D. Reinhardt, H. Langweg, B.C. Witt, M. Fischer et al. (Hrsg.): Sicherheit 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Göttingen 2020 135

Datenschutzgerechte und mehrseitig sichere IT-Plattformen für die medizinische Forschung

Tom Petersen[1]

Abstract:

Die medizinische Forschung ist in vielen Fällen auf die Nutzung von zu Patienten erhobenen Gesundheitsdaten angewiesen. Demgegenüber stehen jedoch die besondere Sensibilität dieser Daten und daraus resultierende Datenschutzanforderungen. Das hier vorgestellte Forschungsvorhaben beschäftigt sich mit dem Entwurf von datenschutzgerechten und sicheren IT-Plattformen für die Erhebung, Speicherung und Bereitstellung von Gesundheitsdaten zu Forschungszwecken, um diesen Zielkonflikt zu lösen. Hierzu werden rechtliche Aspekte, Sicherheitsinteressen beteiligter Akteure und mögliche Architekturen betrachtet sowie technische Maßnahmen vorgestellt, die bei der Erfüllung von Datenschutz- und Sicherheitsanforderungen genutzt werden können.

Keywords: Gesundheitsdaten; Medizinische Forschung; Privacy by Design; DSGVO; Kryptographische Schwellwertschemata; Searchable Encryption

1 Einleitung

In der medizinischen Forschung spielen datenbasierte Verfahren wie u. a. Methoden des maschinellen Lernens für die statistische Beurteilung des Erfolges von Behandlungsmethoden, -geräten oder Medikamenten eine große Rolle [Rü15, OE16]. Für die Forschung relevante medizinische Daten wie beispielsweise Risikofaktoren oder Blutwerte können innerhalb einer medizinischen Einrichtung während der Behandlung erhoben oder anschließend aus der Patientenakte bezogen werden . Die Daten einer einzigen Einrichtung sind jedoch für Forschungszwecke häufig nicht ausreichend, da insbesondere für selten auftretende Krankheiten oder zur Beurteilung regionaler Effekte eine nicht ausreichende Zahl von Fällen vorliegt. Hieraus entsteht der Bedarf, relevante Daten aus vielen Quellen für Forschungszwecke zentral nutzen zu können.

Dem Datenbedarf der medizinischen Forschung und einer zentralen Ablage der benötigten Daten steht jedoch die Sensibilität personenbezogener Gesundheitsdaten gegenüber. Die Veröffentlichung personenbezogener Gesundheitsdaten kann im schlimmsten Fall zu lebenslanger Stigmatisierung oder Benachteiligung Betroffener führen. Die Herausforderung besteht darin, Konzepte und Plattformen für die Datensammlung zu medizinischen

[1] Universität Hamburg, Arbeitsgruppe Sicherheit in Verteilten Systemen, Vogt-Kölln-Straße 30, 22307 Hamburg, Deutschland petersen@informatik.uni-hamburg.de
Die notwendige Patienteneinwilligung und Zweckbindung erhobener Daten seien an dieser Stelle nicht betrachtet.

doi:10.18420/sicherheit2020_13

Forschungszwecken zu erarbeiten, die diese gegensätzlichen Sicherheitsinteressen beachten und vermitteln [Be17].

Das hier vorgestellte Forschungsvorhaben befasst sich mit verschiedenen bei dem Entwurf von Plattformen für die medizinische Forschung relevanten Schwerpunkten, die im Folgenden vorgestellt werden: Abschnitt 2 beschäftigt sich mit rechtlichen Aspekten und Anforderungen. Abschnitt 3 beschreibt (teilweise gegensätzliche) Sicherheitsinteressen, die durch die verschiedenen beteiligten Akteure auftreten. Abschnitt 4 stellt beispielhaft einige Architekturen, die für Plattformen für die medizinische Plattformen genutzt werden können, und Bewertungskriterien für diese vor. Abschnitt 5 geht anschließend auf ergänzende Schutzmaßnahmen ein, die zur datenschutzgerechten Gestaltung bestimmter PLattformaspekte genutzt werden sollen.

2 Rechtliche Betrachtung

Das Spannungsfeld zwischen Datensensibilität und Forschungsnutzen wird bereits in der EU-Datenschutzgrundverordnung (DSGVO) deutlich. In Artikel 9 wird die Verarbeitung von medizinischen Daten ohne ausdrückliche Einwilligung der betreffenden Person grundsätzlich untersagt. Der Einschluss von Patientendaten in einem medizinischen Register kann daher im Normalfall nur durch die explizite Einwilligung des Patienten erfolgen [Po08]. Die Erwägungsgründe 52, 53 und 157 der DSGVO betonen jedoch auch die Wichtigkeit dieser Daten für die im öffentlichen Interesse liegende wissenschaftliche Forschung.

Weitere Artikel der DSGVO bilden einen rechtlichen Rahmen für das beschriebene Forschungsvorhaben, beispielsweise: Artikel 5 schreibt u.a. das Prinzip der Datenminimierung und die Sicherheit der Datenverarbeitung durch geeignete technische und organisatorische Maßnahmen vor. In den Artikeln 12 bis 23 werden die Rechte eines Betroffenen dargestellt. Artikel 25 stellt die Forderung nach Datenschutz durch Technikgestaltung und datenschutzfreundliche Voreinstellungen (Privacy by Design) auf.

Für das Forschungsvorhaben ergeben sich folgende Fragen: Wie können rechtliche Anforderungen wie die Gewährleistung von Betroffenenrechten in einer Plattform für die medizinische Forschung umgesetzt werden? Welche Prinzipien und konkreten technischen Maßnahmen können wie eingesetzt werden, um die Forderung nach Datenschutz durch Technikgestaltung zu erfüllen?

3 Mehrseitige Sicherheit

Dem Spannungsfeld gegensätzlicher Sicherheitsinteressen widmet sich aus technischer Sicht die mehrseitige Sicherheit [RPM96], indem sie versucht, technische Lösungsmöglichkeiten zum Ausgleich dieser möglicherweise gegensätzlichen Sicherheitsinteressen zu entwickeln und nutzbar zu machen.

Im Kontext von Plattformen für die medizinische Forschung müssen hier die Sicherheitsinteressen verschiedener Akteure betrachtet werden: Zu *Patienten* werden auf Basis einer Einwilligung sensible Gesundheitsdaten erfasst, deren Vertraulichkeit gegenüber anderen Akteuren soweit möglich erhalten werden muss. *Medizinisches Personal* in teilnehmenden Einrichtungen verarbeitet die Daten. Der Mitarbeiterzugriff auf Daten sollte protokolliert werden, um absichtliches oder unabsichtliches Fehlverhalten aufdecken zu können. Auf der anderen Seite können diese Daten die Überwachung von Mitarbeitern ermöglichen, was unter dem Aspekt des Mitarbeiterdatenschutzes weitestmöglich verhindert werden sollte. Der *Betreiber* einer Plattform hat ein Interesse an der Qualitätssicherung der erfassten Gesundheitsdaten, um eine hohe Reputation des Datensatzes und darauf basierenden Forschungsarbeiten zu gewährleisten. *Forscher* benötigen Zugriff auf für ihre Forschungsfragen relevante Ausschnitte der erfassten Daten. *Entwickler* und *Administratoren* der Plattform sollten abgesehen von für die Wartung und die Weiterentwicklung der Plattform unerlässlichen Daten keinen Zugriff auf erfasste Daten erhalten.

Für das Forschungsvorhaben ergeben sich folgende Fragen: Welche möglicherweise gegensätzlichen Sicherheitsinteressen haben die Beteiligten? Mithilfe welcher technischen Maßnahmen können diese gewahrt oder vermittelt werden?

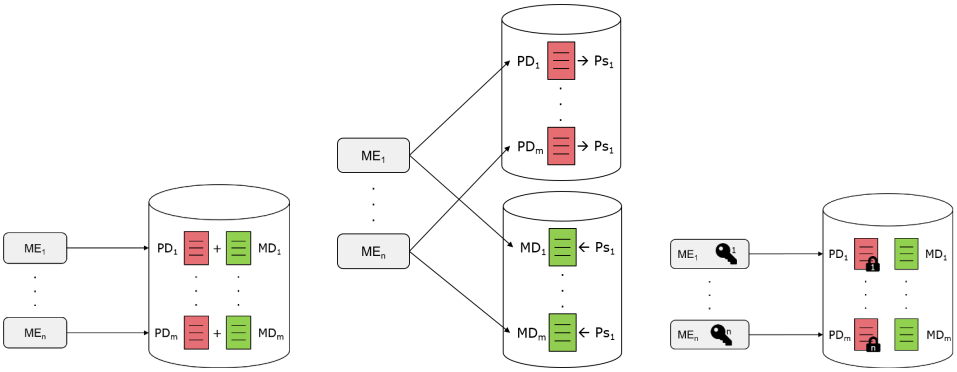
4 Architekturen für IT-Plattformen

In diesem Teil des Forschungsvorhabens sollen verschiedene Architekturen für Plattformen entworfen und evaluiert werden. Die Ausgangssituation ist dabei die folgende: In medizinischen Einrichtungen werden personenbezogene und medizinische Daten zu Patienten erfasst. Die Plattform soll es Forschern ermöglichen, zentral Zugriff auf eine für sie relevante Menge von medizinischen Daten zu erhalten. Zu diesem Zweck entworfene Architekturen beschreiben, wo welche Daten in welcher Form verarbeitet, gespeichert und bereitgestellt werden. Denkbar ist der Einsatz verschiedener Techniken, unter anderem die getrennte Verarbeitung personenbezogener und medizinischer Daten , eine Trennung der Datenhoheit (*informationelle Gewaltenteilung*), die Verwendung von Verschlüsselungsverfahren und die Nutzung von Pseudonymen zur Verknüpfung von Daten.

Die so entworfenen Architekturen können auf bestimmte kontextrelevante Eigenschaften überprüft werden, beispielsweise: Auf welche Daten erhalten beteiligte Akteure Zugriff und verhindert die Architektur nicht erforderlichen Datenzugriff? Ermöglicht die Plattform die Erfassung und Verknüpfung von medizinischen Daten eines Patienten zu mehreren Zeitpunkten (*zeitlicher Längsschnitt*) und in verschiedenen medizinischen Einrichtungen?

Beispielhaft seien hier drei mögliche Architekturen skizziert: Abbildung 1a zeigt eine Architektur, in der medizinische und personenbezogene Daten zentral an einer Stelle im Klartext gespeichert werden. Die Architektur in Abbildung 1b zeigt die getrennte

Die Reidentifikation von Patienten durch medizinische Daten muss im Rahmen des Forschungsvorhabens betrachtet werden, soll in diesem Abschnitt jedoch vorerst vernachlässigt werden.



(a) Zentrale Speicherung aller Daten

(b) Informationelle Gewaltenteilung

(c) Verschlüsselung der personenbezogenen Daten

Abb. 1: Darstellung von drei Beispielarchitekturen

Speicherung von medizinischen und personenbezogenen Daten und die Verknüpfung der Daten mittels Pseudonymen und damit eine Form der informationellen Gewaltenteilung. Abbildung 1c stellt eine Architektur dar, in der die personenbezogenen Daten zentral ausschließlich in verschlüsselter Form gespeichert werden, wobei die notwendigen Schlüssel nur in der entsprechenden medizinischen Einrichtung vorliegen.

Für das Forschungsvorhaben ergeben sich folgende Fragen: Welche sinnvollen Architekturen für Plattformen existieren und welche Eigenschaften erfüllen sie? Wie lassen sich die rechtlichen Anforderungen und Sicherheitsinteressen in entsprechenden Architekturen umsetzen (auch in Verbindung mit entsprechenden technischen Maßnahmen, siehe Abschnitt 5)?

5 Maßnahmen

Methoden aus den Bereichen der Kryptographie und der Privacy Enhancing Technologies (PETs) können dabei helfen, datenschutzgerechte und mehrseitig sichere Plattformen für die medizinische Forschung zu entwickeln. Nach aktuellem Stand sollen drei konkrete Bereiche in diesem Forschungsvorhaben untersucht werden:

5.1 Die Nutzung von kryptographischen Schwellwertschemata zur Durchsetzung des Mehraugenprinzips

Das Mehraugenprinzip ist eine Kontrollmaßnahme, bei der kritische Entscheidungen oder Handlungen nur durch mehrere Personen gemeinsam durchgeführt werden dürfen, um Missbrauch zu erschweren. Häufig wird das Prinzip nur durch organisatorische Maßnahmen umgesetzt. Kryptographische Schwellwertschemata wie Threshold Decryption [DF90] oder

Threshold Signatures [Ge96] erlauben die Durchführung kryptographischer Operationen nur durch die Kooperation mehrerer Personen, die jeweils im Besitz eines Teilschlüssels für die entsprechende Operation sind. Sie können dazu genutzt werden, das Mehraugenprinzip auch kryptographisch zu erzwingen. Für den Einsatz in Plattformen für die medizinische Forschung sind insbesondere praxisrelevante Fragen der Schlüsselverwaltung und möglicher Zugriffsstrukturen[4] zu betrachten, die in bisherigen Veröffentlichungen selten betrachtet werden.

Für das Forschungsvorhaben ergeben sich folgende Fragen: An welchen Stellen der Plattformen kann das Mehraugenprinzip zur Umsetzung von Sicherheits- oder Datenschutzanforderungen genutzt werden? Welche Probleme ergeben sich beim praktischen Einsatz kryptographischer Schwellwertschemata und wie können diese gelöst werden?

5.2 Die Anwendbarkeit von Searchable-Encryption-Schemata

Aufgrund der Sensibilität von Gesundheitsdaten können verschlüsselte Datenbanken eine sinnvolle Maßnahme zum Schutz dieser Daten darstellen (vergleiche Abbildung 1c). Dies verhindert jedoch das direkte Suchen in den Datensätzen bzw. erfordert die vorhergehende Entschlüsselung aller Daten. Ansätze aus dem Bereich der Searchable Encryption [SWP00] können hier als möglicherweise performante Alternative zu diesem Vorgehen in Betracht gezogen werden.

Für das Forschungsvorhaben ergeben sich folgende Fragen: Welche Searchable-Encryption-Schemata bieten sich für den Anwendungskontext an und welche Arten von Suchanfragen erlauben sie? Erlaubt der Einsatz dieser Schemata Rückschlüsse[5] auf die unterliegenden Daten?

5.3 Ansätze zur Anonymitätsmessung von veröffentlichen Gesundheitsdatensätzen

Sollen medizinische Daten für die Forschung verwendet und ggf. veröffentlicht werden, so muss durch Anonymisierung der Daten sichergestellt werden, dass Patienten nicht durch die Daten reidentifiziert werden können. Hierzu können Maße zur Bestimmung der Anonymität wie k-Anonymität [Sw02] oder Differential Privacy [Dw06] genutzt werden.

Im Forschungsvorhaben soll geklärt werden, welche Verfahren sich zur Messung der Anonymität im Kontext von Gesundheitsdaten eignen, welche Randbedingungen es für Ihren Einsatz gibt und wie sie praktisch umgesetzt werden können. Ein besonderer Fokus soll hier auf den Zusammenhang zwischen Anonymisierung und der dezentralen Datensammlung in verschiedenen Krankenhäusern gelegt werden.

[4] Das Mengensystem aller möglichen Mengen von Teilnehmern, die die kryptographische Operation gemeinsam durchführen können, wird Zugriffstruktur genannt.

[5] Es wurde beispielsweise gezeigt, dass statistische Betrachtungen dazu genutzt werden können, mithilfe von Range Queries verschlüsselte Daten zu rekonstruieren [Gr19].

6 Fazit

Mit dem hier vorgestellten Forschungsvorhaben soll zwischen den besonderen Schutzanforderungen von Gesundheitsdaten auf der einen Seite und der Relevanz von IT-Plattformen für den Fortschritt in der medizinischen Forschung auf der anderen Seite vermittelt werden.

Durch den Einsatz geeigneter Architekturen und Maßnahmen wird versucht, den Forderungen des Artikel 25 der DSGVO (Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen) im Kontext von IT-Plattformen für die medizinische Forschung nachzukommen.

Literaturverzeichnis

[Be17] Behrendt, Christian-Alexander; Pridöhl, Henning; Schaar, Katrin; Federrath, Hannes; Debus, Eike Sebastian: Klinische Register im 21. Jahrhundert. Der Chirurg, 88(11):944–949, 2017.

[DF90] Desmedt, Yvo; Frankel, Yair: Threshold cryptosystems. In (Brassard, Gilles, Hrsg.): Advances in Cryptology - CRYPTO '89. Jgg. 435 in Lecture Notes in Computer Science. Springer, S. 307–315, 1990.

[Dw06] Dwork, Cynthia; McSherry, Frank; Nissim, Kobbi; Smith, Adam: Calibrating Noise to Sensitivity in Private Data Analysis. In (Halevi, Shai; Rabin, Tal, Hrsg.): Theory of Cryptography. Springer Berlin Heidelberg, Berlin, Heidelberg, S. 265–284, 2006.

[Ge96] Gennaro, Rosario; Jarecki, Stanisław; Krawczyk, Hugo; Rabin, Tal: Robust threshold DSS signatures. In (Maurer, Ueli, Hrsg.): Advances in Cryptology - EUROCRYPT '96. Jgg. 1070 in Lecture Notes in Computer Science. Springer, S. 354–371, 1996.

[Gr19] Grubbs, P.; Lacharité, M.; Minaud, B.; Paterson, K. G.: Learning to Reconstruct: Statistical Learning Theory and Encrypted Database Attacks. In: IEEE Symposium on Security and Privacy (S&P) 2019. 2019.

[OE16] Obermeyer, Ziad; Emanuel, Ezekiel J: Predicting the future—big data, machine learning, and clinical medicine. The New England journal of medicine, 375(13):1216, 2016.

[Po08] Pommerening, K.; Debling, D.; Kaatsch, P.; Blettner, M.: Register zu seltenen Krankheiten. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz, 51(5):491–499, May 2008.

[RPM96] Rannenberg, Kai; Pfitzmann, Andreas; Müller, Günter: Sicherheit, insbesondere mehrseitige IT-Sicherheit. it+ti - Informationstechnik und Technische Informatik, 38(4):7–10, 1996.

[Rü15] Rüping, Stefan: Big Data in Medizin und Gesundheitswesen. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz, 58(8):794–798, Aug 2015.

[Sw02] Sweeney, Latanya: k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(05):557–570, 2002.

[SWP00] Song, Dawn Xiaoding; Wagner, David; Perrig, Adrian: Practical techniques for searches on encrypted data. In: Proceedings of the 2000 IEEE Symposium on Security and Privacy. IEEE, S. 44–55, 2000.