Pseudonymisierungsdienst auf Basis von Web Services – erfolgreiche Praxis der Generischen Datenschutzkonzepte für medizinische Forschungsnetze

Semler SC[1], Drepper J[1], Pommerening K[2], Schlösser-Faßbender M[3], Schröder M[3], Rienhoff O[3,4]
[1]*TMF e.V., Berlin, Deutschland*
[2]*IMBEI, Univ. Mainz, Deutschland*
[3]*Tembit Software GmbH, Deutschland*
[4]*Inst. f. Medizininformatik, Univ. Göttingen, Deutschland*
sebastian.semler@tmf-ev.de

Einleitung und Fragestellung Für langfristige Verlaufsbeobachtungen in der klinischen Forschung sowie bei dem Aufbau von Biobanken ist eine Pseudonymisierung der Daten erforderlich, da einerseits einfache Anonymisierung keine Follow-Up-Beobachtungen erlaubt, andererseits datenschutzrechtlich ein offener Personenbezug für Forschungsdaten nicht gestattet ist. Innerhalb der Telematikplattform für medizinische Forschungsnetze (TMF) wurden 2003 in Abstimmung mit den Landesdatenschutzbeauftragten Konzepte erarbeitet, wie datenschutzgerechte Pseudonymisierungslösungen gestaltet sein müssen [1]. Diese werden aktuell fortgeschrieben und erweitert. Eine IT-Lösung zur Umsetzung dieser Konzepte konnte jedoch nicht am industriellen Markt gefunden werden, so dass eine Eigenentwicklung innerhalb der TMF notwendig wurde. Über die Realisierung und über den mittlerweile erfolgreichen Einsatz in Forschungsnetzen soll berichtet werden.
Material und Methoden Auf der Basis der generischen Datenschutzkonzepte der TMF wurden Konzepte und Pflichtenhefte zur Spezifikation und Implementierung der notwendigen Komponenten einer Pseudonymisierungslösung erarbeitet. An den Entwicklungen waren mehrere akademische und industrielle Partner beteiligt (u.a. Fraunhofer SIT und ISST, Debold & Lux, Schlumberger, interactive-systems GmbH, Tembit Software GmbH). Nach der Realisierung aussagefähiger Prototypen 2003/2004 konnte 2005/2006 eine komplette Reimplementierung unter Nutzung aktueller Technologien realisiert werden.
Ergebnisse und Diskussion Die generischen Datenschutzkonzepte der TMF definieren ein verteiltes Komponentenmodell für wissenschaftliche Forschungsnetze in der Medizin, welches inbesondere eine aus Datenschutzsicht notwendige Unabhängigkeit des administrativen Zugriffs auf verschiedene Komponenten und Anteile des Datenbestandes fordert. Diese Forderung nach einer verteilten Architektur passt gut zu großen und verteilt organisierten Forschungsnetzen wie den Kompetenznetzen in der Medizin. Kleinere Verbundforschungseinrichtungen können verteilte Infrastrukturen mit Hilfe zentraler Organisationen wie der TMF realisieren. Neben der Notwendigkeit einer räumlich und organisatorisch verteilten Infrastruktur stellt die sehr unterschiedliche Ausstattung der Verbundforschungseinrichtungen mit vorhandenen Softwaresystemen eine weitere wichtige Rahmenbedingung für die passende Implementierung datenschutzgerechter Pseudonymisierungsfunktionen dar. Aufgrund dieser Ausgangslage und der sehr guten Verfügbarkeit von Tools und Frameworks für die Implementierung und Nutzung von Web Services [2], wurde ein Konzept gemäß einer Service Oriented Architecture (SOA) entwickelt.
Letztlich spiegelt diese Wahl auch die umfangreiche Erfahrung der TMF wider: Bestimmte generische Komponenten und damit verbundene Funktionen kommen in verschiedenen Anwendungsfällen und Einrichtungen in gleicher Form aber unterschiedlichen Konstellationen und Prozessabbildungen zum Einsatz. Relevante Anwendungsfälle für Pseudonymisierungsfunktionen sind zum einen klinische Studien, wobei hier die sehr weitgehend regulierten Studien gemäß Arzneimittelgesetz (AMG) und solche nach Vorgaben des Medizinproduktegesetz (MPG) eine Sonderstellung einnehmen. Zum anderen ist aber auch der Aufbau und die Nutzung von Registern, wie auch die Durchführung von Kohortenstudien zu betrachten. Daneben stellen zunehmend auch Biomaterialbanken einen entscheidenden Erfolgsfaktor für medizinische Verbundforschungseinrichtungen dar. Zusätzlich zu diesen häufig rein wissenschaftlich ausgerichteten Szenarien wird in Zukunft aber auch die Integration von, bzw. der Datenaustausch zwischen Versorgungs- und Forschungsstrukturen vermehrt eine zentrale Aufgabe darstellen. All diesen Anwendungsfällen ist gemein, dass sie umfassend und gleichzeitig datenschutzgerecht nur mit Hilfe von Pseudonymisierungslösungen realisierbar sind oder sein werden.
Eine erste wichtige Komponente ist im Regelfall für das Identitätsmanagement zuständig. Hier werden zentral identifizierende Daten der Probanden und Patienten (IDAT) verwaltet und nichttriviale Pseudonyme erster Ordnung generiert und herausgegeben. Um Patientendaten auch bei fehlerhaften Eingaben in unterschiedlichen Softwaresystemen korrekt zuordnen zu können, ist hier oft ein fehlertolerantes Matchingverfahren, optimalerweise auf Basis phonetischer Repräsentation der Daten, gewünscht. All dies stellt der PID-Generator der TMF mit wenigen und einfach nutzbaren Web Services für unterschiedlichste Infrastrukturen und Softwaresysteme zur Verfügung.
Für eine langfristige Speicherung und Nutzung von medizinischen Daten, insbesondere auch dann, wenn der Forschungszweck zum Zeitpunkt der Datenerhebung noch nicht eng eingegrenzt werden kann, ist jedoch aus Sicht des Datenschutzes eine zweite Stufe der Pseudonymisierung und daraus resultierend eine weitere Aufteilung von Datenbeständen und Zuständigkeiten unabdingbar. Dies führt zum notwendigen Einsatz weiterer Softwarekomponenten, die entsprechende Funktionen zur Verfügung stellen. Zentral hierfür ist der eigentliche Pseudonymisierungsdienst, der zwischen einstufig und zweistufig pseudonymisierten Datenbeständen vermittelt. Die zweistufige Pseudonymisierung wird durch symmetrische Verschlüsselung des einstufigen Pseudonyms erreicht. Die zugehörigen medizinischen Daten (MDAT) werden asymmetrisch verschlüsselt und ohne Entschlüsselungsmöglichkeit für den Pseudonymisierungsdienst selbst durchgereicht. Die datenschutzgerechte Realisierung dieser sensiblen und komplexen Kommunikationsanforderung wird durch das Zusammenspiel von drei Komponenten erreicht, die jeweils sehr generische Funktionen zur Verfügung stellen. Die erste Komponente auf der Seite der einstufig pseudonymisierten Daten nimmt die unverschlüsselten medizinischen Daten und das einstufige Pseudonym entgegen, verschlüsselt dann asymmetrisch die medizinschen Daten, signiert den vollständigen Datensatz und schickt ihn an die zweite Komponente, den eigentlichen Pseudonymisierungsdienst. Dieser tauscht das einstufige Pseudonym (PID) gegen dessen symmetrisch verschlüsselte Entsprechung (PSN) und schickt den Datensatz weiter zur dritten Komponente auf der Seite der langfristigen Datenhaltung für übergreifende Forschungsvorhaben (Forschungsdatenbank, FDB). Jede der drei Komponenten stellt genau die hierfür benötigten Funktionen als Web Services zur Verfügung. Auf der Seite der einstufig pseudonymisierten Datenbestände, die häufig für Qualitätssicherungs- und Monitoringprozesse genutzt werden (kurz: Qualitätssicherungsdatenbank, QDB) stehen auf diesem Wege z.B. Funktionen zum Speichern, Löschen, Anonymisieren von medizinischen Daten incl. zweistufiger Pseudonymisierung zur Verfügung (vergl. Abb.). Wichtig ist, dass alle Komponenten und Funktionen so generisch angelegt sind, dass die eigentlichen medizinischen Daten in beliebigem Format (XML, ASCII, Binär) und inhaltlicher Struktur (z.B. CDISC-ODM, -SDTM [3], HL7) übermittelt werden können.
Die einzelnen Komponenten sind in unterschiedlichen Konstellationen in verschiedenen Mitgliedsverbünden der TMF im Einsatz. Gerade die Komponentenorientierung und die technische Umsetzung als Web Services, entsprechend einer Service Oriented Architecture, haben die Aufwände zur Integration dieses komplexen Lösungsansatzes in sehr heterogene Anwendungsumgebungen erheblich gemindert. Aktuell geraten zunehmend auch Anwendungsfälle aus der Patientenversorgung und hier besonders zum Aufbau eletronischer Patientenakten in den Blick. Erste Kontakte und Anfragen hierzu wurden bereits vermittelt. Aufgrund einer modernen und mächtigen Lösung sieht sich die TMF hierfür gut gerüstet.

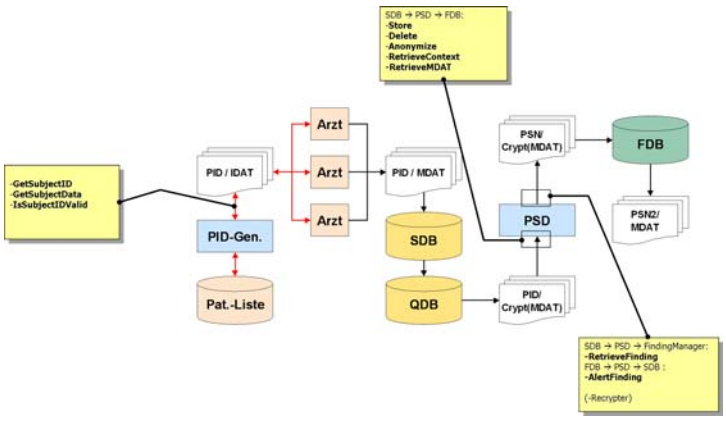

Abb.1: Webservice-Funktionen der Komponenten im Pseudonymisierungsdienst der TMF

Literatur

[1] Reng, CM, Debold, P, Specker, C, Pommerening, K. Generische Lösungen zum Datenschutz für die Forschungsnetze in der Medizin. Schriftenreihe der Telematikplattform für Medizinische Forschungsnetze (TMF). Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft; 2006

[2] Schulz, H, Schüler, P. Weltweites Werkeln: Das bringen Web Services für Anwender und Entwickler. c't 6/02 2002: S. 236

[3] CDISC: Clinical Data Interchange Standards Consortium: www.cdisc.org