ZENTRALISIERTE PSEUDONYMISIERUNG VON MEDIZINISCHEN PATIENTENDATEN

Heurix J[1], Mück T[2], Neubauer T[3]

Kurzfassung

Die elektronische Gesundheitsakte (ELGA) verspricht eine Verbesserung der Kommunikation zwischen Gesundheitsdienstleistern, wodurch die Qualität der medizinischen Versorgung von Patienten verbessert und die Kosten gesenkt werden können, jedoch ein Pool an hochsensiblen Patientendaten erstellt wird, welcher entsprechend vor Missbrauch geschützt werden muss. Dieser Beitrag präsentiert eine neue Architektur für ein zentralisiertes Pseudonymisierungsservice, welche Gesundheitsdaten nur entkoppelt von den persönlichen Stammdaten der Patienten speichert.

1. Einleitung

Im heutigen Gesundheitssystem hat die Verfügbarkeit von aussagekräftigen Informationen einen wesentlichen Einfluss auf die Entscheidungen in der Patientenversorgung und daher auf die Qualität der Behandlung von Patienten. Durch die Einführung der elektronischen Gesundheitsakte kann das Auftreten von unerwünschten Wechselwirkungen von Medikamenten beträchtlich verringert werden, da den Medizinern nun Richtlinien hinsichtlich des Zusammenwirkens von verschiedenen Medikamenten zur Verfügung stehen. Weiters erlaubt die elektronische Gesundheitsakte erhebliche Einsparungen durch die Digitalisierung von diagnostischen Testresultaten und medizinischen Abbildungen [4]. Jedoch sind mit der elektronischen Speicherung von Gesundheitsdaten erhebliche Bedenken bezüglich der Privatsphäre der Patienten verbunden. Mit der zunehmenden Verknüpfung informativer Daten muss der hohen Sensibilität der persönlichen Informationen Rechnung getragen werden. Diese Patientendaten stellen ein lohnendes Ziel für Angreifer dar und sind ebenfalls für Arbeitgeber von großem Interesse, die aufgrund dieser Daten beispielsweise eine Anstellung verweigern könnten. Daher besteht ein zunehmender politischer und gesellschaftlicher Druck, diese Gesundheitsdaten entsprechend zu schützen, um die Vertraulichkeit der Patientendaten zu gewährleisten. Beispielsweise ist seit 2005 die Verarbeitung und Weitergabe von persönlichen Daten durch die Direktive 95/46/EC [5] rechtlich geregelt. Zudem ist das Recht auf Privatsphäre eines Bürgers in Artikel 8 der Konvention zum Schutze der Menschenrechte und Grundfreiheiten [6] festgehalten. Um die Vertraulichkeit der Gesundheitsdaten bei der Verwendung, Weitergabe und Speicherung zu erhalten, wurden diverse Methoden, sogenannte p*rivacy enhancing technologies PETs* (vgl. [7]), vorgeschlagen. Diese Ansätze richten sich jedoch oftmals nicht nach aktuellen gesetzlichen Anforderungen (vgl. [5, 9, 10]), erfüllen nicht die grundlegenden Sicherheitsanforderungen (vgl. [2, 18, 19]) oder sind nicht für klinische Studien (sekundäre Nutzung der Daten) geeignet.

[1] Technische Universität Wien

[2] Sozialversicherungsanstalt der gewerblichen Wirtschaft

[3] Secure Business Austria

Dieser Beitrag präsentiert PIPE (Pseudonymization of Information for Privacy in e-Health) als ein zentralisiertes Pseudonymisierungsservice basierend auf früheren Arbeiten [11, 12, 18]. Das System ist für die entkoppelte Speicherung von medizinischen Daten und Patientenstammdaten verantwortlich, wobei die eigentlichen medizinischen Datensätze von externen Applikationen gewartet und verwendet werden. Der Patient agiert als alleiniger Dateneigentümer, der nach eigenem Ermessen Zugriffsautorisierungen für vertrauenswürdige Parteien definieren kann. Da die Pseudonymisierungsprozedur umfangreiche kryptographische Operationen erfordert und softwarebasierte Kryptomodule anfällig für Schadsoftware sind, setzt PIPE auf ein zentrales Hardwaresicherheitsmodul (HSM) [1] als sichere Umgebung für die benötigten Ver- und Entschlüsselungsoperationen. Benutzerauthentifizierung am System erfolgt durch einen persönlichen Security-Token (z.B. Smart Card).

2. Hintergrund

Grundsätzlich kann die Vertraulichkeit von Patientendaten durch zwei verschiedene präventive Techniken gewährleistet werden, nämlich Anonymisierung und Verschlüsselung, welche jedoch wesentliche Nachteile aufweisen: Während Anonymisierung, also das Entfernen von patientenidentifizierenden Informationen aus medizinischen Datensätzen, unumkehrbar ist und ein Patient daher nicht mehr von Ergebnissen aus klinischen Studien, die diese Daten einschließen, profitieren kann, verhindert die Vollverschlüsselung der Nutzdaten die anonymisierte Sekundärnutzung der Daten ohne ausdrückliche Genehmigung des Patienten (Entschlüsselung mit seinem geheimen Schlüssel), wobei dieser dabei seine Identität preisgeben muss. Unter Berücksichtigung der teils großen Datenmengen kann eine Vollverschlüsselung darüber hinaus bei häufiger Anwendung sehr zeitaufwendig sein. Reaktive Sicherheitsmaßnahmen, die in Zusammenhang mit ELGA oftmals angeführt werden (z.B. Protokollierung), stellen keine ausreichende Sicherheitsmaßnahme dar, um den Datenschutz zu gewährleisten und sollten höchstens in Ergänzung zu anderen Sicherheitsmaßnahmen eingesetzt werden.

Pseudonymisierung hingegen umgeht die Schwächen der gewöhnlichen Anonymisierung und der Vollverschlüsselung der Nutzdaten, indem die Identifikationsdaten im Grunde anonymisiert abgelegt werden, jedoch diese Anonymisierung unter bestimmten Umständen, nämlich bei Kenntnis eines bestimmten Geheimnisses, umkehrbar ist [14]. Nach der Depersonalisierung, der Identifizierung und Trennung persönlicher Daten von den Nutzdaten (vgl. [17]), und anschließender Pseudonymisierung kann für nicht autorisierte Benutzer keine *direkte* Beziehung zwischen den Individuen und deren Daten mehr hergestellt werden. Existierende Ansätze haben verschiedenste Mängel: Die von Pommerening (vgl. [15, 16]) entwickelten Ansätze verwenden eine Kombination aus Hash-Techniken und Verschlüsselung, wobei ein einzelner zentraler Geheimschlüssel verwendet wird, der eine Schwachstelle darstellt, da ein Angreifer, der diesen einzelnen Schlüssel kennt, Zugang zu allen medizinischen Patientendaten erhält. Der von Peterson (vgl. [13]) entwickelte Ansatz hat mehrere gravierende Nachteile: Da alle Schlüssel, die zur Dechiffrierung der medizinischen Daten nötig sind, in der Datenbank gespeichert sind, kann ein Angreifer, der Zugang zu dieser Datenbank erlangt, alle Informationen entschlüsseln. Darüber hinaus ist der Benutzer selbst für die Wahl seines persönlichen Schlüssels verantwortlich, wobei das System überprüft, ob dieser Schlüssel in der Datenbank schon vorhanden ist. Dadurch ist es einem potentiellen Angreifer möglich, einen schon existierenden Schlüssel zu eruieren und somit sofortigen Zugriff auf dazugehörige medizinische Daten zu erlangen. Der Ansatz von Stingl und Slamanig (vgl. [21]) speichert die Daten in einer zentralen Datenbank und verwendet Smart Cards für die Authentifizierung. Das System verwendet mehrere Sub-Identitäten, die durch Pseudonyme repräsentiert werden, wobei eine öffentlich bekannt ist und die restlichen Identitäten geheim bleiben. Das vom deutschen Bundesministerium für Ge-

sundheit geförderte Fraunhofer Institut (vgl. [3, 8]) entwickelte eine Architektur, die sich auf eine vollständige Verschlüsselung der Nutzdaten stützt, was für klinische Studien nicht praktikabel ist.

3. Architektur

Das zentralisierte Pseudonymisierungsservice PIPE stützt sich auf zwei sichere Instanzen, den Security-Token des Benutzers und das zentrale Hardwaresicherheitsmodul. Der Security-Token agiert als Authentifizierungsmittel, wobei der geheime Schlüssel des Tokens (äußerer privater Schlüssel) nur durch Eingabe eines PINs zugänglich ist. Dadurch wird eine „Zwei-Faktor-Authentifizierung" erreicht. Darüberhinaus dient er als clientseitiges Kryptomodul. Das HSM hat in PIPE zwei Aufgaben: Zunächst werden dort alle zentrale, kryptographische Operationen durchgeführt, wodurch sich die clientseitigen Operationen am Token nur auf die während der Benutzerauthentifizierung (sowie Session-Schlüssel-Operationen) beschränken. Weiters agiert das HSM als sichere Umgebung für den inneren symmetrischen Schlüssel (siehe Hüllenmodell), da dieser nur innerhalb der HSM im Klartext präsent ist und so gegen unberechtigten Zugriff geschützt ist. Ziel dieser HSM-Architektur ist es, dass ein Angreifer unter keinen Umständen, d.h. auch wenn er physischen Zugang zu der Datenbank erlangt, eine Verbindung zwischen Gesundheitsdaten und zugehörige Patienten herstellen kann, auch wenn die meisten kryptographischen Schlüssel in der Datenbank abgelegt werden. Dies wird durch das mehrstufige Hüllenmodell (*Abbildung 1*) erreicht, welches aus drei Schichten besteht, die die verschiedenen Aspekte des Sicherheitsmodells realisieren und die der Benutzer passieren muss, um die medizinischen Daten mittels zugewiesener Pseudonyme abrufen zu können.

Die äußere Hülle, die Authentifizierungsschicht, wird durch das äußere asymmetrische Schlüsselpaar gebildet, das sich im Besitz des Benutzers auf dem Security-Token befindet (S* in *Abbildung 1*). Die innere Hülle ist die Rechteschicht, die durch das innere asymmetrische Schlüsselpaar und den inneren symmetrischen Schlüssel des Benutzers realisiert werden (K* in *Abbildung 1*), wobei der innere private Schlüssel mit dem äußeren öffentlichen Schlüssel und der innere symmetrische Schlüssel mit dem inneren privaten Schlüssel chiffriert in der Datenbank abgelegt sind. Die innerste Schicht stellt die pseudonymisierte Datenschicht dar (CD in *Abbildung 1*), die medizinische Daten mit den (Klartext-) Pseudonymen referenziert. Die Zugehörigkeit der Pseudonyme zu den Benutzern wird durch Verschlüsselung der Relation Pseudonym/Benutzer-ID mit dem inneren symmetrischen Schlüssel gesichert. Das HSM übernimmt alle zentrale, kryptographische Operationen (siehe Umrandungen in *Abbildung 1*). Datenzugriff erfordert die folgenden Schritte: Zunächst muss sich der Benutzer am PIPE-Server, welcher aus dem HSM und der PIPE-Logik besteht, authentifizieren. Dieser Prozess basiert auf einem *Challenge/Response*-Verfahren mit einer Zufallszahl als Challenge und schließt das äußere asymmetrische Schlüsselpaar des Benutzers und seine ID auf dem Security-Token und dem asymmetrischen Schlüsselpaar des PIPE-Servers mit ein. Nach erfolgter Authentifizierung wird der (verschlüsselte) innere private Schlüssel aus der Datenbank an den Benutzer gesendet und im Security-Token mit dem äußeren privaten Schlüssel entschlüsselt. Zusätzlich wird ein symmetrischer Session-Schlüssel[1] ausgetauscht, der für die Vertraulichkeit der Daten beim Transfer zwischen Client und Server sorgt. Dann wird der innere private Schlüssel an das HSM transferiert, um den inneren symmetrischen Schlüssel[2] zu entschlüsseln, welcher wiederum für die

[1] Austausch erfolgt wiederum mittels asymmetrischer Verschlüsselung, sodass dieser nur im Security Token und im HSM im Klartext vorliegt.

[2] Die Verwendung des zusätzlichen symmetrischen Schlüssels anstelle der direkten Chiffrierung der Pseudonyme mit dem inneren privaten Schlüssel verbessert die Geschwindigkeit der Verschlüsselungsoperationen und erhöht die allgemeine Sicherheit, da der (entschlüsselte) innere symmetrische Schlüssel nur serverseitig im HSM verwendet wird.

Entschlüsselung der Pseudonym/Benutzer-Relation nötig ist. Mit den Klartextpseudonymen kann nun der gewünschte Nutzdatensatz abgerufen werden. Der innere symmetrische Schlüssel verbleibt nun bis zum Ende der Session im HSM und steht für weitere Operationen zur Verfügung.

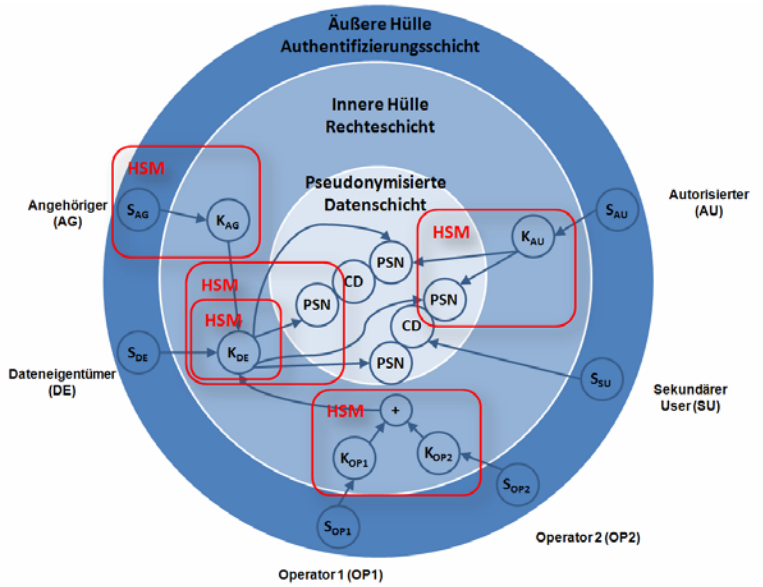

Abbildung 1: PIPE-Hüllenarchitektur (ohne Administrator)

Das Hüllenmodell unterstützt verschiedene Benutzertypen (Rollen), die verschiedene Datenzugriffsrechte und Aufgaben besitzen:

- Dateneigentümer: Der Dateneigentümer (Patient) hat die volle Kontrolle über seine Daten und hat auch die Möglichkeit, Datenzugriffsautorisierungen über einige bestimmte Datensätze (Autorisierte) oder über die Gesamtheit seiner Daten (Angehörige) zu definieren und ebenso zu widerrufen. Für den Datenzugriff verwendet der Eigentümer sogenannte Root-Pseudonyme, die mit Ausnahme von ausgewählten Angehörigen nur ihm bekannt sind.

- *Angehöriger*: Der Angehörige ist eine vertrauenswürdige Partei, der der innere private Schlüssel des Eigentümers zur Verfügung gestellt wird, wodurch der Angehörige in weiterer Folge die Root-Pseudonyme des Eigentümers mitverwenden kann und somit auf sämtliche existierende Daten sowie automatisch auf alle zukünftig hinzugefügten Datensätze autorisiert wird. Der innere private Dateneigentümerschlüssel wird mit dem inneren symmetrischen Schlüssel des Angehörigen verschlüsselt in der Datenbank abgelegt. Zusätzlich werden die IDs beider Parteien mit den jeweiligen inneren symmetrischen Schlüsseln gesichert, sodass die Relation nur für diese beiden Parteien ersichtlich ist. Um die Autorisierung zu widerrufen, muss der Dateneigentümer diese Relation aus der Datenbank entfernen.

- *Autorisierter*: Im Gegensatz zum Angehörigen wird dem Autorisierten (z.B. Gesundheitsdienstleister) nur Zugriff auf bestimmte Daten gewährt und er erhält auch keinen Zugang zu dem inneren privaten Eigentümerschlüssel. Für den Autorisierten werden hingegen neue Pseudonyme, sogenannte Shared-Pseudonyme, generiert, die sowohl dem Autorisierten als auch dem Dateneigentümer bekannt sind und mit den jeweiligen Datensätzen referenziert werden. Für jede so durchgeführte Autorisierung wird ein eigenes Pseudonym generiert. Analog zum Angehörigen werden die Pseudonyme und die IDs des Eigentümers und des Autorisierten mit

den beiden inneren symmetrischen Schlüsseln verschlüsselt. Widerruf der Autorisierung geschieht durch Löschen des Pseudonyms.

- *Administrator*: Der Administrator hat in PIPE keine Zugriffsrechte auf die Daten der Dateneigentümer (oder genauer, kann die einzelnen Gesundheitsdatensätze nicht den einzelnen Patienten zuweisen), da er weder den inneren privaten Schlüssel noch Shared-Pseudonyme mit den Dateneigentümern teilt. Seine Aufgaben beschränken sich auf administrative Tätigkeiten wie dem Management der Benutzerinstanzen inklusive Vergabe der Security-Token.

- *Operator*: Da der Security-Token beschädigt oder verloren werden und daher der innere private Schlüssel aufgrund des fehlenden äußeren privaten Schlüssels nicht mehr entschlüsselt werden kann, muss eine entsprechende Sicherheitskopie dieses Schlüssels in der Datenbank gespeichert werden. Um Missbrauch dieser Kopie zu verhindern, wird in PIPE das Konzept des *Secret Sharing* [20] angewandt, bei welchem der geheime Schlüssel in mehrere Teile zerlegt wird und eine bestimmte Anzahl dieser Teile benötigt wird, um den Schlüssel wiederherzustellen (*threshold*). Die Aufgabe der Operator ist es nun, diese Teile zusammen mit der ID des jeweiligen Benutzers *aufzubewahren* (mit ihren inneren symmetrischen Schlüsseln verschlüsselt in der Datenbank abzuspeichern). Diese ID kann vorher mit dem geheimen Schlüssel des HSM (Logikschlüssel) verschlüsselt werden, welcher niemandem außer dem HSM zu Verfügung steht, um den Operatoren die Kenntnis des Benutzers zu verwehren und somit den möglichen Missbrauch weiter einzuschränken.

- *Sekundärer User*: Der sekundäre User (z.B. eine Forschungseinrichtung) hat die Genehmigung, medizinische Daten für Forschungszwecke einzusehen, kann allerdings nicht die Verbindung zwischen medizinischen Daten und den persönlichen Daten des entsprechenden Patienten herstellen. Da die Informationen bereits pseudonymisiert gespeichert sind, wird dem sekundären User direkter Zugang zu den (medizinischen) Daten gewährt. Der Zugang wird mitgeloggt, um den Patienten über alle Abfragen zu informieren.

Wie schon erwähnt, werden verschiedene Pseudonymtypen benötigt, um die verschiedenen Zugriffsberechtigungsstufen von Dateneigentümer und Autorisierter zu realisieren. Einerseits gibt es die vom Dateneigentümer verwendeten *Root-Pseudonyme*, die nur ihm (sowie etwaigen Angehörigen) bekannt sind, andererseits die *Shared-Pseudonyme*, die individuelle Autorisierungen darstellen und für jedes Eigentümer/Autorisierter/Datensatz-Tupel einzigartig sind. Darüber hinaus wird eine weitere Unterteilung benötigt, um die Stammdaten von den Gesundheitsdaten logisch zu trennen, da auch die Stammdatensätze pseudonymisiert und somit von der Benutzer-ID entkoppelt gespeichert werden. Die *Identification-Pseudonyme* sind mit den Stammdaten referenziert, wobei die *Health-Pseudonyme* sich auf die medizinischen Datensätze beziehen. Diese Pseudonyme gehen eine 1:1-Parent/Child-Beziehung ein, die die Relation Patient/medizinische Daten darstellen und mit dem inneren symmetrischen Schlüssel verschlüsselt vor unberechtigtem Zugriff geschützt sind. Identification- und Health-Pseudonyme existieren sowohl als Root- als auch als Shared-Pseudonyme.

4. Zusammenfassung

Durch die Einführung der elektronischen Gesundheitsakte können Gesundheitsdaten in strukturierter und erweiterbarer Form gesammelt und für autorisierte Gesundheitsdienstleister zur Verfügung gestellt werden. Problematisch sind jedoch die Bedenken bezüglich der Gewährleistung der Vertraulichkeit bei Bereitstellung der Daten in elektronischer Form, besonders bei zentraler Speiche-

rung. Dieser Beitrag präsentierte eine Methode für die zentralisierte Pseudonymisierung von Gesundheitsdaten und gab einen Überblick über die Sicherheitsarchitektur und die unterstützten Rollen. Dieser Ansatz sieht den Patienten als alleinigen Dateneigentümer vor und stützt sich auf ein zentrales Hardwaresicherheitsmodul für die sichere Ausführung der nötigen kryptographischen Operationen, sowie auf persönliche Security-Token für die Benutzerauthentifizierung.

5. Literatur

[1] ATTRIDGDE, J., An overview of hardware security modules, Technischer Bericht, SANS Institute, 2002.
[2] BARROWS, R. C. & CLAYTON, P. D., Privacy, confidentiality, and electronic medical records, Journal of the American Medical Informatics Association 13, 139–148, 1996.

[3] CAUMANNS, J., Der Patient bleibt Herr seiner Daten, Informatik-Spektrum, 321–331, 2006.

[4] ERNST, F. R. & GRIZZLE, A. J., Drug-related morbidity and mortality: Updating the cost-of-illness model, Journal of the American Pharmacists Association 41(2), 192–199, 2001.

[5] EUROPÄISCHE UNION, Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Official Journal of the European Communities L 281, 31–50, 1995.

[6] EUROPARAT, European Convention on Human Rights, Loukis G. Loucaides (Hrsg.), BRILL, 2007.

[7] FISCHER-HÜBNER, S., IT-Security and Privacy: Design and Use of Privacy-Enhancing Security Mechanisms, Springer, 2001.

[8] FRAUNHOFER INSTITUT, Spezifikation der Lösungsarchitektur zur Umsetzung der Anwendungen der elektronischen Gesundheitskarte, 2005.

[9] HINDE, S., Privacy legislation: a comparison of the US and European approaches, Computers and Security 22(5), 378–387, 2003.

[10] HORNUNG, G., GÖTZ, C. F.-J. & GOLDSCHMIDT, A. J. W., Die künftige Telematik-Rahmenarchitektur im Gesundheitswesen, Wirtschaftsinformatik 47, 171–179, 2005.

[11] NEUBAUER, T. & KOLB, M. An Evaluation of Technologies for the Pseudonymization of Medical Data Springer Studies in Computational Intelligence, 2009

[12] NEUBAUER, T. & MÜCK, T., Ein System zur Pseudonymisierung von Gesundheitsdaten, Tagungsband e-Health2008 & eHealth Benchmarking 2008, Österreichische Computer Gesellschaft, 2008.

[13] PETERSON, R. L., Patent: Encryption system for allowing immediate universal access to medical records while maintaining complete patient control over privacy, US Patent US 2003/0074564 A1, 2003.

[14] PFITZMANN, A. & KÖHNTOPP., M., Anonymity, Unlinkability, Unobservability, Pseudonymity, and Identity Management - A Consolidated Proposal for Terminology, in Lecture Notes in Computer Science, Springer, 2005.

[15] POMMERENING, K., Medical Requirements for Data Protection, in Proceedings of IFIP Congress (2), 533–540, 1994.

[16] POMMERENING, K. & RENG, M., Medical and Care Compunetics 1, IOS Press, Kapitel: Secondary use of the Electronic Health Record via Pseudonymisation, 441–446, 2004.

[17][15] RECTOR, A., ROGERS, J., TAWEEL, A., INGRAM, D., KALRA, D., MILAN, J., SINGLETON, P., GAIZAUSKAS, R., HEPPLE, M., SCOTT, D. & POWER, R., Clef - joining up healthcare with clinical and post-genomic research, in Proceedings of UK e-Science All Hands Meeting, 203–211, 2003.

[18] RIEDL, B., NEUBAUER, T., GOLUCH, G., BOEHM, O., REINAUER, G. & KRUMBOECK, A., A secure architecture for the pseudonymization of medical data, in Proceedings of the Second International Conference on Availability, Reliability and Security, 318–324, 2007.

[19] SCHABETSBERGER, T., AMMENWERTH, E., GÖBEL, G., LECHLEITNER, G., PENZ, R., VOGL, R. & WOZAK, F., What are functional requirements of future shared electronic health records?, Connecting Medical Informatics and Bio-Informatics, 1070–1075, 2005.

[20] SHAMIR, A., How to share a secret, Commun. ACM 22(11), 612-613, 1979.

[21] STINGL, C. & SLAMANIG D., Berechtigungskonzept für ein e-Health-Portal, e-Health 2007 – Medical Informatics meets eHealth 227, 138-140, Österreichische Computer Gesellschaft 2007.