit können, ebenso wie im Ansatz dieser Arbeit, Änderungen/ Ergänzungen, die nach der Datenübernahme an Behandlungsdaten eines Patienten vorgenommen werden in den Forschungsdaten nachgezogen werden. Dieser Fall tritt z.B. dann ein, wenn Patientendaten aus Folgebehandlungen übertragen werden sollen, nachdem die erste Übernahme bereits vorgenommen wurde. Im Detail ergeben sich allerdings Unterschiede: Im „SD-Ansatz“ ist eine Re-identifikation eines Patienten zurück aus den Forschungsdaten nicht möglich; es kann also nicht für einen einzelnen Patienten im Klinikum angefragt werden, ob Änderungen an den dortigen Daten vorgenommen wurden: Eine u.U. sukzessive, nachfragebasierte Datenübernahme, die aus dem Forschungssystem getriggert wird, ist nicht möglich. Da im „SD-Ansatz“ keine Re-identifikation möglich ist, können einer Entität in den Forschungsdaten auch keine Daten aus anderen Einrichtungen direkt zugeordnet werden. Ein einrichtungsübergreifendes Identitätsmanagement für Forschungszwecke ohne die Einbeziehung der führenden klinischen Systeme wird mit dem „SD-Ansatz“ nicht unterstützt. Für eine Übernahme von Daten aus anderen Einrichtungen müssten die entsprechenden Patienten zunächst im führenden klinischen System (KIS) identifiziert und danach die SD-ID errechnet werden. Dies könnte zu datenschutzrechtlichen Fragen führen, da für Zusammenführung und Zugriff im klinischen Umfeld Rechtsgrundlagen vorhanden sein müssen.

  Durch das in dieser Arbeit vorgestellte separate Identitätsmanagement für Forschungssysteme ist ein Rückgriff auf das führende klinische System für eine Zusammenführung von Daten aus verschiedenen Quellsystemen nicht erforderlich: Auf der Basis der Patienteneinwilligung kann das Forschungssystem Daten übernehmen und für eine Zusammenführung klinische Identifikatoren verwenden. Es ist allerdings notwendig, dass die identifizierenden klinischen Daten durch komplexe Sicherheitsmaßnahmen

(doppelte Pseudonymisierung, Kryptographie, räumliche und organisatorische Trennung von Komponenten) geschützt werden.

Ein Vorteil des „SD-Ansatzes“ besteht somit darin, dass keine Speicherung der Abbildung zwischen der Patienten-IDs aus dem KIS und den korrespondierenden Pseudonymen in den de-identifizierten Daten notwendig ist. Dies vereinfacht die Umsetzung, da kein separates Identitätsmanagement sowie eine zweistufige Pseudonymisierung implementiert werden muss.

Mit dem „SD-Ansatz“ ist die Einführung eines Erfassungsprozesses für Biomaterialien, die im Rahmen der Routinebehandlung analysiert und befundet werden, einfacher umzusetzen als mit dem in der Arbeit vorgestellten. In den Routineprozessen zur Befundung sind den Biomaterialien üblicherweise bereits Patienten-IDs aus dem KIS zugeordnet. Davon ausgehend muss im einfachsten Fall im Rahmen des Einlagerungsprozesses des Biomaterials für Forschungszwecke lediglich eine Umetikettierung an den Biomaterialbehältern durchgeführt werden. Das Pseudonym für das neue Etikett wird aus der bereits vorhandenen Patienten-ID errechnet.

Im Gegensatz dazu ist im vorgestellten Ansatz zuerst eine Verknüpfung zwischen Patienten- und Probenidentität in den Forschungssystemen für identifizierende- und Probendaten zu erstellen. Im Folgeschritt können dann die vom Forschungssystem zur Verwaltung von Biomaterialien vergebenen IDs den einzelnen Biomaterialien zugeordnet werden. Ein Vorteil des Ansatzes besteht allerdings darin, dass dadurch eine doppelte Codierung über Pseudonyme zwischen Patienten- und Probenidentitäten implementiert wird. Eine Umsetzung der Doppelcodierung wird zunehmend von Ethikkommissionen bei der Genehmigung von Forschungsvorhaben verlangt.

- Vergleich mit dem GANI_MED Projekt [Schack2010]:

  Eine zentrale Komponente des GANI_MED Projekts ist ein Research Data Warehouse, in das medizinische Daten aus der Krankenversorgung übernommen werden. Im Gegensatz zu Vanderbilt, das einen Opt-Out-Ansatz verfolgt, ist die rechtliche Basis generell die Patienten- Einwilligung. Vor einer Übernahme klinischer Daten in das Forschungs-Warehouse erfolgt eine Pseudonymisierung. Wie üblich werden identifizierende Daten und zugewiesene Pseudonyme in einer separaten Treuhänderstelle verwaltet.

In GANI_MED werden vom KIS versandte HL7 Nachrichten mitprotokolliert, um einen separaten Master Patient Index (MPI) aufzubauen. Dies entspricht der weiter oben beschriebenen bzw. diskutierten Option „Zugriff auf den HL7-Strom“. Die notwendigen Maßnahmen und die evtl. resultierenden Einschränkungen wurden oben beschrieben. Deswegen wird in den Münchener Projekten die oben beschriebene Option 1 präferiert, die einen Direktzugriff auf das führende klinische System und evtl. Folgezugriffe auf weitere klinische Systeme vorsieht.

- Vergleich mit dem i2b2 System aus Havard [Murphy2010]:

  In das i2b2 Warehouse werden medizinische Daten aus der Krankenversorgung repliziert. Diese Daten werden de-identifiziert und in forschungsfragestellungsspezifische Datensammlungen übernommen. Die Datensammlungen sind in i2b2 gespeichert. Patientenidentitäten werden in einer Identitätsmanagementkomponente im i2b2-System verwaltet. In i2b2 ist im Gegensatz zum hier vorgestellten Ansatz keine zweistufige Pseudonymisierung zwischen Patientenidentität und zugeordneten medizinischen Daten vorgesehen. So lassen sich personenbeziehbare Daten aus der Identitätsmanagementkomponente und medizinische Daten aus den Forschungsdatensammlungen bzw. den replizierten Daten aus der Krankenversorgung über den internen i2b2 Identifikator miteinander verknüpfen. Eine Trennung der i2b2 Datenbanken um diese verschiedenen organisatorischen Einheiten unterstellen zu können wird nicht explizit unterstützt.

- Vergleich mit dem De-identified Information Warehouse (DIW) Ansatz des Ohio State University Medical Centers [Erdal2012]:

  Das DIW enthält de-identifizierte Daten die aus dem Warehouse, welches in der Krankenversorgung eingesetzt wird, übernommen werden. Ein zentraler Unterschied zwischen dem in der Arbeit vorgestellten Konzept und dem von Erdal et al. besteht darin, dass im DIW nicht vorgesehen ist einen Patienten in den Suchergebnissen des DIWs identifizieren zu können und um zusätzliche forschungsfragestellungsspezifische Daten zu ergänzen. Eine Re-Identifikation eines Patienten auf Basis der im Suchergebnis zu einem Patienten enthaltenen Identifikatoren wird bei Erdal et al. versucht

auszuschließen. Die Zusammenführung unterschiedlicher Identitäten eines Patienten erfolgt nicht im DIW. Eine Zusammenführung von Patientenidentitäten hat im Quellsystem zu erfolgen. Die dann führende Patientenidentität kann mit den zusammengeführten Daten in das DIW übernommen werden. Wie dann mit der im DIW angelegten nicht mehr führenden Patientenidentität und den damit assoziierten Fallidentitäten und Daten verfahren wird, ist nicht beschrieben. Erdal et al. setzten für eine Übernahme von Daten zu einem Patienten u. Fall aus unterschiedlichen Quellsystemen, die im DIW einer Patienten- und Fallidentität zugeordnet sein sollen, voraus, dass die unterschiedlichen Patientenidentitäten in den Quellsystemen über den gleichen Identifikator verfügen. Das in dieser Arbeit vorgestellte Konzept sieht vor, dass Patienten in den Quellsystemen sowohl auf Basis der führenden Identifikatoren aus der Krankenversorgung als auch auf Grundlage ihrer demografischen Daten identifiziert und ihre Identitäten und Daten gezielt ins Forschungssystem übernommen werden können. Dadurch ist es möglich Patienten zu identifizieren und deren Daten im Forschungssystem zusammenzuführen, die nicht über denselben Identifikator in den Quellsystemen verfügen. Dies ist insbesondere dann von Nutzen wenn Daten zu einem Patienten über Institutionsgrenzen hinweg erfasst werden sollen um damit bspw. gesamte Krankheitsverläufe zu charakterisieren.

- Vergleich mit dem „Honest Broker“ Ansatz der Michigan Clinical Research Collaboratory [Boyd2009]:

  Der „Honest Broker“ dient bei Boyd et al. dazu Nachrichten zwischen IT-Systeme aus der Krankenversorgung und Forschungssystemen zu routen. IDs von Patienten aus den unterschiedlichen Systemen können im „Honest Broker“ miteinander verknüpft werden. Ein signifikanter Unterschied zwischen den Ansätzen besteht darin, dass der „Honest Broker“ passiv auf Nachrichten wartet und diese dann weiterleitet. So kann nicht aktiv über den „Honest Broker“ mit einem Direktzugriff nach Patientenidentitäten bzw. Fallidentitäten in den angeschlossenen Systemen gesucht werden, was die Auswertung von Suchergebnisse durch den Anwender und die gezielte Übernahme in Forschungssysteme verhindert.

- Vergleich mit weiteren Ansätzen:

Das IHE PIX-Profil beschreibt einen Ansatz für den Aufbau eines MPIs, beim dem ebenfalls Informationen zu Patienten aus HL7 Nachrichten extrahiert werden. Hierfür empfängt der PIX „Patient Identifier Cross-reference Manager“ u.a. HL7 Nachrichten, die vom KIS bei der Aufnahme und der Aktualisierung demographischer Daten sowie bei der Zusammenlegung von Patienten versandt werden. Der darüber aufgebaute MPI ordnet jedem Patienten eine eindeutige ID zu. Diese kann dann in den IT-Systemen zur Unterstützung der Krankenversorgung für den jeweiligen Patienten übernommen werden. Die Auswertung von HL7 Nachrichten, die um Attribute erweitert sind, mit denen Informationen zum Vorliegen von Einwilligungserklärungen propagiert werden, ist im IHE PIX-Profil nicht vorgesehen. MPIs auf Basis des IHE PIX-Profils, berücksichtigen damit die oben beschriebenen Aspekte des Datenschutzes im Rahmen der Forschung unzureichend. Eine Erfassung von Fall- und Einsendungs-IDs, anhand derer Patienten im MPI identifiziert werden könnten, ist im PIX-Profil ebenfalls nicht vorgesehen.

IBM stellt mit dem „Enterprise Master Person Index“ (EMPI) [Initiate2010] ein Werkzeug für das Management von Patientenidentitäten bereit. Der EMPI speichert demographische Daten der Patienten. Für die Zusammenführung verschiedener Identitäten eines Patienten wird im EMPI ein probabilistisches Record Linkage Verfahren eingesetzt. Die Menge an Attributen, die im EMPI gespeichert werden, kann zwischen verschiedenen Einsatzszenarien variieren. Mit der minimalen Anzahl an Attributen lassen sich Patienten im EMPI identifizieren und es kann festgestellt werden wo weitere elektronische Akten des Patienten existiert. Anwendung findet der EMPI im eHealth Bereich. Dort soll er im Rahmen der Behandlung den Austausch von elektronischen Patientenakten unterstützen [Shelley2010]. Dies weicht vom hier vorgestellten Ansatz stark ab: die patientenbezogene prospektive Erfassung von Daten und Bioproben hat im Rahmen dieser Arbeit einen hohen eigenen Stellenwert: die Datenübernahme dient dem Aufbau eines abgesicherten separaten Identitätsmanagements und ergänzt ansonsten lediglich die Datenerfassung. Zudem liegt bei der Datenübernahme ein besonderer Fokus auf der Berücksichtigung vorhandener Einwilligungserklärungen. Ein direkter Zugriff auf die Systeme für die Datenextraktion wird dabei bevorzugt. Zugriffe auf die HL7-Ströme wie sie

auch der EMPI unterstützt, sind im hier vorgestellten Ansatz optional vorgesehen.

In Peyton und Hus Framework [Peyton2010] für ein föderiertes Identitäts-management ist vorgesehen, mit Hilfe eines MPIs de-identifizierte medizinische Daten aus unterschiedlichen Repositories zu einem Patienten zusammenzuführen. Demographische Daten und Pseudonyme zu einem Patienten werden bei Peyton und Hu im MPI-System abgelegt. In einem Repository ist dann den medizinischen Daten eines Patienten ein Pseudonym zugeordnet. Ein Angreifer müsste zwei Systeme kompromittieren (MPI u. Repository), um die identifizierenden und medizinischen Daten unerlaubt zusammenführen zu können. Im Gegensatz dazu ist in dieser Arbeit eine doppelte Pseudonymisierung vorgesehen. Der Angreifer müsste dann Zugriff auf drei Systeme haben (MPI, Pseudonymisierungsdienst, Repository), um die Zuordnungen zwischen Patienten und ihre medizinischen Daten herstellen zu können. Des Weiteren unterscheiden sich die beiden Ansätze darin, in welchem Prozessschritt mit welchen Berechtigungen auf die Daten des Patienten zugegriffen wird. In dieser Arbeit ist vorgesehen nur dann (demographische) Daten aus der Krankenversorgung in die Forschung zu übernehmen, wenn entsprechende Einwilligungserklärungen der Patienten vorliegen. Bei MPI-basierten Ansätzen werden demographische Daten der Patienten zentral gespeichert. Für den Aufbau eines MPIs ist es häufig notwendig bereits bei der Aufnahme eines Patienten, also beim Anlegen der Identität eines Patienten in den IT-Systemen zur Unterstützung der Krankenversorgung, die Daten abzugreifen und in den MPI zu übernehmen (s. GANI_MED Projekt). Dadurch kommt es vor, dass für Forschungszwecke zu einem Zeitpunkt auf Patientendaten zugegriffen wird, an dem nicht zwangsläufig bereits eine unterschriebene Einwilligungserklärung des betroffenen Patienten vorliegt. Auch beim Anlegen von Patientenidentitäten im Forschungssystem bzw. im MPI-System lassen sich Unterschiede identifizieren. Peyton und Hu schlagen ein „Record Linkage“ Verfahren vor, um festzustellen, ob für einen Patienten bereits eine Identität im MPI vorhanden ist. Existiert eine Identität so wird diese weiterverwendet. Die Entscheidung darüber erfolgt vollautomatisiert. Das vollautomatisierte Zusammenführen von Identitäten eines Patienten mit Hilfe von „Record Linkage“- Verfahren ist unpräzise. Im hier beschriebenen Konzept ist vorgesehen, dass der Anwender die entsprechende Patientenidentität

zunächst im KIS sucht. Erst wenn er diese dort gefunden hat, erfolgt das Anlegen einer Identität im Forschungssystem. Falls mehrere Identitäten im KIS gefunden werden entscheidet der Anwender, welche die richtige ist. Dafür muss er ggf. neben den demographischen Daten auch medizinische Daten im KIS einsehen. Anhand der vom KIS vergebenen und ins Forschungssystem mitübernommenen ID kann präzise festgestellt werden, ob eine Identität des Patienten dort bereits vorhanden ist.

Eine Alternative für das Zusammenführen verschiedener Identitäten eines Patienten beschreiben Au und Coll [Au2008]. Den Patienten werden von einer vertrauenswürdigen dritten Stelle IDs zugewiesen. Die IDs werden in Devices gespeichert, die den Patienten gehören. Jede elektronische Krankenakte eines Patienten in unterschiedlichen Einrichtungen erhält eine solche ID. Der Patient entscheidet, ob seine Krankenakten zusammengeführt werden können. Dazu veranlasst er die vertrauenswürdige Stelle ein Zertifikat auszustellen, mit dem die Verknüpfung zweier IDs bestätigt wird. Damit können Patientenakten über Einrichtungsgrenzen hinweg mit geringem Aufwand zusammengeführt werden. Au und Coll berücksichtigen nicht das Management von Identitäten eines Patienten innerhalb einer Institution. Im Gegensatz zu dem hier vorgestellten Ansatz wird nicht auf eine praxistaugliche Umsetzbarkeit fokussiert. Unberücksichtigt bleibt insbesondere die benötigte Infrastruktur einschließlich der Devices sowie die zu etablierenden Prozesse für das Management der Identifikatoren und der Bestätigung von deren Zusammenlegung mit Zertifikaten einer vertrauenswürdigen dritten Stelle.

In Konzept von Deng et al. [Deng2009] wird vorausgesetzt, dass einem Patient eine globale ID (z.B. Sozialversicherungsnummer) zugeordnet ist, die auch in der Krankenversorgung verwendet werden darf. Dies ist nicht zwangsläufig gegeben. Sollte eine globale ID für die Krankenversorgung vorhanden sein, so können rechtliche Rahmenbedingungen die Verwendung des Identifikators darauf beschränken. Dies erschwert eine Übertragung der Vorgehensweise um Identitäten von Patienten zum Zwecke der Forschung miteinander zu verknüpfen. Die lokale ID für einen Patienten ergibt sich bei Deng et al. durch die symmetrische Verschlüsselung der globalen ID die zuvor um ein Präfix erweitert wird. Das Präfix ist das Ergebnis der Anwendung einer Hashfunktion auf einrichtungsspezifische Informationen. Daher unterscheiden sich lokale IDs eines Patienten in verschiedenen

Einrichtungen. Bei der Umsetzung dieses Konzepts ergibt sich gegenüber dem in der Arbeit beschriebenen Ansatz der Nachteil, dass die erzeugte lokale Patienten-ID der elektronischen Akte zugeordnet werden muss. Üblicherweise verfügen elektronische Patientenakten bereits über eigene Identifikatoren die von den EMR-Systemen verwaltet werden. Entweder muss dann das EMR-System erweitert werden bzw. es ist ein zusätzlicher Abbildungsschritt zwischen der lokalen Patienten-ID und der bereits einer elektronischen Patientenakte zugeordneten „EMR-ID“ notwendig.

5.1.2 Weiterverwendung von klinischen Daten in der medizinischen Forschung

Der hier vorgestellte Ansatz, Formulare, die für Forschungsfragestellungen konzipiert worden sind, mit Versorgungsdaten vorzubefüllen, weist eine Reihe von Vorteilen auf:

- Die Forschungssysteme, die als Zielsysteme für die Datenübernahme verwendet werden, besitzen i.a. ausgereifte Funktionalitäten zur Erstellung und Anpassung von elektronischen Formularen (EDC-Systeme), die genutzt werden können. Dies trägt zur Flexibilität des Konzepts bei. I.a. lassen sich solche Forschungssysteme rasch aufbauen und einfach adaptieren.
- Durch die Weiterverwendung von Formularmetadaten aus den Quellsystemen und der Ableitung von Transformationsvorschriften für die Datenübernahme werden Mechanismen bereitgestellt, welche das Aufsetzen von forschungsbezogenen Datensammlungen weiter beschleunigen können. Die Transformationsvorschriften selbst, können wiederum weiterverwendet werden. So könnten Bibliotheken für diese Transformationsvorschriften aufgebaut werden.
- Im vorgestellten Ansatz müssen Abbildungsregeln nur für Attribute definiert werden, die in einer spezifischen Forschungsdatensammlung benötigt werden. Im Gegensatz dazu müssen bei institutionsweiten Warehouselösungen bereits zu Beginn zahlreiche Abbildungsregeln für Attribute festgelegt werden, die aus Routinesystemen übernommen werden sollen.
- Im vorgestellten Ansatz können identifizierende und medizinische Daten in unterschiedlichen Systemen gespeichert werden. Die gezielte Trennung der Daten erfolgt während des Übernahmeprozesses. Damit können

Datenschutz- und Pseudonymisierungsanforderungen leichter umgesetzt werden.

- Bei der Extraktion der Daten mit einem Direktzugriff (s. Option1) kann überprüft werden, ob eine Einwilligungserklärung des Patienten vorliegt. Damit kann sichergestellt werden, dass nur Daten extrahiert werden, für deren Verwendung in der Forschung eine Zustimmung des Patienten erfolgt ist.
- Durch die differenzierte Vergabe von Zugriffsberechtigungen in den Forschungssystemen kann der Zugriff auf die gesammelten Forschungsdaten im Rahmen der Erfassung und des Managements der Daten auf einen definierten Anwenderkreis beschränkt werden. Solange die Daten nicht aus den Systemen exportiert werden, können Aspekte, die das geistige Eigentum der Forscher an den erfassten Daten betreffen, berücksichtigt werden. In dem sichergestellt wird, dass nur bestimmte Personen Berechtigungen für den Export der erfassten Daten erhalten, können Prozesse für die Auswertung und Nutzung der Daten gezielt unterstützt werden. Damit können Aspekte des Geistigen Eigentums sowie des Datenschutzes auch im Rahmen der Nutzung der Forschungsdaten sichergestellt werden.
- Die Einschränkung des Nutzerkreises erleichtert auch die Kontrolle der in Einwilligungserklärungen aufgeführten Verwendungszwecke der Forschungsdaten im Rahmen der Erfassung und des Managements der Daten in den Forschungssystemen. Dies kann auch den Erhalt von Datenschutzfreigaben für einrichtungsübergreifende Forschungs-datensammlungen vereinfachen.

Der vorgestellte Ansatz weist gewisse Einschränkungen auf:

- In unterschiedlichen Forschungsdatensammlungen können dieselben Daten aus der Behandlung benötigt werden. In diesem Fall muss eine mehrmalige Replikation der Daten erfolgen.
- Die Anzahl an Kommunikationsverbindungen kann mit der Zunahme der Anzahl an Forschungsdatensammlungen ansteigen. Wesentliche Probleme sind hierdurch nicht zu erwarten.
- Die Konsistenzsicherung auf Typ- und Instanzebene erfolgt nicht automatisiert. Änderungen an Formularen in den klinischen IT-Systemen müssen in den Forschungssystemen nachgezogen werden, ebenso auch Abbildungsregeln. Werden Änderungen an übernommenen Daten im

Forschungssystem durchgeführt, so werden diese nicht im Behandlungssystem nachgeführt. Im Rahmen der betrachteten Projekte war der Bedarf gering, da eine Fokussierung auf Kerndaten aus dem klinischen Bereich erfolgt ist (Demographische Daten, IDs, Falldaten, Diagnosen, Maßnahmen). Typ- oder Instanzänderungen sind hierbei selten.

Positiv ist die sehr gute Eignung des beschriebenen Ansatzes für ein eigenes Identitätsmanagement im Forschungskontext sowie für das Management, Tracking und die Doppelpseudonymisierung von Bioproben für die Forschung. Beides wird durch einen Warehouseansatz nicht abgedeckt.

Im Vergleich zu anderen Projekten, in denen Daten aus der Krankenversorgung in die klinische Forschung übernommen werden, ergeben sich folgende Aspekte:

Im „SD-Ansatz“ von Vanderbilt ist keine Re-identifikation eines Patienten aus den Forschungsdaten vorgesehen. Aus dieser bewussten Einschränkung ergeben sich sehr gute Datenschutzcharakteristika, allerdings kann nicht mehr aus dem Forschungssystem heraus bestimmt werden, für welche Patienten des KIS Daten übernommen wurden bzw. ob Änderungen/Ergänzungen an übernommenen Daten im KIS nach dem Zeitpunkt des Transfers stattgefunden haben. Für die Aktualisierung der Daten im Forschungssystem müssen Lösungen bereitgestellt werden. Eine Möglichkeit besteht im erneuten Transfer der Daten aller Patienten. Dies ist in der Praxis allerdings sehr aufwendig, da z.B. im BioVU-Projekt Daten von über 1,9 Mio. Patienten jedes Mal erneut übernommen werden müssten. Eine alternative Vorgehensweise setzt voraus, dass Änderungen an Daten über die Zeit im KIS/KAS mitprotokolliert und zum Zeitpunkt des erneuten Datentransfers nur die aktuellen Daten übernommen werden bzw. alle Daten zu den Patientenidentitäten an denen eine einzelne Änderung durchgeführt wurde. Dazu ist es notwendig, dass das KIS entsprechende Protokollierungsmechanismen mit dazugehörigen Exportfunktionalitäten zur Verfügung stellt. In der Praxis kann dies bei proprietären KIS Systemen sehr aufwendig zu implementieren sein. Der in dieser Arbeit vorgestellte Ansatz benötigt dagegen lediglich Funktionen für den Export von Patientendaten. Solche (Teil-) Funktionalitäten, bspw. für Entlassbriefe, (Labor-) Befunde, abrechnungsrelevante Daten, Tumorregisterdaten etc. werden häufig bereits von den Quellsystemen bereitgestellt. Hieraus ergibt sich gegenüber dem SD-Ansatz eine deutliche Vereinfachung. Bei Änderungen auf Typebene, z.B. die Hinzunahme neuer Formularfelder in den Routinesystemen und der

korrespondierenden Felder in den Forschungssystemen, sind im „SD-Ansatz“ und dem hier vorgestellten Ansatz ähnliche Anpassungen im Transferprozess durchzuführen.

Die einfache Anpassbarkeit von Forschungssystemen an sich ändernde Anforderungen für spezifische Forschungsfragestellungen wird im vorgestellten Konzept explizit berücksichtigt. Dies erleichtert das Aufsetzen und Anpassen von Forschungssystemen zur prospektiven Datenerfassung. Im Gegensatz dazu fokussieren Warehouse-Ansätze (s. das Beispiel GANI_MED) darauf, bereits in separaten Systemen erfasste Daten zu replizieren und Anwendern definierte Sichten darauf zu ermöglichen. Die Erweiterung um neue Parameter, wie sie im Forschungskontext notwendig wird, ist vergleichsweise aufwendig.

Auch in i2b2 werden Daten aus der Krankenversorgung möglichst umfänglich für die Forschung bereitgestellt [Murphy2010]. Im Gegensatz dazu sieht der hier vorgestellte Ansatz vor, Daten forschungsfragestellungsspezifisch durch Forscher zu erfassen und gezielt um Daten aus der Krankenversorgung zu ergänzen. Sowohl der Umfang als auch die Granularität der erfassten Daten gehen dabei häufig über das in der Krankenversorgung übliche Maß hinaus. Daher ist es Teil des hier skizzierten Konzepts nur hochstrukturierte Daten aus der Krankenversorgung selektiv zu übernehmen. Ob im i2b2-System bei der Replikation der Daten aus der Krankenversorgung das Vorliegen unterschriebener Einverständniserklärungen der Patienten berücksichtigt wird, ist nicht beschrieben.

Der Schwerpunkt des „De-identified Information Warehouses“ liegt auf der Bereitstellung von Daten aus der Krankenversorgung für Forscher und dabei sicherzustellen, dass eine Re-identifikation eines Patienten auf Basis von Suchergebnissen ausgeschlossen ist. Daher wird auch auf eine Übernahme von Freitexten aus der Krankenversorgung in das DIW verzichtet. Beim DIW ergeben sich einerseits sehr gute Datenschutzcharakteristika, anderseits führt das zu ähnlichen Problemen bzgl. der Aktualisierung der Daten wie für den „Synthetic Derivative“-Ansatz beschrieben.

Bei Ansätzen, in denen die zwischen IT-Systemen der Krankenversorgung und der Forschung ausgetauschten Nachrichten ausgewertet, manipuliert und weitergeleitet werden (s. „Honest Broker“ des Michigan Centers of Biological Information, GANI_MED), sind die zur Verfügung stehenden Daten auf die Inhalte der Nachrichten beschränkt. Daraus ergibt sich in der Praxis der Nachteil, dass Daten die für die Forschung relevant sind nicht zwangsläufig zwischen den Systemen ausgetauscht werden müssen. Dies ist insbesondere dann der Fall, wenn einzelne zentrale

klinische Informationssysteme ein breites Spektrum an Funktionalitäten anbieten und diese genutzt werden. So kann ein einzelnes klinisches Informationssystem z.B. Funktionen bereitstellen mit denen Daten auf den Krankenstationen, dem OP, in der Apotheke, der Radiologie und Kardiologie verwaltet werden. Bei Einsatz eines solchen holistischen Systems sinken die Anzahl an IT-Systemen in der Krankenversorgung und damit auch die zwischen den Systemen ausgetauschten Nachrichten. Ein weiterer Nachteil der nachrichtenbasierten Ansätze besteht darin, dass der Grad an Strukturiertheit der Daten, die zwischen den Systemen ausgetauscht werden, reduziert sein kann. So kann z.B. ein Befund der im Ursprungssystem (semi)-strukturiert erfasst wurde in der Nachricht in Freitextform übertragen werden.

Die CDISC „eSource Data Interchange“ (eSDI)-Arbeitsgruppe beschreibt Szenarien zur Weiterverwendung von Daten aus der Krankenversorgung in klinischen Studien unter Berücksichtigung der Anforderungen aus der FDA 21 CFR Part 11. Das hier vorgestellte Konzept kann für die Umsetzung des eSDI „Extraction and Investigator Verification“-Szenarios herangezogen werden. Im „Extraction and Investigator Verification“-Szenario erfolgt die Erfassung von Daten die in klinischen Studien verwendet werden partiell im EHR-System. Im Übernahmeprozess werden diese Daten aus dem EHR-System extrahiert. Ein Forscher verifiziert die extrahierten Daten und prüft, ob diese mit den Daten im EHR-System übereinstimmen. Danach werden sie im EDC-System gespeichert. Das Vorbefüllen von Formularen in Forschungssystemen ist zentraler Bestandteil des hier vorgestellten Ansatzes. Wird als Forschungssystem ein FDA 21 CFR Part 11 konformes CDMS eingesetzt, kann das Szenario umgesetzt werden. Die Validierung der zu übernehmenden Daten erfolgt im CDMS selbst. Hierzu unterstützen CDMS Systeme einen zweistufigen Übernahmeprozess. In der ersten Stufe werden die zu transferierenden Daten in einen „Puffer“ zwischengespeichert. In der zweiten Stufe erfolgt vor der endgültigen Übernahme die Validierung der im Puffer befindlichen Daten durch den Anwender.

Zusätzlich zur Übernahme medizinischer Daten aus den Routinesystemen wird in der hier vorgelegten Arbeit vorgeschlagen, auch Metadaten aus diesen Systemen weiterzuverwenden. Die Metadaten werden bei der Erzeugung neuer Formulare/eCRFs in den Forschungssystemen genutzt, in dem die Attribute (Name, Datentyp) und deren Ausprägungen sowie die korrespondierenden Feldbezeichnungen nicht erneut definiert werden müssen. Die extrahierten Metadaten werden in das CDISC-ODM-Format übernommen. Durch den Import der

CDISC-ODM-Spezifikation von eCRFs in das Forschungssystem wird die Erstellung der Formulare vereinfacht. Die Entwicklung von Funktionen zur Extraktion von Metadaten aus Komponentensystemen ist mit erheblichem Aufwand verbunden. Daher ist es eine naheliegende Vorgehensweise, existierende Ansätze weiterzuverwenden, um die den Komponentensystemen zugrunde liegenden Datenmodelle in einem ersten Schritt z.B. in RDF zu überführen. In einem zweiten Schritt erfolgt dann eine selektive Übernahme der Metadaten aus der XML-Repräsentation des RDF Modells in die CDISC-ODM-XML-Repräsentation. Eine Alternative zur Abbildung von relationalen Datenmodellen auf RDF besteht darin, die Datenmodelle auf ausdrucksmächtigere Beschreibungssprachen für Ontologien abzubilden und dann eine Überführung in CDISC-ODM vorzunehmen.

Liegt dem Ausgangssystem das relationale Modell zu Grunde so lassen sich in der Literatur verschiedene Ansätze für die (semi-) automatisierte Abbildung von relationalen Datenmodellen auf RDF finden. Beispiele hierfür sind D2RQ [D2RQ] und Virtuoso RDF Views [Blakeley2007], [OpenLink]. Eine Übersicht über Ansätze und Werkzeuge findet sich in [RDB2RDF]. Für die Erzeugung einer Ontologie wurde z.B. im Rahmen des DataGenie Projekts ein Plugin für Protégé entwickelt, mit dem aus einer relationalen Datenbank automatisiert eine Protégé-spezifische Ontologie geniert werden kann [DataGenie]. Hierbei wird jede Tabelle auf eine Klasse und jede Spalte auf eine Eigenschaft der entsprechenden Klasse abgebildet. Mit RelationalOWL steht eine spezielle OWL Ontologie für die Beschreibung von Schemainformationen aus relationalen Datenbanken zur Verfügung [Laborda2005].

Zwei domänenspezifische Aspekte erhöhen die Anforderungen an die Erstellung von (semi-) automatisierten Abbildungen auf RDF bzw. eine Ontologie: (a) Im klinischen Umfeld werden häufig Systeme vorgefunden, die generische Entity-Attribute-Value-Schemata verwenden, also auf die übliche explizite Modellierung verzichten. (b) Es gibt außerdem Informationssysteme, die spezielle Tabellentypen auf relationale Datenbanken aufsetzen und den Zugriff sowie die Verteilung von Daten auf diese Tabellen innerhalb der Applikation verwalten (z.B. SAP Pool- und Clustertabellen). In beiden Fällen wird die automatisierte Extraktion von Metadaten erschwert, wobei diese Fälle in der Literatur häufig nicht betrachtet werden.

In diesen Fällen könnten die Abbildungen zwischen den Datenmodellen der Ausgangssysteme auf die RDF-basierten Datenmodelle bzw. Ontologien nicht mehr automatisiert durchgeführt werden: Es müsste durch den Entwickler festgelegt werden, wie Entitäten, Beziehungen zwischen diesen Entitäten und Attribute im EAV-Datenmodel bzw. in Pool- und Clustertabellenstrukturen identifiziert und auf

RDF bzw. eine spezifische Ontologie abgebildet werden können. Dieser Ansatz wäre in der Praxis zeitintensiv, da eine hohe Anzahl von Entitäten und Attributen abzubilden sind. Zur Vereinfachung könnten Konzepte bzw. Werkzeuge entwickelt werden, mit denen die Abbildung von Metadaten aus EAV Schemata mit impliziter Struktur nach RDF erleichtert wird. Bspw. könnten Inferenzregeln bei der Erstellung der Anfragen und der Anfrageergebnisse nach Metadaten aus den EAV-Schemata eingesetzt werden.

Eine automatisierte Transformation der in der Domäne anzutreffenden Datenmodelle, Schemata und Formate auf RDF bzw. Ontologien wird bisher nur unzureichend mit Werkzeugen unterstützt. Deshalb wurde im Rahmen der hier vorgelegten Arbeit eine Extraktion und Überführung von Daten und Metadaten aus verschiedenen Ausgangssystemen nach CDISC-ODM untersucht. Dieses wird zwar von Zielsystemen in der Domäne implementiert, es existiert für die Datentransformation aber keine durchgängige Werkzeugunterstützung. In der Praxis ist es also oftmals ebenfalls zeitintensiv Daten aus den Ausgangssystemen in CDISC-ODM zu übernehmen. Aus diesem Grund wurde im Rahmen dieser Arbeit ein prototypisches Transformationswerkzeug entwickelt. Mit Hilfe dieses Werkzeugs konnte erfolgreich ein semi-automatischer Prozess für die Extraktion von (Meta)-Daten aus dem wesentlichen Quellsystem (i.s.h.med) implementiert werden. Die Ausgangslage bildet dabei eine XML-Repräsentation des gewünschten PMDs. Die Hauptaufgabe des Nutzers bei der Extraktion von (Meta)-Daten aus Formularen eines Dokumenttyps liegt darin, Felder für die Extraktion auswählen. Hierzu wird eine Baumansicht der im Formular enthaltenen Datenfelder zur Verfügung gestellt. Ausgehend von diesen Informationen ist das Werkzeug in der Lage, automatisiert entsprechende Metadaten zum Formular aus i.s.h.med abzurufen und eine Beschreibung dieser Informationen in ODM zu generieren. Im Hintergrund erzeugt das Werkzeug Transformationsregeln (als XSLT-Ausdrücke) die die spätere Transformation der eigentlichen Daten nach ODM stark vereinfacht. Für einen Praxiseinsatz muss das Werkzeug noch um zusätzliche Funktionalitäten erweitert werden. Bei einer hohen Anzahl an Feldern im Ausgangsdokument ist die aktuelle Baumdarstellung manchmal unübersichtlich. Hier wird vorgeschlagen eine Funktion zu implementieren, die es ermöglicht, einzelne Teilformulare auszublenden, um so die Übersichtlichkeit zu erhöhen. Weitere wichtige Funktionalitäten beinhalten das Rückgängig machen von bereits durchgeführten Arbeitsschritten und eine Versionierung der generierten CDISC-ODM Beschreibungen sowie Transformations-regeln.

Ein wichtiges Charakteristikum des hier vorgestellten Konzepts für die Weiterverwendung von klinischen Daten in der Forschung ist die Handhabung von Rollen und Berechtigungen: die Berechtigungsprüfungen finden bei der Extraktion von Daten aus den Routinesystemen statt. Dieser (unter 4.1.3 beschriebene) Vorgang ist komplex, ermöglicht aber danach die Verwaltung der Daten auf hohem Sicherheitsniveau. So kann sichergestellt werden, dass nur die Daten extrahiert werden für die Zugriffsberechtigungen des Anwenders vorliegen. Während des Direktzugriffs bei der Datenextraktion kann eine Abfrage auf das Vorliegen einer Einwilligungserklärung erfolgen. Dadurch werden nur die Daten derjenigen Patienten in Forschungssysteme übernommen, die einer Weiterverwendung zugestimmt haben. Für eine anschließende Nutzung können verschiedene Sicherheits-mechanismen (erneute Pseudonymisierung, k-Anonymisierung, pseudonymisierte Nutzung von Daten und Bioproben) bereitgestellt werden. Was diese Nutzung von Daten und Proben unter verschiedenen Sicherheitsaspekten betrifft, sind Warehousekonzepte häufig weniger flexibel als der hier vorgestellte Ansatz.

Die Berücksichtigung von Sicherheitsfragen als Merkmal des hier vorgestellten Ansatzes wurde bereits mehrfach angesprochen, etwa die getrennte Speicherung von identifizierenden Daten, medizinischen Daten und Probendaten in unterschiedlichen Datenbanken, die verschiedenen Organisationseinheiten unterstellt sind. Damit wird sichergestellt, dass eine Zuordnung von medizinischen und Probendaten aus den jeweiligen Subsystemen zu den identifizierenden Daten nur durch die Auflösung von Pseudonymen über den Pseudonymisierungsdienst erfolgen kann. In diesem Fall muss sich ein Angreifer unerlaubten Zugang auf mindestens zwei getrennte Subsysteme verschaffen, um medizinische und identifizierende Daten gemeinsam und zugeordnet sehen zu können. Befindet sich der Pseudonymisierungsdienst auf einem weiteren separaten System, müssten sogar drei Systeme kompromittiert werden. In dem bei der Zuordnung der Subsysteme die drei Verantwortlichen so gewählt werden, dass mindestens ein nicht am Forschungsvorhaben beteiligter Partner involviert ist, kann das Risiko reduziert werden, dass Forscher sich zusammenschließen und alle in den Subsystemen vorhandenen Daten ohne Einsatz der Anwendung und den damit durchgeführten Berechtigungsprüfungen zusammenführen.

Sollte ein Angreifer Zugriff auf die medizinischen bzw. die Probendaten erhalten, hängt das Risiko für eine Re-identifikation eines Patienten von den erfassten Daten ab. Das Risiko steigt bei Daten, welche sowohl in den medizinischen Daten als auch

anderswo erfasst und öffentlich zugänglich sind. Bei Probendaten hängt das Risiko für eine Re-identifikation ebenfalls von den erfassten Daten ab. So haben z.B. Lin et al. gezeigt, dass weniger als 100 SNPs bereits genügen, um ein einzelnes Individuum innerhalb der Weltbevölkerung eindeutig identifizieren zu können, vorausgesetzt es liegen entsprechende Vergleichsproben vor [Lin2004]. Aus Genom-, bzw. Exomdaten kann direkt auf physische Attribute wie z.B. das Geschlecht, die Blutgruppe, die Farbe der Haut und Erbkrankheiten des Patienten geschlossen werden [Lowrance2007].

Der Aufwand für einen Angreifer, um die Daten aus den getrennten Datenbanken zusammenführen zu können, ist groß. Ein Schwachpunkt des beschriebenen Ansatzes stellen die Client-Rechner der Anwender dar auf denen die Zusammenführung der Daten unter Wahrung von Berechtigungen erfolgt. Kompromittiert der Angreifer einen dieser Rechner und gelangt an die Zugangsdaten des Anwenders, so kann er an dessen Stelle auf die Daten zugreifen. Verhindern lässt sich dies durch den Einsatz von Einmalpasswörtern. Hierzu müssen die Anwender mit entsprechenden Kennwortgeneratoren ausgestattet werden.

5.1.3 Integration von Anwendungen auf Präsentationsebene

Zentraler Bestandteil der oben beschriebenen Sicherheitsarchitektur ist die clientseitige Zusammenführung von Daten, die in getrennten Teilsystemen gespeichert werden (z.B. zur Verwaltung von identifizierenden-, medizinischen- und Probendaten). Dazu kann es notwendig sein die Benutzerschnittstellen der Teilsysteme auf Präsentationsebene im Client auf dem Computer des Anwenders integrieren zu können.

Für die Integration von Benutzerschnittstellen, die auf Web-Technologie basieren, existieren neben den im Kapitel 4.3 beschriebenen Ansatz für GUI-Wrapper weitere Ansätze. Diese weisen verschiedene Vor- und Nachteile gegenüber dem im Rahmen der Arbeit beschriebenen Ansatz auf. Ein Ansatz besteht im Einsatz von html-Frames bzw. eingebetteten html Frames (Inlineframes (Iframes)). Dabei würde der Typ1-GUI-Wrapper einen html Frame bzw. eine html-Seite mit einem Iframe bereitstellen. In diesen würde die auf Web-Technologie basierende, zu integrierende Benutzerschnittstelle gerendert werden. Die Verwendung eines neuen Frames bzw. Iframes führt dazu, dass die darzustellende html-Seite geändert wird. Diese setzt sich dann aus der vom zu integrierenden System bereitgestellten html-Seite plus dem Frame bzw. Iframe zusammen. Dabei wird der ursprünglichen html-Seite durch

den GUI-Wrapper eine weitere Ebene in die zu rendernde „Gesamt-html-Seite“ eingefügt (im „Document Object Model“ (DOM)-Knotenbaum der Seite kommt ein weiterer Knoten hinzu). Dies hat Auswirkungen auf den dynamischen Inhalt der „Gesamt-html-Seite“. Wird beim Ausführen des dynamischen Inhalts (JavaScript) auf Knoten des DOM-Knotenbaumes zugegriffen und werden diese Knoten durch die Angabe des absoluten Pfades vom Wurzelknoten aus adressiert, so kann dies zu Fehlern führen. Dies liegt daran, dass auf Knoten zugegriffen wird, auf die bei der Ausführung des dynamischen Inhalts in der originalen html-Seite nicht oder in anderem Kontext zugegriffen wird. Systeme deren Benutzerschnittstellen auf Web-Technologie basieren und Frames bzw. IFrames verwenden, können folglich nur dann integriert werden, falls die darzustellenden Seiten über keinen dynamischen Inhalt verfügen oder bei denen die Adressierung der Knoten im DOM-Knotenbaum nicht über absolute Pfadangaben erfolgt. Durch die Verwendung von Frames bzw. Iframes ist es mit den beschriebenen Einschränkungen möglich, die Benutzerschnittstelle der zu integrierenden Systeme einzubinden. Ein Kontextwechsel, z.B. zu einem Patienten in der integrierten Anwendung, ist mit diesem Ansatz alleine nicht realisierbar. Für die Durchführung des Kontextwechsels könnte ein, auf JavaScript bzw. ActionScript basierendes, Skript eingesetzt werden, an das z.B. die ID des aufzurufenden Patienten übergeben wird. Dieses Skript würde die Navigation zu diesem Patienten innerhalb der in den Frames bzw. Iframes integrierten Anwendungen durchführen. Der Nachteil bei diesem Ansatz besteht darin, dass Sicherheitsmechanismen in Browsern dies erheblich erschweren. Um Angriffe basierend auf „Cross Site Scripting“ (XSS) zu verhindern, implementieren Browser das „Same Origin Policy“ (SOP) Sicherheitskonzept. Dabei erlauben Browser den Zugriff auf die Objekte einer Webseite nur dann mit JavaScript bzw. ActionScript, wenn der Skript-Code und die Webseite von derselben Quelle stammen. Dies hat zur Folge, dass bei einem auf Frames bzw. Iframes und JavaScript basierendem Ansatz für die Integration einer Anwendung Änderungen an der Software selbst vorgenommen werden müssen. Dazu werden auf dem Webserver, der die html-Seiten für die Benutzerschnittstelle der zu integrierenden Anwendung zur Verfügung stellt, eine entsprechende html-Seite und JavaScript hinterlegt. Die Navigation z.B. zu einem Patienten in der integrierten Anwendung erfolgt durch ausführen dieses JavaScripts. Hierbei stammen die neu hinzugefügte html-Seite, JavaScript-Code für die Navigation und die Webseiten der Benutzerschnittstelle der zu integrierenden Anwendung vom selben Server. Der Kontextwechsel wird dann dadurch initiiert, dass die zusätzlich auf dem Server

hinterlegte html-Seite vom Browser des Anwenders aufgerufen wird. Dabei werden in der URL die für den Kontextwechsel benötigten Parameter codiert. Diese werden vom auf der Seite eingebetteten JavaScript ausgelesen und der Kontextwechsel durchgeführt. Der Nachteil dieses Ansatzes besteht darin, dass Manipulationen am Webserver der zu integrierenden Anwendung durchgeführt werden müssen. Dadurch können sich Probleme ergeben. Handelt es sich bei der zu integrierenden Anwendung um Software, bei deren Installation entsprechende Qualifications durchgeführt wurden, müssen diese ggf. wiederholt werden. Hersteller können im Fehlerfall darauf verweisen, dass Manipulationen an der Software durchgeführt wurden.

Ein alternativer Ansatz zur Integration von auf Webtechnologie basierenden Benutzerschnittstellen besteht darin, einen GUI-Wrapper zu entwickeln, der eine Engine zum Rendern von html, CSS und JavaScript selbst implementiert bzw. Bibliotheken verwendet, die eine solche Engine realisieren. Der Vorteil dieses Ansatzes besteht in der hohen Flexibilität. Sicherheitsmechanismen wie XSS könnten für die zu integrierende Anwendung nicht berücksichtigt werden. Nachteile ergeben sich bzgl. der Größe des Wrappers. Sollte der Wrapper bei jedem Kontextwechsel vom Server heruntergeladen und ausgeführt werden müssen wären längere Wartezeiten nicht auszuschließen. Häufig setzen die zu integrierenden Anwendungen einen auf dem Markt etablierten Browser voraus. Selbst entwickelte Browserfunktionalitäten im Wrapper müssten sich für einen solchen Browser ausgeben. Die rasche Weiterentwicklung im Bereich der Webtechnologie und deren zunehmende Komplexität erschweren diesen Ansatz zusätzlich. Die Installation des Wrappers auf den Computern der Anwender stellt eine andere Möglichkeit für das Deployment des Wrappers dar. Dazu muss der Anwender ggf. über Administratorberechtigungen auf dem lokalen Arbeitsplatzrechner verfügen, was üblicherweise nicht der Fall ist.

Die Integration von Webanwendungen auf Präsentationsebene mit automatisiertem Kontextwechsel, bei dem lokal auf dem Computer des Anwenders Komponenten installiert sind, kann auch mit Hilfe von Skripten zur GUI-Automatisierung erfolgen (bspw. AutoIt [AutoIt]). Um lokal auf einem Computer installierte Komponenten aus dem Browser heraus aufrufen zu können, müssen bestimmte Voraussetzungen erfüllt sein. So erlaubt dies z.B. der Microsoft Internet Explorer. Allerdings sind die Sicherheitsvorkehrungen im Browser herabzusetzen. Es muss die Ausführung von ActionScript erlaubt werden. Außerdem müssen vertrauenswürdige Bereiche definiert werden (Webseiten von denen heruntergeladene Skripte ausgeführt werden

dürfen). Beim Einsatz von GUI-Skripten zur Durchführung eines Kontextwechsels in Browsern ergibt sich das Problem, dass die auf Webseiten dargestellten GUI-Elemente (Felder, Listen, Button, usw.) nicht über das GUI-Skript adressiert werden können. Dies bedeutet, es stehen keine Handler auf diese Elemente zur Verfügung. Der Fokus auf diese Elemente aus dem Skript heraus kann durch die Simulation von Tastatureingaben oder durch Positionierung und Auswahl mit der Maus erlangt werden. Dies ist sehr fehleranfällig. Absolute Positionsangaben für Mausbewegungen schlagen bei unterschiedlichen Auflösungen (Bildschirm, Browseransicht vergrößert/verkleinert) fehl. Tastatureingaben haben den Nachteil, dass der Anwender während der Ausführung des Skripts den Fokus auf eine andere Anwendung setzen kann. Dadurch werden die vom Skript simulierten Tastatureingaben in die vom Betriebssystem verwaltete Ereignisschlange für diese Anwendung eingereiht und der Fokus erreicht nicht das GUI-Element im Browser. Weitere simulierte Eingaben durch das Skript werden an das falsche GUI-Element geleitet. Die Eingabe schlägt fehl. Durch das Blockieren sämtlicher Benutzereingaben während der Ausführung des GUI-Skripts kann eine Fokusänderung durch den Anwender verhindert werden. Dies hat zur Folge, dass umfangreiche Maßnahmen bei der Ausführung des GUI-Skripts ergriffen werden müssen um ein Unterbrechen oder Abstürzen des Skripts zu verhindern.

Der im Rahmen dieser Arbeit entwickelte und umgesetzte Ansatz zur kontextbezogenen Präsentationsintegration von Anwendungen, deren Benutzerschnittstelle auf Webtechnologie basiert, weist diese skizzierten Nachteile nicht auf. Die Webseiten der zu integrierenden Anwendungen werden nicht mit Hilfe von Frames/Iframes in die „Gesamt-html-Seite“ eingebettet. Die Einbettung erfolgt dadurch, dass auf der „Gesamt-html-Seite“ ein Bereich vorgesehen ist, in dem der Wrapper seinen darzustellenden Inhalt rendert. Dies geschieht z.B. durch die Einbettung eines Applets in die „Gesamt-html-Seite“. Der DOM-Baum der „Gesamt-html-Seite“ und der html-Seiten der zu integrierenden Anwendung bleiben unverändert. Die zu integrierende Anwendung bekommt auf der Gesamtseite ein eigenes „Window“ zugewiesen. Dadurch adressieren absolute Pfadangaben für Zugriffe auf Knoten des DOM-Baumes durch JavaScript aus den zu integrierenden html-Seiten, die richtigen Knoten.

Bei der Integration einer Anwendung, basierend auf dem in dieser Arbeit vorgestellten Ansatz, müssen keine Eingriffe an der zu integrierenden Anwendung vorgenommen werden. Damit entfallen Maßnahmen bezüglich einer zusätzlichen Installation-Qualification, des zu integrierenden Systems. Ebenso gibt es

Unterschiede in den Dateigrößen der Wrapper. Wrapper, die auf dem in der Arbeit vorgeschlagene Ansatz beruhen, nutzen die lokal auf dem Rechner des Anwenders installierten Renderengines der Browser. Damit sind sie wesentlich kleiner als ein Wrapper, der eine eigene Renderengine implementiert. Dies hat geringere Downloadzeiten zur Folge. Außerdem ergeben sich keine Probleme bzgl. der Anforderungen der zu integrierenden Software an die sich auf dem Markt befindlichen Browser.

Gegenüber dem Einsatz von lokal auf den Rechnern installierten GUI-Skripten ergibt sich der Vorteil, dass bei der Installation und beim Aufruf des Wrappers auf bewährte Ansätze zurückgegriffen werden kann. Das Deployment des Wrappers kann auf den Download und den Start eines Applets bzw. einer Webstartanwendung reduziert werden. Im Gegensatz dazu müssen GUI-Skripte separat heruntergeladen werden. Zusätzlich setzt deren Aufruf aus Browsern heraus reduzierte Sicherheitseinstellungen in den Browsern voraus. Weitere Vorteile des Ansatzes gegenüber GUI-Skripten ergeben sich bzgl. der Einbettung in die Benutzeroberfläche der Anwendung, aus der heraus der Kontextwechsel initiiert wird. Beim Einsatz von GUI-Skripten öffnet sich ein neues Browserfenster. Die Einbettung einer Anwendung mit dem Wrapper erfolgt im selben Fenster. Durch die programmatische Anbindung der Renderengine des lokal installierten Browsers, über die Ereignisse abgefragt und JavaScript „eingeschleust“ werden können, gestaltet sich der Kontextwechsel im Vergleich zur Verwendung von GUI-Skripten, robuster und weniger fehleranfällig.

Ein Nachteil des erarbeiteten Ansatzes ergibt sich bei der Erstellung der Beschreibung für den Zustandsautomaten. Es müssen z.B. die Felder der Webformulare in der Benutzeroberfläche der zu integrierenden Anwendungen identifiziert werden, in die Werte eingetragen werden sollen. Dazu ist es ggf. notwendig, den html-Quellcode der Seite des Zielsystems zu analysieren. Muss JavaScript „eingeschleust“ werden, um eine bestimmte Benutzerinteraktion mit der zu integrierenden Anwendung durchzuführen, ist dieses ggf. zu erstellen.

Für die kontextbezogene Präsentationsintegration von Anwendungen, die ihre Benutzeroberflächen mit lokal auf den Computern der Anwender installierten Fat-Clients visualisieren, stehen wenige Ansätze zur Verfügung. Ein Ansatz besteht in der Verwendung der oben beschriebenen externen GUI-Skripte. Dabei sind die Probleme bzgl. des Deployments, Starts und der Parameterübergabe an die Skripte dieselben. Bei der Adressierung der GUI-Elemente in den zu integrierenden Fat-Client-Benutzerschnittstellen ergeben sich Unterschiede. Setzt sich die

Benutzeroberfläche des Fat-Clients aus heavyweight Komponenten zusammen, so lassen sich diese einfach identifizieren und adressieren. Bei heavyweight-Komponenten werden im Gegensatz zu lightweight-Komponenten vom Betriebssystem bereitgestellte Methoden benutzt, um die grafische Oberfläche zu zeichnen. Dies hat bei Windows-Betriebssystemen zur Folge, dass Windowhandler für die einzelnen GUI-Elemente und Bezeichner für Toplevel GUI-Elemente der Benutzeroberfläche vom Betriebssystem verwaltet werden. Über die Bezeichner kann von den GUI-Skripten aus auf die einzelnen GUI-Elemente zugegriffen und Benutzerinteraktionen simuliert werden. Werden lightweight Komponenten zur Realisierung der Benutzerschnittstelle verwendet, ergeben sich bzgl. der Adressierung und des Zugriffs auf die GUI-Elemente ähnliche Probleme, wie sie bei der kontextbezogenen Integration von webtechnologiebasierten Benutzeroberflächen und GUI-Skripten zu finden sind.

Verglichen mit GUI-Skripting-Lösungen beschränkt sich das Deployment und der Start des vorgestellten Wrappers zur Integration von Fat-Clients auf das Herunterladen von Applets bzw. einer Webstart-Anwendung. Die Integration des Fat-Clients in die Oberfläche der aufrufenden Anwendung kann im gleichen Fenster erfolgen. Ein separates Fenster, wie es bei Einsatz von GUI-Skripten erscheinen würde, entfällt. Der Fat-Client wird mit dem Wrapper ein „Teil“ der Benutzerschnittstelle der aufrufenden Anwendung. Im Falle von heavyweight-basierten Komponenten des Fat-Clients gewährleistet der beschriebene Wrapper eine sichere Adressierung der GUI-Elemente, in dem er direkt auf die vom Betriebssystem verwalteten Windowhandler zugreift. Damit kann im Gegensatz zu GUI-Skripten bei denen die Identifikation eines GUI-Elements anhand des Namens erfolgt gewährleistet werden, dass ein Ereignis für das gewünschte GUI-Element ausgelöst wird. Bei lightweight-Komponenten ergeben sich auch bei dem vorgestellten Ansatz Probleme bzgl. der Identifikation des richtigen GUI-Elements. Zur Lösung dieses Problems sind weitere Ansätze zu untersuchen. Beispielsweise könnten Ansätze zur Erkennung von GUI-Elementen basierend auf ihrer Form und Beschriftung eingesetzt werden [Sikuli]. Ein weiterer Nachteil des vorgestellten Ansatzes besteht darin, dass er Betriebssystem abhängig ist. So sind die vom Betriebssystem bereitgestellten Methoden für den Zugriff auf einzelne Windowhandler betriebssystemspezifisch. Die Machbarkeit wurde bislang nur an Microsoft Windows XP und Microsoft Windows 7 nachgewiesen.

Der vorgestellte Ansatz, Anwendungen auf Präsentationsebene zu integrieren, kann von verwandten Konzepten abgegrenzt werden. Bei Verwendung des CCOW-Standards [CCOW] ergibt sich der Vorteil, kontextbezogene Wechsel zwischen verschiedenen Systemen durchführen zu können. Abweichend vom vorgestellten Ansatz kann aus jeder über den CCOW Kontext-Manager integrierten Anwendung der Kontextwechsel initiiert werden: Es ist keine „Master-Anwendung“ notwendig, aus der heraus der Wechsel initiiert wird. In dem eine Anwendung den CCOW-Standard implementiert, stellt sie explizit Schnittstellen für die Steuerung der Anwendung nach außen bereit. Damit wird die Möglichkeit, Kontextwechsel von einer Anwendung durchzuführen, Teil der Funktionalität der Anwendung. Die notwendigen Abläufe innerhalb der Anwendung werden in den vom Betriebssystem verwalteten Prozessen für diese Anwendung ausgeführt. Im Gegensatz dazu werden für den vorgestellten GUI-Wrapper separate Prozesse vom Betriebssystem initiiert. Über diese wird der Kontextwechsel in der zu integrierenden Anwendung gesteuert. Zugriffe auf die zu steuernde Anwendung erfolgen über die vom Betriebssystem bereitgestellten Schnittstellen, um auf Prozesse einer Anwendung von fremden Prozessen aus zuzugreifen. Dies hat zur Folge, dass die Zugriffsmöglichkeiten u.a. aus Sicherheitsgründen eingeschränkt sind. Ein erheblicher Nachteil des CCOW-Standards besteht in der Schwierigkeit diesen umzusetzen. So unterstützen die im Klinikum befindlichen (Legacy-)Anwendungen den Standard nicht, und die Implementierung des Standards gestaltet sich aufwendig. Um (Legacy-) Anwendungen im Nachhinein CCOW-fähig zu machen ist ggf. die Zusammenarbeit mit Softwareherstellern notwendig [Berger2009]. Die Verwendung des CCOW-Standards für die kontextbezogene Integration von in der Routine eingesetzten Anwendungen ist in der Praxis nicht praktikabel.

Andere Ansätze zur Integration von Anwendungen auf Präsentationsebene zielen primär darauf ab, Funktionalitäten der Anwendungen über Webservice-Schnittstellen zur Verfügung zu stellen [Canfora2008], [Zhang2008]. Die Eingaben in diese Anwendungen erfolgen über die Benutzerschnittstellen der Anwendungen. Die Ausgaben der Anwendungen werden aus den Bildschirmmasken ausgelesen bzw. es wird über die Benutzerschnittstellen veranlasst, Daten lokal zu speichern. Der Zugriff für den Aufruf von Funktionen in den Anwendungen erfolgt nicht wie üblich über APIs der Anwendungen. Daten für die Eingabe und Ausgabe werden über Webservice-Schnittstellen ausgetauscht.

Der Ansatz von Canfora et al. zum Integrieren der Benutzerschnittstellen unterscheidet sich von dem hier vorgestellten Ansatz darin, dass im Wrapper die

Positionen (absolut und relativ) der Felder und Labels auf den einzelnen Bildschirmmasken hinterlegt sind. Die Eingabe von Daten in die Felder erfolgt durch die Simulation von Tastatureingaben. Die Auswahl der Felder für die Eingabe ebenso. Der im Rahmen der Arbeit entwickelte Ansatz identifiziert die einzelnen Felder auf der Benutzeroberfläche sowie das Hauptfenster der Benutzeroberfläche anhand des vom Betriebssystem vergebenen Prozess-Identifikators und den damit „assoziierten“ Window-Handlern. Dadurch ergeben sich Vorteile gegenüber dem Ansatz von Canfora et al. Die Identifikation der Eingabefelder in den Bildschirmmasken ist sicher möglich, da Angaben von Positionen der Felder in Bildschirmmasken von den Bildschirmauflösungen abhängig sind. Canfora et al. weisen die Umsetzung ihres Ansatzes nur an einem Textterminal nach, welches als Benutzerschnittstelle für eine Anwendung dient. In diesem Textterminal erfolgt die Positionsbestimmung von Labels anhand von Spalten und Zeilenangaben. Für die Eingabe von Daten in ein Feld ist es bei dem in der Arbeit vorgestellten Ansatz nicht notwendig den Eingabefokus auf dieses Feld durch die Simulation von Tastatureingaben zu lenken. Im vorgestellten Ansatz werden Eingaben durch die Angabe des entsprechenden Window-Handlers und mit Hilfe des Aufrufs einer Betriebssystemfunktion an ein Feld übergeben. Im Ansatz von Zheng et al. beruht die Navigation in der zu integrierenden Anwendung auf GUI-Skripten. Daraus ergeben sich die oben genannten Vor- und Nachteile.

Eine weitere Alternative liegt darin, GUI-Elemente zu kopieren und Benutzerinteraktionen in den zu integrierenden Anwendungen aufzuzeichnen [Marhaim2010]. Dies weist gegenüber dem vorgestellten Ansatz bei der Integration von Fat-Client-Anwendungen Nachteile auf. So verfügen die in der Krankenversorgung eingesetzten Informationssysteme über hoch differenzierte Benutzerschnittstellen, deren Bildschirmmasken sich aus sehr vielen GUI-Elementen zusammensetzen. Darüber hinaus sind die Navigationsmöglichkeiten innerhalb der Ausgangsanwendung vielfältig. Das Kopieren dieser GUI-Elemente und ihr Anordnen innerhalb neuer Bildschirmmasken wäre sehr aufwendig. Erschwerend kommt hinzu, dass die eingesetzten GUI-Elemente von den Systemen selbst bereitgestellte Widgets umfassen (z.B. SAPGui), was für die kopierten GUI-Elemente zu einem geänderten „Look and Feel“ führen würde. Ein weiterer Nachteil besteht darin, dass davon ausgegangen wird, die in den kopierten GUI-Elementen auftretenden Ereignisse, an die Originalelemente weiterleiten zu können. In der Routine befinden sich überwiegend Microsoft Windows Betriebssysteme (XP, Vista, 7) im Einsatz. Sicherheitskonzepte in diesen Betriebssystemen sehen vor, nur berechtigten

Prozessen Zugriffe auf die Ereignisschlangen anderer Prozesse zu gewähren, um diese manipulieren zu können. Betriebssystemprozesse, die für die Ausführung verschiedener Anwendungen (z.B. Anwendung mit Original-GUI, Anwendung mit kopierter GUI) laufen, können dies nicht. Das Aushebeln dieser Sicherheitskonzepte, z.B. in dem Interrupthandler oder Betriebssystemtreiber manipuliert werden (s. Keylogger usw.), ist für den Praxiseinsatz eine ungeeignete Vorgehensweise. Die Machbarkeit des von Marhaim et al. vorgestellten Ansatzes wurde nur prototypisch an Webanwendungen getestet, in dem GUI-Elemente dieser Anwendungen kopiert und eine Kommunikation mit den Originalelementen durchgeführt wurden. Für die kontextbezogene Integration von Webanwendungen ist der Aufwand der benötigt wird, um diese durch das Kopieren von GUI-Elementen nachzubauen, ist größer als beim Ansatz dieser Arbeit, bei dem „lediglich“ die GUI-Elemente identifiziert werden müssen.

Ein weiterer Ansatz sieht vor, die Integration von Informationssystemen durch die Weiterverwendung von GUI-Komponenten aus existierenden Anwendungen zu ermöglichen [Yu2007]. Dazu sollen Adapter für die verschiedenen Typen von GUI-Komponenten bereitgestellt werden. Diese Adapter kommunizieren über eine Middleware. Dazu propagieren sie eintretende Ereignisse und Operationen an andere Adapter. Für die Integration von Fat-Client-Anwendungen ergeben sich gegenüber dem vorgestellten Ansatz ähnliche Probleme, wie bei dem Ansatz, der auf kopierten GUI-Elementen beruht. Die Machbarkeit des Ansatzes wird anhand einer Anwendung demonstriert, die sich aus Komponenten anderer Anwendungen zusammensetzt. Die einzelnen Komponenten basieren alle auf Webtechnologie (Java Applet, Ajax, ActiveX). Mit grafischen Komponenten aus Fat-Clients wird die Machbarkeit nicht nachgewiesen. Der von Yu et al. vorgestellte Ansatz unterscheidet sich zu dem vorgestellten Ansatz darin, dass einzelne Komponenten aus einer GUI und nicht die gesamte GUI integriert werden. Dies setzt voraus, dass sich die zu integrierende GUI aus Komponenten zusammensetzt, die sich separieren lassen. Der Schwerpunkt des Ansatzes liegt auf der statischen Zusammensetzung „neuer“ GUIs. Wie dynamische Aspekte (z.B. die Navigation für den Aufruf eines Applets), die für eine kontextbezogene Integration Voraussetzung sind, unterstützt werden, wird nicht ausgeführt.

Der vorgestellte Ansatz zur Integration von Webanwendungen unterscheidet sich zu Mashup-Anwendungen insbesondere darin, dass diese darauf fokussieren, einzelne Komponenten existierender Webanwendungen zu „neuen“ Anwendungen zusammenzusetzen und nicht gesamte Anwendungen zu integrieren. Dabei wird bei

Mashup-Anwendungen u.a. auf APIs zugegriffen, die von den zu integrierenden Anwendungen dafür bereitgestellt werden. Die Visualisierung geschieht dann in der Mashup-Anwendung selbst. Werden grafische Elemente existierender Webanwendungen in die neue Mashup-Anwendung integriert, kann dies unter Zuhilfenahme von Frames/Iframes erfolgen. Dabei ergeben sich oben beschriebene Probleme bzgl. der Adressierung einzelner Elemente mit JavaScript in der zu integrierenden Anwendung. Werden die Webseiten der „neuen“ Mashup-Anwendung auf einem Mashup-Server aus den einzelnen Seiten der originalen Anwendungen zusammengesetzt und auf dem Browser des Endanwenders visualisiert, ergeben sich Probleme bzgl. der Same Origin Policy, wenn vom Browser Daten aus der Originalanwendung nachgeladen werden. Um diese zu Umgehen werden bei Mashup-Anwendungen u.a. Ansätze gewählt, die Lücken in den SOP-Implementierungen der Browser ausnutzen, wie z.B. „Dynamic Script Tagging“. Dazu unterstützen die zu integrierenden Anwendungen z.B. JSONP.

6 Ausblick

Im Rahmen dieser Arbeit wurde ein Konzept für das Management verschiedener Identitäten von Patienten, Fällen und Proben in einer verteilten Umgebung erarbeitet. Es umfasst die Unterstützung der verteilten Suche nach Patienten in Systemen zur Krankenversorgung und Forschung sowie die Übernahme demographischer und weiterer identifizierender Daten aus verschiedenen, insbesondere klinischen Systemen. Unter Berücksichtigung datenschutzrechtlicher Vorgaben können Identitätsinformationen zusammengeführt und persistent gemacht werden. Hierdurch wurden neue Möglichkeiten für die Integration von Krankenversorgungs- und Forschungsdaten erarbeitet, wobei ein besonderer Schwerpunkt auf ein eigenes, abgesichertes Identitätsmanagement von Forschungsdaten gelegt wurde, das gleichzeitig auch Identitätsinformationen aus klinischen Systemen nutzbar macht. Daneben wurde ein Konzept umgesetzt, auf dessen Grundlage Systeme auf Präsentationsebene integriert werden können, wodurch sich Kontextwechsel initiieren lassen, die weitere Informationen in den Ausgangssystemen zugänglich machen.

Glossar

ADT	Arbeitsgemeinschaft Deutscher Tumorzentren	Dachverband der Tumorzentren in Deutschland.
BAPI	Business Application Programming Interface	Standardisierte Programmierschnittstelle über die auf SAP Business Objekte zugegriffen werden kann.
BDSG	Bundesdatenschutzgesetz	Regelt den Datenschutz im öffentlichen Bereich und in der Privatwirtschaft. Legt u.a. Voraussetzungen fest, nach denen eine Verarbeitung personenbezogener Daten zulässig ist. [Leiner2006]
CTSA	Clinical and Translational Science Award	Förderprogramm des NIH um Erkenntnisse aus der (Grundlagen-)Forschung in praktische Lösungen für den Patienten überzuführen.
CCOW	Clinical Context Object Workgroup	Definiert Standards welche die visuelle Integration von Anwendungen im Gesundheitswesen unterstützen. [CCOW]
CDA	Clinical Document Architecture	XML-basierte Dokumentenarchitektur für die elektronische Dokumentation und Kommunikation von medizinischen Informationen. [Haas2004]
CDISC	Clinical Data Interchange Standards Consortium	Non-profit Organisation die Standards für die Erfassung, den Austausch und die Archivierung von klinischen Daten und Metadaten für die Forschung etabliert. [CDISC]
CDMS	Clinical Data Management System	IT-System zur elektronischen Erfassung und dem Management von Daten im Rahmen klinischer Studien. Sie werden auch im Rahmen von Arzneimittelstudien eingesetzt, die strengen Regularien unterliegen. Im Rahmen dieser Arbeit sind diese Systeme (als Untergruppe der EDC-Systeme) von nachgeordneter Bedeutung.

COM	Component Object Model	Ermöglicht die Kommunikation zwischen Softwarekomponenten in Microsoft Windows Betriebssystemen
CRF	Case Report Form	Erhebungsbogen in dem Daten zu einem Patienten im Rahmen von Studien dokumentiert werden
CTMS	Clinical Trial Management System	IT-System, welches die Planung, die Vorbereitung, die Durchführung und die Auswertung von Studien unterstützt.
DDE	Double Data Entry	Verfahren mit dem durch doppelte Eingabe derselben Daten eine höhere Datenqualität erreicht werden soll. Findet u.a. Anwendung bei der Datenerfassung in klinischen Studien.
DICOM	Digital Imaging and Communications in Medicine	Kommunikations- und Interoperabilitätsstandard für medizinische Bilddaten [Haas2004]
DIMDI	Deutsches Institut für Medizinische Dokumentation und Information	Herausgeber von Klassifikationen zur Kodierung von Diagnosen und Operationen in Deutschland [DIMDI]. Klassifikationen bilden Grundlage für die Abrechnung der erbrachten Leistungen von Krankenhäusern gegenüber Krankenkassen.
DOM	Document Object Model	Programmierschnittstelle mit der auf Elemente eines HTML bzw. XML Dokuments zugegriffen werden kann
DRG	Diagnosis Related Group	Klassifikationssystem zur aufwandbezogenen Zusammenfassung von Behandlungsfällen. Grundlage für die Vergütung von Behandlungsleistungen in zahlreichen Ländern und auch in Deutschland. [Haas2004], [DIMDI], [InEK], [GDRG]
EAV	Entity Attribute Value	Generisches Datenmodell ohne explizite Modellierung von Attributen

eCRF	Electronic Case Report Form	Elektronischer Erhebungsbogen für die Erfassung von Daten in Studien
EDC	Electronic Data Capture	Dienen der elektronischen Erfassung strukturierter patientenbezogener Daten im Forschungskontext. EDC-Funktionalitäten werden auch von CDMS Systemen bereitgestellt. Im Rahmen dieser Arbeit wird an verschiedenen Stellen „Forschungssystem“ als Überbegriff von EDC-Systemen verwendet.
EMR	Electronic Medical Record	Elektronische Akte eines Patienten in einer Institution
FDA	Food and Drug Administration	Lebensmittelüberwachungs- und Arzneimittelzulassungsbehörde der USA
FDA 21 CFR Part 11	FDA 21 Code of Federal Regulations Part 11	Umfasst Regularien der FDA für elektronische Akten und elektronische Signaturen.
GCP	Good Clinical Practice	Internationaler, nach ethischen und wissenschaftlichen Gesichtspunkten aufgestellter Standard für den Entwurf, die Durchführung, die Protokollierung und die Auswertung von klinischen Studien, an denen Versuchspersonen teilnehmen. [GCP]. GCP-Konformität ist mit hohem Aufwand verbunden. Sie wird bei Arzneimittelstudien, nicht dagegen bei Registern gefordert.
HAPI	HL7 application programming interface	Stellt Programmierschnittstellen für die Verarbeitung von HL7-Nachrichten mit Java zur Verfügung
HL7	Health Level 7	ANSI akkreditierte Organisation welche Standards für IT-Systeme im Gesundheitswesen entwickelt. Der HL7 Messaging Standard wird für den Datenaustausch zwischen Systemen zur Unterstützung der Krankenversorgung eingesetzt

ICD	International Classification of Diseases	Eine von der WHO herausgegebene einachsige monohierachische Klassifikation für Diagnosen [Leiner2006]. Sie wird in Deutschland als ICD-10-GM zur Diagnosendokumentation und damit Grundlage der Diagnosis Related Groups verwendet [s.h. DIMDI]
ICPM	International Classification of Procedures in Medicine	Eine von der WHO herausgegebene Klassifikation für Prozeduren und Maßnahmen [Leiner2006], [ICPM]
IHE	Integrating the Healthcare Enterprise	Initiative von Anwendern und Hersteller um den Informationsaustausch zwischen IT-Systemen im Gesundheitswesen zu verbessern. [IHE]
IQ	Installation Qualification	Mit dieser wird sichergestellt, dass die Installation von Software und Hardware entsprechend der Spezifikation erfolgt ist
IS-H	Industry Solution Healthcare	Krankenhausspezifische Erweiterung des SAP R/3-Systems zur Unterstützung des Patientenmanagements und der Abrechnung. [IS-H]
JAX-WS	Java API for XML Web Services	Programmierschnittstelle für die Erstellung von SOAP basierten Webservices.
JCo	Java Connector	Middleware Komponente welche die Kommunikation mit SAP-Servern ermöglicht.
JNI	Java Native Interface	Ermöglicht die Integration von Java Code mit C und C++ Code.
JSON	JavaScript Object Notation	Vom Menschen lesbares programmiersprachenunabhängiges Datenaustauschformat.

KAS	Klinisches Arbeitsplatzsystem	System welches von Ärzten und Pflegekräften auf Stationen im Rahmen ihrer Patientenversorgungstätigkeiten eingesetzt wird. Es dient auch als zentrales Repository für klinische Daten.
KIS	Krankenhaus-informationssystem	Zentrales IT-System im Krankenhaus. Stellt u.a. Funktionalitäten für das Management von demografischen Daten der Patienten, Falldaten (insbes. Datums- und organisatorische Angaben zu stationären und ambulanten Aufenthalten) sowie abrechnungsrelevanten Daten (nach ICD und OPS codierte Diagnosen und Maßnahmen sowie DRGs) zur Verfügung. Diese Funktionalitäten werden auch als Patientendatenmanagement bezeichnet. Das KIS ist eine Obermenge von Patientendatenmanagementsystem und KAS.
LIMS	Laboratory Information Management System	Die Funktionalität ist dem LIS grundsätzlich ähnlich, allerdings wird der Begriff „LIMS“ i.a. im Forschungskontext, „LIS“ im Krankenversorgungskontext eingesetzt. Aufgrund der in der Arbeit beschriebenen Trennung zwischen Forschungs- und Versorgungsumgebung, handelt es sich technisch um unterschiedliche Systeme. Typischerweise verarbeitet ein LIMS im Ggs. zum LIS auch Daten aus Genomanalysen.
LIS	Laboratory Information System	System im Krankenversorgungsbereich, dient der Ergebnisdokumentation und Unterstützung der Abläufe im Labor. Dies umfasst das Probenmanagement, die Messwertauswertung sowie die Ergebniszusammenfassung. I.a. werden die LIS-Daten in das KAS repliziert.
LSDG	Landesdaten-schutzgesetz(e)	Pendants zum BDSG auf Länderebene. Erweitern und verschärfen teilweise die Vorschriften des BDSG. [Leiner2006]

MPI	Master Patient Index	Verwaltet für jeden Patienten einen eindeutigen Identifikator. Verschiedene Identitäten eines Patienten in verschiedenen IT-Systemen referenzieren diesen Identifikator. Dient der einrichtungsübergreifenden Identifikation; in Deutschland nicht flächendeckend vorhanden.
NCI	National Cancer Institute	Teil des NIH. Koordiniert die nationalen Krebsprogramme in den USA.
NIH	National Institutes of Health	Eine dem U.S. Department of Health and Human Services unterstellte Behörde die für medizinische Forschung zuständig ist.
ODM	Operational Data Model	CDISC Format für den Austausch von Daten und Metadaten für die klinische Forschung [ODM].
OPS	Operationen- und Prozedurenschlüssel	Wird verwendet um medizinische Operationen und Prozeduren zu verschlüsseln; neben der ICD eine Hauptgrundlage der DRGs in Deutschland [DIMDI]
PACS	Picture Archiving and Communication System	System zur Speicherung und Kommunikation umfangreicher Bilddaten mit speziellen Speicher- und Zugriffsstrategien [Haas2004]
PDV	Patientendaten-verwaltung	Synonym zu Patientendatenmanagement. Bildet die Kernfunktionalität eines KIS. Umfasst Funktionalitäten für die Verwaltung demographischer Daten, für die Dokumentation von Fällen sowie für die Abrechnung erbrachter Leistungen.
PIX	Patient Identifier Cross Referencing	IHE-Profil für die Verknüpfung von Patientenidentifikatoren aus unterschiedlichen Domänen. [PIX]

PMD	Parametrisierbare medizinische Dokumente	Framework zur Entwicklung von Eingabemasken in i.s.h.med.
RD	Research Derivative	De-identifizierte elektronische Patientenakte deren ID sich aus Anwendung einer Hashfunktion auf die EMR-ID errechnet. Mapping zwischen EMR-ID und RD-ID wird persistiert.
RDF	Resource Description Framework	Standard für den Austausch von Daten im Web. [RDF]
Register		„Medizinisches Dokumentationssystem mit einer Aufgabenstellung, vor allem im Bereich der medizinisch-wissenschaftlichen Forschung“ [Leiner2006]. Im Rahmen dieser Arbeit dienen die betrachteten Forschungssysteme bzw. EDC-Systeme i.a. dem Aufbau von Registern. Register werden zur Generierung von Hypothesen (und auch von klinischen Studien) verwendet.
RFC	Remote Function Call	Entfernte Funktionsaufrufe zwischen SAP-Systemen sowie zwischen Nicht-SAP Systemen und SAP-Systemen.
RFD	Retrieve Form for Data Capture	IHE-Profil das beschreibt wie Daten in einer Anwendung des Benutzers für die Weiterverwendung in einer anderen Anwendung erfasst werden können.
RIS	Radiologie-informationssystem	System welches die Abläufe in der Radiologie unterstützt. Dazu zählen insbesondere die Verwaltung der zu untersuchenden Patienten sowie die Befunderstellung und –verwaltung.
SCXML	State Chart extensible Markup Language	Sprache zur Beschreibung eines ereignisbasierten Zustandsautomaten.

SD	Synthetic Derivative	De-identifizierte elektronische Patientenakte deren ID sich aus Anwendung einer Hashfunktion auf die EMR-ID errechnet. Wurde im Rahmen des Vanderbilt BioVU Projekts entwickelt
SNP	Single Nucleotide Polymorphism	„Ein SNP ist eine Variation an einer einzelnen Position in einer DNA-Sequenz zwischen Individuen. Wenn mehr als 1% einer Population nicht das gleiche Nukleotid an einer spezifischen Position in der DNA-Sequenz tragen, dann kann diese Variation als ein SNP klassifiziert werden.“ [Scitable]
SOP	Standard Operating Procedure	Standardisierte Arbeitsanleitungen für die Durchführung von Studien.
SOP	Same Origin Policy	Sicherheitskonzept welches den Zugriff aus einer Webseite auf Objekte anderer Websites einschränkt.
SWT	Standard Widget Toolkit	GUI-Framework für Java zur Erstellung von Benutzerschnittstellen.
TMF	Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.	Soll Forscher bei der Lösung von Problemen in der vernetzen medizinischen Forschung unterstützen.
WSDL	Web Services Description Language	XML-Format zur Beschreibung von Services in Form von Endpunkten an die Informationen als Nachrichten gesendet und empfangen werden können.
XDS	Cross-Enterprise Document Sharing	IHE-Profil in dem beschrieben wird wie das Management von Dokumente zu einem Patienten zwischen verschiedenen Institutionen erfolgen kann. [XDS]
XSS	Cross-site scripting	Ausnutzung einer Sicherheitslücke bei der Angreifer Code hinter einem Link, der als vertrauenswürdig erscheint, hinterlegen.

Literaturverzeichnis

[Aarts2004] Aarts J, Doorewaard H, Berg M. Understanding implementation: the case of a computerized physician order entry system in a large Dutch university medical center. J Am Med Inform Assoc. 2004 May-Jun; 11(3):207-16.

[ACM] Association for Computing Machinery. [homepage on the internet]. [cited 2012 Aug 6]. Available from: http://www.acm.org/

[ACM1] The Public Policy Committee of ACM. Understanding Identity and Identification. [article on the internet]. [cited 2009 May 25]. Available from: http://usacm.acm.org/usacm/Issues/identity.pdf

[AMG2005] Gesetz über den Verkehr mit Arzneimitteln (Arzneimittelgesetz - AMG). Arzneimittelgesetz in der Fassung der Bekanntmachung vom 12. Dezember 2005. Available from: http://www.gesetze-im-internet.de/bundesrecht/amg_1976/gesamt.pdf

[Au2008] Au R, Croll P. Consumer-Centric and Privacy-Preserving Identity Management for Distributed E-Health Systems. Proceedings of the 41st Annual Hawaii International Conference on System Sciences. 2008.

[AutoIt] AutoIt v3 – Automate and Script Windows Tasks [homepage on the internet]. Jonathan Bennett; 2009. [cited 2010 Jan 6]. Available from: http://www.autoitscript.com/autoit3/

[BAYKRG 2007] Bayerisches Krankenhausgesetz (BayKrG) in der Fassung der Bekanntmachung vom 28. März 2007. [cited 2010 Jan 6]. Available from: http://www.stmf.bayern.de/kommunaler_finanzausgleich/allgemeines/krankenhausfoerderung/Bayerisches%20Krankenhausgesetz.pdf

[BBMRI] BBMRI. Biobanking and Biomolecular Resources Research Infrastructure [homepage on the Internet]. BBMRI; [cited 2009 Jul 17]. Available from: http://www.bbmri.eu/

[BDSG1990] Bundesdatenschutzgesetz (BDSG). Available from: http://www.gesetze-im-internet.de/bundesrecht/bdsg_1990/gesamt.pdf

[Berger2009] Berger RG, Baba J. The realities of implementation of Clinical Context Object Workgroup (CCOW) standards for integration of vendor disparate clinical software in a large medical center. Int J Med Inform. 2009 Jun;78(6):386-90.

[Blakeley 2007] Blakeley C. RDF Views of SQL Data (Declarative SQL Schema to RDF Mapping). OpenLink Software, 2007.

[Bloomrosen 2008] Bloomrosen M, Detmer D. Advancing the Framework: Use of Health Data-A Report of a Working Conference of the American Medical Informatics Association. J Am Med Inform Assoc. 2008 Nov-Dec;15(6):715-22.

[Boyd2009] Boyd AD, Saxman PR, Hunscher DA, Smith KA, Morris TD, Kaston M, Bayoff F, Rogers B, Hayes P, Rajeev N, Kline-Rogers E, Eagle K, Clauw D, Greden JF, Green LA, Athey BD. The University of Michigan Honest Broker: a Web-based service for clinical and translational research and practice. J Am Med Inform Assoc. 2009 Nov-Dec;16(6):784-91.

[Canfora 2008] Canfora G, Fasolino AR, Frattolillo G, Tramontana P. A wrapping approach for migrating legacy system interactive functionalities to Service Oriented Architectures. Journal of Systems and Software. 2008;81(4):463-48.

[CCOW] The HL7 CCOW Standard [homepage on the internet]. HL7 Australia; [cited 2010 Jan 6]. Available from: http://www.hl7.org.au/CCOW.htm

[CDISC] CDISC [homepage on the internet]. CDISC; [cited 2011 Jan 6]. Available from: http://www.cdisc.org/

[Collins2011] Collins FS. Reengineering Translational Science: The Time Is Right. Sci Transl Med. 2011 Jul 6;3(90):90.

[COM4J] Com4j [homepage on the internet]. [cited 2010 March 8]. Available from: http://com4j.java.net/

[CTSA] Clinical and Translational Science Awards [homepage on the internet]. NIH; [cited 2010 Jan 6]. Available from: http://www.ctsaweb.org/

[DataGenie] Protégé DataGenie. Available from: http://protege.cim3.net/cgi-bin/wiki.pl?DataGenie

[Deloitte 2012] Deloitte. 2012 European Summit on Trustworthy Reuse of Helath Data. Current eHealth data landscape within the EU. IMIA Conference; Brussels; May 2012.

[Deng2008] Deng M, Scandariato R, Cock D de, Preneel B, Joosen W. Identity in federated electronic healthcare. In: Wireless Days, 2008. WD '08. 1st IFIP. Piscataway, N.J.: IEEE; 2008. p. 1–5.

[Deng2009] Deng M, De Cock D, Preneel B. Towards a cross-context identity management framework in e-health. Online Information Review. 2009;33(3):422-42.

[Denny2010] Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, Wang D, Masys DR, Roden DM, Crawford DC. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics. 2010 May 1;26(9):1205-10.

[DICOM] DICOM Homepage [homepage on the internet]. Medical Imaging & Technology Alliance; [cited 2010 Jan 6]. Available from: http://medical.nema.org/

[DIMDI] DIMDI. Deutsches Institut für Medizinische Dokumentation und Information [homepage on the internet]. [cited 2012 Aug 24]. Available from: http://www.dimdi.de/static/de/index.html

[DRG] InEK – Institut für das Entgeltsystem im Krankenhaus [homepage on the internet]. [cited 2011 July 6]. Available from: http://www.g-drg.de/cms/

[D2RQ] D2RQ Platform. [cited 2011 Jan 24]. Available from: http://www4.wiwiss.fu-berlin.de/bizer/D2RQ/spec/.

[Embi2009] Embi PJ, Payne PR. Clinical research informatics: challenges, opportunities and definition for an emerging domain. J Am Med Inform Assoc. 2009 May-Jun;16(3):316-27.

[Erdal2012] Erdal BS, Liu J, Ding J, Chen J, Marsh CB, Kamal J, Clymer BD. A database de-identification framework to enable direct queries on medical data for secondary use. Methods Inf Med. 2012;51(3):229-41.

[eSDI2005] Clinical Data Interchange Standards Consortium, Electronic Source Data Interchange (eSDI) Group. Leveraging the CDISC Standards to Facilitate the use of Electronic Source Data within Clinical Trials. Version 0.5, 16th September 2005 [serial on the internet]. 2005 Sep 16 [cited 2009 Jul 17]. Available from: http://www.ehealthinformation.ca/documents/eSDIv05.pdf

[Ethik2004] Nationaler Ethikrat: Biobanken in der Forschung - Stellungnahme. Saladruck, Berlin, 2004 [cited 2010 Aug 9] Available from: http://www.ethikrat.org/dateien/pdf/NER_Stellungnahme_Biobanken.pdf

[Ethik2010] Deutscher Ethikrat 2010. Humanbiobanken für die Forschung - Stellungnahme. Deutscher Ethikrat Berlin. Available from: http://www.ethikrat.org/dateien/pdf/stellungnahme-humanbiobanken-fuer-die-forschung.pdf

[GCP] ICH. Good Clinical Practice: Consolidated Guideline [homepage on the Internet]. ICH; [cited 2009 Jul 17]. Available from: http://www.ich.org/cache/compo/276-254-1.html

[GDRG] InEK – Institut für das Entgeltsystem im Krankenhaus [homepage on the internet]. [cited 2011 July 6]. Available from: http://www.g-drg.de/cms/

[Gulcher 2000] Gulcher JR, Kristjánsson K, Gudbjartsson H, Stefánsson K. Protection of privacy by third-party encryption in genetic research in Iceland. Eur J Hum Genet. 2000 Oct;8(10):739-42.

[Haas2004] Haas P. Medizinische Informationssysteme und Elektronische Krankenakten. 1st ed. Berlin Heidelberg: Springer; 2004.

[HAPI] HAPI. HL7 application programming interface. Available from: http://hl7api.sourceforge.net/

[Harris2009] Harris PA, Taylor R, Thilke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009; 42:377-81.

[Harrop2009] Harrop M. Identity Management - An overview of the status of some of the key IdM work. 4th ETSI (European Telecommunications Standards Institute) Security Workshop. France 2009. [cited 2012 Sep 17]. Available from: http://docbox.etsi.org/Workshop/2009/200901_SECURITYWORKSHOP/TheCottinghamGroup_Harrop_IdentityManagement.pdf

[Hibernate] Hibernate.org [homepage on the internet]. Red Hat Middleware; 2009. [cited 2010 Jan 6]. Available from: https://www.hibernate.org/

[HIPPA] Health Insurance Portability and Accountability Act. [homepage on the internet]. U.S. Department of Health & Human Services; [cited 2010 Jan 6]. Available from: http://www.hhs.gov/ocr/privacy/

[HITECH] Department of Health and Human Services. Health Information Technology for Economic and Clinical Health Act. [homepage on the internet]. [cited 2012 Nov 15]. Available from: http://healthit.hhs.gov/portal/server.pt/community/healthit_hhs_gov__home/1204

[HL7] HL7 Home [homepage on the internet]. HL7; [cited 2010 Jan 6]. Available from: http://www.hl7.org

[ICD] WHO International Classification of Diseases (ICD) [homepage on the internet]. WHO; [cited 2010 Jan 6]. Available from: http://www.who.int/classifications/icd/en/

[ICPM] WHO International Classification of Health Interventions (ICHI) [homepage on the internet]. WHO; [cited 2010 Jan 6]. Available from: http://www.who.int/classifications/ichi/en/

[IHE] IHE.net Home [homepage on the internet]. IHE International; [cited 2010 Jan 6]. Available from: http://www.ihe.net/

[InEK] InEK – Institut für das Entgeltsystem im Krankenhaus [homepage on the internet]. [cited 2011 July 6]. Available from: http://www.g-drg.de/cms/

[Initiate2010] IBM Initiate. EMPI: A Building Block For Interoperability. 2010. IBM AG; [cited 2012 Mai 15]. Available from: http://www-07.ibm.com/solutions/au/healthcare/presentations/downloads/hic_empi.pdf

[ISH] SAP AG. SAP Deutschland - SAP für das Gesundheitswesen – Nutzen [homepage on the internet]. SAP AG; [cited 2009 Jul 17]. Available from: http://www.sap.com/germany/industries/healthcare/businessbenefits/index.epx

[ishmed] Siemens Deutschland. i.s.h.med [homepage on the internet]. Siemens Deutschland; [cited 2009 Jul 17]. Available from http://www.medical.siemens.com/webapp/wcs/stores/servlet/CategoryDisplay~q_catalogId~e_-3~a_categoryId~e_1018818~a_catTree~e_100010,1008631,1010512,1018818~a_langId~e_-3~a_storeId~e_10001.htm

[ITU] ITU. International Telecommunication Union [homepage on the internet]. ITU; [cited 2012 Aug 6]. Available from: http://www.itu.int/en/Pages/default.aspx

[ITU1] ITU. International Telecommunication Union. ITU-T X.1252. Baseline identity management terms and definitions. [cited 2012 Mai 17]. Available from: http://www.itu.int/rec/T-REC-X.1252-201004-I

[Java-WS] Java Web Services At a Glance [homepage on the internet]. Sun Microsystems, Inc.; 2010. [cited 2010 Jan 6]. Available from: http://java.sun.com/webservices/

[JCo] SAP. Java Connector [homepage on the internet]. [cited 2011 Feb 6]. Available from: http://help.sap.com/saphelp_dm40/helpdata/de/6f/1bd5c6a85b11d6b28500508b5d5211/content.htm

[Jena] Jena - A Semantic Web Framework for Java [homepage on the internet]. [cited 2011 Jan 6]. Available from: http://jena.sourceforge.net/

[Jenia] Jenia.org [homepage on the internet]. Jenia.org; 2010. [cited 2010 Jan 6]. Available from: http://www.jenia.org/

[Josang2005] Jøsang A, Pope S. User Centric Identity Management. Asia Pacific Information Technology Security Conference (AusCERT); Australia; 2005 May 22-26.

[Josang2007] Jøsang A, AlZomai M, Suriadi S. Usability and privacy in identity management architectures. Proceedings of the fifth Australasian symposium on ACSW frontiers; 2007; Darlinghurst, Australia; p. 143-152.

[JSF] JavaServer Faces Technology [homepage on the internet]. Sun Microsystems, Inc.; 2010. [cited 2011 Feb 6]. Available from: http://java.sun.com/javaee/javaserverfaces/

[Karp2008] Karp DR, Carlin S, Cook-Deegan R, Ford DE, Geller G, Glass DN, Greely H, Guthridge J, Kahn J, Kaslow R, Kraft C, Macqueen K, Malin B, Scheuerman RH, Sugarman J. Ethical and practical issues associated with aggregating databases. PLoS Med. 2008 Sep 23;5(9):e190.

[Keller2006] Keller H, Krüger S. ABAP Objects: ABAP-Programmierung mit SAP NetWeaver. SAP PRESS; 2006.

[Killcoyne 2009] Killcoyne S, Boyle J. Managing Chaos: Lessons Learned Developing Software in the Life Sciences. Comput Sci Eng. 2009 Nov;11(6):20-29.

[Kohlmayer 2010] Kohlmayer F, Lautenschläger RR, Wurst SHR, Klopstock T, Prokisch H, Meitinger T, Eckert C, Kuhn KA. Konzept für ein deutschlandweites Krankheitsnetz am Beispiel von MitoREGISTER. GI Jahrestagung (2), 2010. pp. 746-51.

[Krafzig2005] Krafzig D, Banke K, Slama D. Enterprise SOA – Service-Oriented Architecture Best Practices. Indianapolis: Prentice Hall; 2005.

[Kuhn2001] Kuhn KA, Giuse DA. From hospital information systems to health information systems. Problems, challenges, perspectives. Methods Inf Med. 2001;40:275-87.

[Laborda 2005] de Laborda CP, Conrad S. Relational OWL A Data and Schema Representation Format Based on OWL. In Second Asia-Pacific Conference on Conceptual Modelling (APCCM2005), vol 43 of CRPIT, p.89 -96, Newcastle, Australia, 2005.

[Lamla2010] Lamla G, Blaser R, Prasser F, Wurst SHR, Rechl H, Gradinger R, Kuhn KA. Eine Umsetzung des IHE Single Source Konzepts für die translationale Forschung bei Knochen- und Weichteilsarkomen. GI-Edition: Lecture Notes in Informatics; GI Symposium 2010

[Langanke 2011] Langanke M, Brothers KB, Erdmann P, Weinert J, Krafczyk-Korth J, Dörr M, Hoffmann W, Kroemer HK, Assel H. Comparing different scientific approaches to personalized medicine: research ethics and privacy protection. Per Med. 2011 Jul;8(4):437-444.

[LDSG] Datenschutzgesetze. [cited 2011 Feb 6]. Available from: http://www.baden-wuerttemberg.datenschutz.de/recht/default.htm

[Leiner2006] Leiner F, Gaus W, Haux R. Medizinische Dokumentation. Schattauer 2006.

[Lenz2004] Lenz R, Kuhn KA. Towards a continuous evolution and adaptation of information systems in healthcare. Int J Med Inform. 2004;73(1):75-89.

[Lin2004] Lin Z, Owen AB, Altman RB. Genetics. Genomic research and human subject privacy. Science. 2004; 305(5681):183.

[Littman 2007] Littman BH, Di Mario L, Plebani M, Marincola FM. What's next in translational medicine? Clin Sci (Lond). 2007 Feb;112(4):217-27.

[LOINC] Logical Observation Identifiers Names and Codes (LOINC) [homepage on the internet]. Regenstrief Institute, Inc.; [cited 2010 Jan 6]. Available from: http://loinc.org/

[Lowrance 2007] Lowrance WW, Collins FS. Ethics. Identifiability in genomic research. Science. 2007; 317(5838):600-2.

[Lunshof 2008] Lunshof JE, Chadwick R, Vorhaus DB, Church GM. From genetic privacy to open consent. Nat Rev Genet. 2008 May;9(5):406-11.

[Macro2009] InferMed. InferMed - Clinical trial software, remote data entry, clinical decision support, clinical guidelines [homepage on the internet]. InferMed; [cited 2009 Jul 17]. Available from: http://www.infermed.com/index.php/macro/features.

[Mak2011] Mak HC. Trends in computational biology—2010. Nat Biotechnol. 2011 Jan;29(1):45-9.

[Malin2005] Malin BA. An evaluation of the current state of genomic data privacy protection technology and a roadmap for the future. J Am Med Inform Assoc. 2005 Jan-Feb;12(1):28-34.

[Malin2010] Malin B, Karp D, Scheuermann RH. Technical and policy approaches to balancing patient privacy and data sharing in clinical and translational research. J Investig Med. 2010; 58(1):11-8.

[Marheim 2010] Marheim I, Mordechai E, Bartolini C, Bergman R, Ariel O, Peltz C. A visual tool for rapid integration of enterprise software applications. Proceeding ICWE'10 Proceedings of the 10th international conference on Web engineering Springer-Verlag Berlin, Heidelberg 2010

[MedDRA] MedDRA MSSO Welcome [homepage on the internet]. MedDRA MSSO; [cited 2010 Jan 6]. Available from: http://www.meddramsso.com/

[Mendis2012] Mendis ME, Phillips LC, Kuttan R, Donahoe J, Churchill S, Kohane I, Murphy SN. The Architecture behind the Identity Management Cell within i2b2. 2012 AMIA Summit on Translational Bioinformatics; San Francisco, CA, USA. p.168.

[Mirhaji09] Mirhaji P, Zhu M, Vagnoni M, Bernstam EV, Zhang J, Smith JW. Ontology driven integration platform for clinical and translational research. BMC Bioinformatics. 2009 Feb 5;10 Suppl 2:S2.

[Murphy 2010] Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, Kohane I. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc. 2010 Mar-Apr;17(2):124-30.

[MyFaces] MyFaces [homepage on the internet]. Apache Software Foundation; 2010. [cited 2010 Jan 6]. Available from: http://myfaces.apache.org/

[m4] Munich biotech cluster m4 [homepage on the internet]. [cited 2011 Aug 05]. Available from: http://www.m4.de/

[m4Ethik] Kuhn KA, Wichmann HE, et al. Arbeitsgruppe Ethikkonzept m4. Ethik- und Datenschutzkonzept der Biobank Alliance und des Data Integration System des m4 Spitzenclusters München (m4 BA/DIS). 2012.

[NCI2007] National Cancer Institute Office of Biorepositories and Biospecimen Research. Biospecimen Basics: An Overview of the National Cancer Institute Best Practices for Biospecimen Resources. [serial on the internet]. 2007 Jun [cited 2009 Jul 17]. Available from: http://www.allirelandnci.org/pdf/NCI_Best_Practices_060507.pdf

[NIH2011] NIH. Translational Research – Overview [homepage on the internet]. [cited 2011 March 6]. Available from: http://commonfund.nih.gov/clinicalresearch/overview-translational.aspx

[ODM] CDISC. Specification for the Operational Data Model. [homepage on the internet]. CDISC; [cited 2011 Jan 17]. Available from: http://www.cdisc.org/stuff/contentmgr/files/0/919cb4ef843829170d470b37eb662aeb/misc/odm1_3_0_final.htm

[OpenLink] OpenLink Virtuoso Platform. Automated Generation of RDF Views over Relational Data Sources. Available from: http://docs.openlinksw.com/virtuoso/rdfrdfviewgnr.html.

[OPS] DIMDI - OPS - Operationen- und Prozedurenschlüssel [homepage on the internet]. DIMDI; [cited 2010 Jan 6]. Available from: http://www.dimdi.de/static/de/klassi/prozeduren/ops301/index.htm

[Part11] Guidance for Industry. Part 11, Electronic Records; Electronic Signatures — Scope and Application. [article on the internet]. U.S. Department of Health and Human Services, Food and Drug Administration; 2003 Aug [cited 2010 Jan 6]. Available from: http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm125125.pdf

[PASNet] Nexus AG. PAS-NET [homepage on the internet]. Nexus AG; [cited 2010 Jan 06]. Available from: http://www.nexus-medos.de/web/0/inter/?act=art&act2=show&art_id=dc_2007_07_26_en_8f503c6b85d77ff

[Payne2009] Payne PR, Embi PJ, Sen CK. Translational informatics: enabling high-throughput research paradigms. Physiol Genomics. 2009 Nov 6;39(3):131-40.

[Peyton2010] Peyton L, Hu J. Federated identity management to link and protect healthcare data. IJEB. 2010;8(3): 214-232.

[PWC2009] PricewaterhouseCoopers. Transforming healthcare through secondary use of health data [presentation on the internet]. [cited 2010 Dec 20]. Available from: http://www.pwc.com/us/en/healthcare/publications/secondary-health-data.jhtml

[PIX2010] IT Infrastructure Technical Framework Volume 2a (ITI TF-2a) Transactions Part A – Sections 3.1 – 3.28. Aug. 2010. [cited 2010 Dec 6]. Available from: http://www.ihe.net/Technical_Framework/upload/IHE_ITI_TF_Rev7-0_Vol2a_FT_2010-08-10.pdf

[Pommerening2005] Pommerening K, Reng M, Debold P, Semler S. Pseudonymisierung in der medizinischen Forschung - das generische TMF-Datenschutzkonzept. GMS Med Inform Biom Epidemiol. 2005;1(3):Doc17.

[Powell2005] Powell J, Buchan I. Electronic health records should support clinical research. J Med Internet Res. 2005 Mar 14;7(1):e4.

[Prasser 2011] Prasser F, Wurst SHR, Lamla G, Kohlmayer F, Blaser R, Schmelcher D, Vögele B, Kuhn KA. Informatics and Translational Medical Research: Challenges and Developments. it - Information Technology. 2011;53(5):217-26.

[RDF] W3C. Resource Description Framework (RDF): Concepts and Abstract Syntax [homepage on the internet]. W3C; [cited 2009 Jul 17]. Available from: http://www.w3.org/TR/rdf-concepts/

[RDB2RDF] W3C RDB2RDF Incubator Group Available from: http://www.w3.org/2005/Incubator/rdb2rdf/RDB2RDF_SurveyReport.pdf

[Rec2006] Council of Europe. Recommendation Rec(2006)4 of the Committee of Ministers to member states on research on biological materials of human origin. France 2006. Available from: https://wcd.coe.int/wcd/ViewDoc.jsp?id=977859

[Reng2006] Reng CM, Debold P, Specker C, Pommerening K. Generische Lösungen zum Datenschutz für die Forschungsnetze in der Medizin. Medizinisch Wissenschaftliche Verlaggeselsschaft Berlin. 2006.

[RFD2007] IHE. IHE Technical Frameworks: Retrieve Form for Data Capture (RFD) [serial on the internet]. 2007 Aug 15 [cited 2009 Jul 17]. Available from: http://static.ihe.net/Technical_Framework/upload/IHE_ITI_TF_Supplement_RFD_TI_2007_08_15.pdf

[Roden2008] Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balser JR, Masys DR. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin Pharmacol Ther. 2008 Sep;84(3):362-9.

[Safran2007] Safran C, BLoomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, Detmer DE. Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. J Am Med Inform Assoc. 2007;14(1):1-8.

[Schack 2010] Schack C, Birkle M, Havemann C, Heinze O, Krafczyk-Korth J, Möller A, Ostrzinski S, Reinecke P, Bergh B, Hoffmann W. GANI_MED - Forschungsplattform. GMDS 2010. [cited 2012 March 17]. Available from: http://www.gmds2010.de/cms/wordpress/wp-content/uploads/folien/246.pdf

[Schnell2009] Schnell R, Bachteler T, Reiher J. Privacy-preserving record linkage using Bloom filters. BMC Med Inform Decis Mak. 2009 Aug 25;9:41.

[Scitable] Scitable by Nature Education. Glossary [homepage on the internet]. [cited 2012 Aug 28]. Available from: http://www.nature.com/scitable/definition/snp-295

[SCXML] State Chart XML (SCXML) [homepage on the internet]. State Machine Notation for Control Abstraction; [cited 2010 Jul 08]. Available from: http://www.w3.org/TR/scxml/

[Shelley2010] Shelley M, Schlyer P. The Enterprise Master Person Index – Delivering better eHealth in Europe, the Middle East and Africa (EMEA). 2010. IBM AG; [cited 2012 Nov 15]. Available from: http://www-05.ibm.com/de/healthcare/literature/ibm-empi.pdf

[Sikuli] Project Sikuli. [homepage on the internet]. Available from: http://sikuli.org/

[SNOMED] IHTSDO – International Health Terminology Standards Development Organisation [homepage on the internet]. IHTSDO; [cited 2010 Jan 6]. Available from: http://www.ihtsdo.org/

[Sonntag 2005] Sonntag KC. Implementations of translational medicine. J Transl Med. 2005 Aug 30;3:33.

[Sung2003] Sung NS, Crowley WF Jr, Genel M, Salber P, Sandy L, Sherwood LM, Johnson SB, Catanese V, Tilson H, Getz K, Larson EL, Scheinberg D, Reece EA, Slavkin H, Dobs A, Grebb J, Martinez RA, Korn A, Rimoin D. Central challenges facing the national clinical research enterprise. JAMA. 2003 Mar 12;289(10):1278-87.

[Sweeney 2002] Sweeney L. k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems.2002; 10 (5):557-70.

[SWT] The Standard Widget Toolkit (SWT) [homepage on the internet]. Eclipse Foundation; [cited 2010 Jul 8] Available from: http://www.eclipse.org/swt/

[TätigDS 2006] Bayerischer Landtag. 22. Tätigkeitsbericht des Landesbeauftragten für den Datenschutz; [cited 2011 Dec 23] Available from: http://www.datenschutz-bayern.de/tbs/tb22/tb22.pdf

[Tiles2] Apache Tiles 2 [homepage on the internet]. Apache Software Foundation; 209. [cited 2010 Jan 6] Available from: http://tiles.apache.org/

[Tracy2008] Tracy RP. 'Deep phenotyping': characterizing populations in the era of genomics and systems biology. Curr Opin Lipidol. 2008 Apr;19(2):151-7.

[Wears2005] Wears RL, Berg M. Computer technology and clinical work: still waiting for Godot. JAMA. 2005 Mar 9;293(10):1261-3.

[Web services] Web of Services – W3C [homepage on the internet]. W3C; [cited 2010 Jan 6]. Available from: http://www.w3.org/standards/webofservices/

[Weng2012] Weng C, Appelbaum P, Hripcsak G, Kronish I, Busacca L, Davidson KW, Bigger JT. Using EHRs to integrate research with patient care: promises and challenges. J Am Med Inform Assoc. 2012 Sep-Oct;19(5):684-7.

[Wichmann 2011] Wichmann HE, Kuhn KA, Waldenberger M, Schmelcher D, Schuffenhauer S, Meitinger T, Wurst SH, Lamla G, Fortier I, Burton PR, Peltonen L, Perola M, Metspalu A, Riegman P, Landegren U, Taussig MJ, Litton JE, Fransson MN, Eder J, Cambon-Thomsen A, Bovenberg J, Dagher G, van Ommen GJ, Griffith M, Yuille M, Zatloukal K. Comprehensive catalog of European biobanks. Nat Biotechnol. 2011 Sep 8;29(9):795-7.

[Wurst2010] Wurst SHR. Dataspace Integration in der medizinischen Forschung. Technische Universität München. 2010.

[XDS2010] IHE. IT Infrastructure Technical Framework Volume 2b (ITI TF-2b) Transactions Part B – Sections 3.29 – 3.43. Aug. 2010. [cited 2010 Dec 6]. Available from: http://www.ihe.net/Technical_Framework/upload/IHE_ITI_TF_Rev7-0_Vol2b_FT_2010-08-10.pdf

[Yu2007] Yu J, Benatallah B, Saint-Paul R, Casati F, Daniel F, Matera M. A framework for rapid integration of presentation components. Proceeding WWW '07 Proceedings of the 16th international conference on World Wide Web ACM New York, NY, USA 2007

[Yuille2008] Yuille M, van Ommen GJ, Bréchot C, Cambon-Thomsen A, Dagher G, Landegren U, Litton JE, Pasterk M, Peltonen L, Taussig M, Wichmann HE, Zatloukal K. Biobanking for Europe. Brief Bioinform. 2008 Jan;9(1):14-24.

[Zhang2008] Zhang B, Bao L, Zhou R, Hu S, Chen P. A Black-Box Strategy to Migrate GUI-Based Legacy Systems to Web Services. Proceeding SOSE '08 Proceedings of the 2008 IEEE International Symposium on Service-Oriented System Engineering IEEE Computer Society Washington, DC, USA 2008.

Publikationsverzeichnis

Schlundt JW, **Lamla G**, Kuhn KA. Prozessunterstützung im Gesundheitswesen – Potentiale und Herausforderungen. Datenbank Spektrum. 2006;(17):7-11.

Wurst SHR, **Lamla G**, Schlundt J, Karlsen R, Kuhn KA: A Service-oriented Architectural Framework for the Integration of Information Systems in Clinical Research. In: Lee DJ, editor. Proceedings of the Twenty-First IEEE International Symposium on Computer-Based Medical Systems; 2008: Jyväskylä, Finland: IEEE Computer Science Press; 2008. p. 161-3.

Kuhn KA, Wurst SHR, Schmelcher D, **Lamla G**, Kohlmayer F, Wichmann HE. Integration von Biobanken für Forschungsaufgaben. GI-Edition: Lecture Notes in Informatics; GI Symposium 2009.

Wurst SHR, **Lamla G**, Prasser F, Kemper A, Kuhn KA. Einsatz von Dataspaces für die inkrementelle Informationsintegration in der Medizin. GI-Edition: Lecture Notes in Informatics; GI Symposium 2009.

Kuhn KA, Wurst SHR, Schmelcher D, **Lamla G**, Kohlmayer F, Wichmann HE. Integration von Biobanken für Forschungsaufgaben. GI-Edition: Lecture Notes in Informatics; GI Symposium 2009.

Wurst SHR, **Lamla G**, Prasser F, Kemper A, Kuhn KA. Einsatz von Dataspaces für die inkrementelle Informationsintegration in der Medizin. GI-Edition: Lecture Notes in Informatics; GI Symposium 2009.

Kuhn KA, Wurst SHR, Schmelcher S, **Lamla G**, Kohlmayer F. Identifying Biobanks, Subjects, and Specimens. BBMRI WP5, D5.2. 2009 Jul 31.

Dias A, Fisterer B, **Lamla G**, Kuhn KA, Hartvigsen G, Horsch A. Measuring physical activity with sensors: a qualitative study. Stud Health Technol Inform. 2009;150:475-9.

Lamla G, Blaser R, Prasser F, Wurst SHR, Rechl H, Gradinger R, Kuhn KA. Eine Umsetzung des IHE Single Source Konzepts für die translationale Forschung bei Knochen- und Weichteilsarkomen. GI-Edition: Lecture Notes in Informatics; GI Symposium 2010.

Wichmann HE, Kuhn KA, Waldenberger M, Schmelcher D, Schuffenhauer S, Meitinger T, Wurst SH, **Lamla G**, Fortier I, Burton PR, Peltonen L, Perola M, Metspalu A, Riegman P, Landegren U, Taussig MJ, Litton JE, Fransson MN, Eder J, Cambon-Thomsen A, Bovenberg J, Dagher G, van Ommen GJ, Griffith M, Yuille M, Zatloukal K. Comprehensive catalog of European biobanks. Nat Biotechnol. 2011 Sep 8;29(9):795-7.

Prasser F, Wurst SHR, **Lamla G**, Kuhn KA, Kemper A. Inkrementelle ontologiebasierte Informationsintegration für die translationale medizinische Forschung. GI-Edition: Lecture Notes in Informatics; GI Symposium 2011.

Prasser F, Wurst SHR, **Lamla G**, Kohlmayer F, Blaser R, Schmelcher D, Vögele B, Kuhn KA. Informatics and Translational Medical Research: Challenges and Developments. it - Information Technology. 2011;53(5):217-26.

Schleinkofer T, Villain S, **Lamla G**, Praßer F, Kuhn KA, Mansmann U. S3ULMU - Prototyp-Infrastruktur für die IT-Unterstützung klinischer Studien am Klinikum der Universität München. GI-Edition: Lecture Notes in Informatics; GI Symposium 2012.