Pseudonyme – ein Kompromiß zwischen Anonymisierung und Personenbezug

Klaus Pommerening[1]

Fragestellung

Bei der Auswertung medizinischer Daten, sei es für Forschungszwecke, sei es zur Qualitätssicherung, ist der Personenbezug der Daten nicht eigentlich nötig. Er wird aber trotzdem oft mitgeführt, um verschiedene Daten eines Falls zusammenführen oder Daten nacherheben zu können. Das ist mit den Vorgaben der Datenschutzgesetze nur so lange vereinbar, wie es zur Erfüllung der jeweiligen Zweckbestimmung unumgänglich ist. Daß dies jedoch meistens nicht der Fall ist, zeigen die folgenden Ausführungen.

Durch die Einführung von Pseudonymen kann der Personenbezug so verschleiert werden, daß faktische Anonymität entsteht, ohne die gewünschte Verwertung der Daten nennenswert zu behindern. Bei

- *Anonymität* besteht ein Bezug nur auf eine Gesamtmenge von Individuen,
- *Pseudonymität* besteht ein Bezug zu einem einzelnen Individuum, dessen Identität allerdings nicht erkennbar ist,
- *Personenbezug* ist das einzelne Individuum in seiner Identität erkennbar.

Bekannt ist die Verwendung von Pseudonymen durch Buchautoren. Was diesen zugestanden wird, sollte auch Bürgern – etwa in ihrer Eigenschaft als Patienten – mit ihren Daten gewährt werden.

Die angemessene Realisierung verschiedener Typen von Pseudonymen in der Informationstechnik beruht auf kryptographischen Methoden. Mit diesen lassen sich oft scheinbar konkurrierende Ziele vereinbaren – auf der Grundlage verblüffender mathematischer Konstruktionen. Bei geeigneter Umsetzung bleibt die Zusammenführbarkeit der Daten im erwünschten Maß trotz Pseudonymität erhalten [6]; Transaktionen unter Pseudonym können beweissicher gestaltet werden [1, 3]; unter kontrollierten Umständen kann das Pseudonym aufgehoben werden. Beispielhaft werden zwei Anwendungen im Bereich des Gesundheitswesens vorgestellt: Eingesetzt werden Pseudonyme bereits bei der Krebsregistrierung; vorgeschlagen wird der Einsatz bei der Krankenkassenabrechnung nach dem GSG.

Methodische Grundlagen

Typen von Pseudonymen: Pseudonyme kann man etwa nach der *Erzeugungsart* unterscheiden:

- *deterministisch:* Das Pseudonym wird durch eine schlüsselabhängige Einweg- oder Hashfunktion[2] von einer vertrauenswürdigen zentralen Instanz aus invarianten Daten, etwa Identitätsdaten, erzeugt.
- *willkürlich:* Das Pseudonym wird nach einem festen Einweg-Algorithmus vom Benutzer aus einem Geheimnis, einer Paßphrase, erzeugt;
- *zufällig:* Das Pseudonym wird frei gewählt oder nach einem Zufallsverfahren erzeugt. Solche Pseudonyme können als Einmal-Pseudonyme etwa zur Zusammenführung verschiedener Datenquellen zu Forschungszwecken dienen [5]. Ansonsten sind sie nur wiederverwendbar, wenn sie in einer Referenzliste gespeichert werden.

Ein weiteres Unterscheidungskriterium ist die *Verknüpfbarkeit*: Wer kann verschiedene Pseudonyme einer Person zusammenführen? Wer ist Geheimnisträger? Von wem und unter welchen Umständen kann das Pseudonym aufgehoben werden? Im Idealfall ist der Besitzer des Pseudonyms auch der Erzeuger und gleichzeitig der einzige Geheimnisträger, der das Pseudonym lüften kann; dies entspricht der Idee der informationellen Selbstbestimmung. Bei man-

[1] Institut für Medizinische Statistik und Dokumentation der Johannes-Gutenberg-Universität, 55101 Mainz
[2] Unter einer Hash-Funktion wird hier stets eine kryptographisch sichere, d. h., nicht effizient umkehrbare Prüfsummen-Funktion verstanden.

cher Anwendung, z. B. für die Krebsregistrierung, ist dies allerdings nicht wünschenswert. Verknüpfbarkeit ist mit dem Zweck eines Pseudonyms aber nur zu vereinbaren, wenn die Fähigkeit dazu auf eine vertrauenswürdige Instanz beschränkt bleibt, die entweder eine Referenzliste („Codebuch“) führt oder die Pseudonyme nach einem deterministischen Verfahren erzeugt. Der Vorteil eines Codebuchs ist seine einfache Konzeption und Verständlichkeit, der Nachteil seine Größe und damit seine erschwerte Geheimhaltung; bei einem deterministischen Verfahren muß dagegen nur der Schlüssel geschützt werden.

Auch die *Beweissicherheit* wird durch Einschalten einer vertrauenswürdigen Instanz erreicht, die entweder eine Referenzliste führt oder ein vom Besitzer willkürlich oder zufällig erzeugtes Pseudonym durch blinde Unterschrift beglaubigt, wie im nächsten Abschnitt beschrieben.

Blinde Unterschrift: Unverknüpfbare, aber trotzdem beweissichere Pseudonyme werden nach Chaum [1] mit dem *Prinzip* der blinden elektronischen Unterschrift gewonnen:

a) *Unterschrift:* Ein (elektronisches) Dokument wird unterschrieben, ohne daß der Unterschreibende dessen Inhalt erkennen kann. Die Unterschrift bestätigt also nicht den Inhalt des Dokuments, sondern nur die Tatsache der Vorlage durch eine bestimmte Person zu einem bestimmten Zeitpunkt. Ein Analogon wäre die Unterschrift auf der Rückseite eines Papierdokuments.

b) *Prüfung:* Dokument und Unterschrift werden vorgelegt. Der Prüfende kann erkennen, ob die Unterschrift zum Dokument gehört und rechtmäßig erlangt wurde.

c) *Anonymität:* Niemand, auch nicht der Unterzeichner, kann, wenn er Dokument und Unterschrift vorgelegt bekommt, diese dem Besitzer zuordnen oder den Unterschriftsvorgang rekonstruieren.

Die *Realisierung* dieser scheinbar paradoxen Anforderungen geht von der gewöhnlichen elektronischen Unterschrift aus. Der Unterzeichner N (wie „Notar“) hat eine Funktion S, mit der er zu einem Dokument D (das etwa durch eine Zeichenkette oder eine Datei in binärer Form repräsentiert wird) eine Unterschrift $S(D)$ erzeugen kann; diese Funktion hängt von einem Schlüssel ab, der Geheimnis von N ist. Zur Überprüfung durch jedermann dient ein zugehöriger öffentlicher Schlüssel. Will nun die Besitzerin A (wie „Alice“) eines Dokuments D dieses von N unterschreiben lassen, ohne daß N den Inhalt erfährt, so transformiert sie es in eine unleserliche Gestalt $C(D)$, wobei die Transformation C (wie „Camouflage“) von einer Zufallszahl abhängt, die A als Geheimnis behält. Durch N's Unterschrift wird $S(C(D))$ erzeugt. Danach entfernt A die Camouflage durch Rücktransformation $C'(S(C(D))) \stackrel{?}{=} S(D)$. Damit das Verfahren funktioniert, muß hier die Gleichheit stehen, also C' eine Art Umkehrtransformation von C sein; C' muß aus C und bekannten Parametern leicht bestimmbar sein. Solche Transformationen lassen sich mit den bekannten Schemata zur elektronischen Unterschrift, etwa nach dem RSA- oder dem ElGamal-Verfahren, tatsächlich leicht finden; für die mathematischen Einzelheiten sei auf [1, 2] verwiesen.

Anwendungsbeispiele

Krebsregister: Das Verfahren der Krebsregistrierung mit Hilfe von Pseudonymen, wie es im Mainzer Pilotprojekt [6] durchgeführt wird, erfüllt folgende *Anforderungen:*

1. Das Register muß in der Lage sein, durch Abgleich mehrfache Meldungen desselben Falls zu erkennen („record linkage“).
2. Die Abgleichsprozedur soll Synonym- und Homonym-Fehlerraten minimieren, um eine brauchbare Datenqualität zu gewährleisten.
3. Die Register der verschiedenen Bundesländer sollen ihre Datenbestände abgleichen können.

4. Unter gewissen kontrollierten Umständen soll die Aufdeckung eines Pseudonyms möglich sein (z. B. um Folge-Informationen für eine epidemiologische Studie erheben zu können).
5. Der Besitzer selbst soll sein Pseudonym nicht aufdecken können. – Die Entscheidung über die Aufklärung des Patienten bleibt dem behandelnden Arzt vorbehalten.

Realisierung: Zentrale Instanz des Verfahrens ist die sogenannte Vertrauensstelle, die die Rolle einer vertrauenswürdigen Instanz spielt. Forderung 1 wird durch deterministische Erzeugung eines Pseudonyms mit Hilfe einer schlüsselabhängigen Hash-Funktion erfüllt. Um Forderung 4 zu erfüllen, wird das Pseudonym um einen zweiten Teil ergänzt, der aus den Identitätsdaten durch asymmetrische Verschlüsselung gewonnen wird. Der Schlüssel dafür ist ein Geheimnis der Vertrauensstelle; der Rückschlüssel, der zur kontrollierten Reidentifizierung dient, wird von einer externen Aufsichtsstelle verwahrt. Damit ist auch Forderung 5 erfüllt. Um Forderung 3 nachzukommen, wird bei der Bildung des ersten Teils des Pseudonyms im ersten Schritt eine bundeseinheitliche schlüsselfreie Hash-Funktion, im zweiten ein länderspezifischer Schlüssel verwendet. Beim länderübergreifenden Abgleich wird die Verschlüsselung in den beiden beteiligten Vertrauensstellen aufgehoben, so daß die reinen Hash-Werte wieder sichtbar sind; diese werden dann mit einem gemeinsamen Einmalschlüssel verschlüsselt, der speziell für dieses eine Abgleichsverfahren gebildet wird. Forderung 2 steht in direktem Konflikt zur Anonymität. Hier wurde durch Zerlegung des Hash-Wertes in mehrere „Kontrollnummern“ versucht, einen Kompromiß zu finden, der die Anonymität nur so weit wie unbedingt nötig schwächt; diese Schwächung wird durch sehr restriktive Handhabung des Zugangs zu den Pseudonymen kompensiert. Für die Einzelheiten des Verfahrens sei auf [5, 6] verwiesen.

Krankenkassenabrechnung: Als zweites Beispiel wird anhand des vollständigen Verfahrensablaufs gezeigt, wie sich der Personenbezug bei der Krankenkassenabrechnung vermeiden läßt. Das Verfahren wurde bereits in [4] grob skizziert. Hier spielt die Rechtssicherheit und damit die Beglaubigung von Pseudonymen durch blinde Unterschrift eine wichtige Rolle.

Sollen die Patienten gegenüber den Krankenkassen hinter Pseudonymen versteckt werden, sind folgende *Anforderungen* zu erfüllen:

1. Die Krankenkassen müssen bei der Abrechnung der Behandlung zweifelsfrei erkennen, daß die Leistungen für eines ihrer Mitglieder erbracht wurden.
2. Die Krankenkassen sollen keine personenbezogenen Krankheitsgeschichten sammeln können.
3. Die Krankenkassen sollen aber zur Kalkulation ihrer Risiken einzelfallbezogene Auswertungen über Krankheitsverläufe und Kosten für bestimmte Krankheitsbilder erstellen können.

Ein mögliches Verfahren zur *Realisierung* sieht so aus[3]: Die Patientin geht zur Krankenkasse und erhält eine Versichertenkarte. Sie wählt zu Hause oder an vertrauenswürdiger Stelle, etwa bei ihrem Hausarzt, eine Kontrollzahl, camoufliert sie und überträgt sie auf die Karte. Dann läßt sie sie von der Krankenkasse (blind) unterschreiben; danach entfernt sie die Camouflage wieder. Das Pseudonym wird auf der Versichertenkarte durch eine PIN geschützt, hinter der ein kryptographisches Verschlüsselungsverfahren steckt; solche Verfahren gehören zu den üblichen Standards der Chipkartentechnik.

Bei ärztlicher Behandlung legt die Patientin die Versichertenkarte vor und schaltet das Pseudonym durch Eingabe ihrer PIN frei; der Arzt übernimmt das Pseudonym, prüft es auf Gültigkeit und verwendet es zur Abrechnung. Es ersetzt also die Versichertennummer. Die Krankenkasse erkennt durch Prüfung der Unterschrift, daß die behandelte

[3] Das Verfahren ist zum besseren Verständnis vereinfacht dargestellt; die Rolle der kassenärztlichen Vereinigung bleibt außer Acht, ebenso die Möglichkeit, auch den Arzt vor der Krankenkasse durch ein Pseudonym zu schützen. Ein umfassenderer Vorschlag, der dies auch berücksichtigt, wird zur Zeit von der GMDS-Arbeitsgruppe „Datenschutz in Krankenhausinformationssystemen“ erarbeitet.

Patientin bei ihr versichert ist, kann mit dem Arzt abrechnen und die Daten der Patientin zusammenführen. Sie kann die Daten aber nicht der konkreten Patientin zuordnen. Da der behandelnde Arzt die Zuordnung zwischen Pseudonym und Identität sowieso erkennt und der Schweigepflicht unterliegt, schadet es nichts, wenn für die Pseudonymerzeugung sein Praxiscomputer eingesetzt wird.

Ein entsprechendes Verfahren zur Abrechnung von Rezepten wurde bereits von B. Struif [7] vorgestellt, zusammen mit einer funktionierenden Musterimplementation.

Das Verfahren zur pseudonymen Abrechnung ist auch auf private Krankenversicherungen übertragbar. Es würde diese zu einer fairen Praxis beim Vertragsabschluß zwingen. Da die Regelung für alle Anbieter einzuführen wäre, wäre sie auch wettbewerbsneutral.

Die Verwendung von Pseudonymen zur Krankheitskostenabrechnung eröffnet die Möglichkeit, eventuell sogar weitere Daten zu übermitteln, ohne den Datenschutz weiter auszuhöhlen. Hier bietet sich also für die Krankenkassen die Chance, durch verbesserte Datenlage zu aussagefähigeren Ergebnissen bei der Auswertung ihrer Daten zu gelangen.

Andere mögliche Anwendungen: elektronisches Geld, alle Arten von anonymen Berechtigungsausweisen, anonyme Verträge, elektronische Wahlen.

Diskussion

Durch die Einführung von Pseudonymen läßt sich in vielen Anwendungsbereichen der Informationstechnik in der Medizin ein tragbarer Kompromiß zwischen dem informationellen Selbstbestimmungsrecht und dem Datenhunger von Forschung und Gesundheitswesen finden. Kryptographische Pseudonyme stellen eine Grundtechnik des praktischen Datenschutzes dar. Sie sollten, wo immer möglich, eingesetzt werden.

Literatur

1. Chaum, D.: Security without identification: Transaction systems to make Big Brother obsolete. Communications of the ACM 28, 1985, 1030–1045.
2. Horster, P.; Petersen, H.: Classification of blind signature schemes and examples of hidden and weak blind signatures. In: *EUROCRYPT ‘94*. Berlin; Springer Verlag 1994.
3. Pfitzmann, B.; Waidner, M.; Pfitzmann, A.: Rechtssicherheit trotz Anonymität in offenen digitalen Systemen. Datenschutz und Datensicherung 14, 1990, 243–253 & 305–315.
4. Pommerening, K.: Datenschutz in Krankenhausinformationssystemen. In: Brüggemann, H. H.; Gerhardt-Häckl, W.: *Verläßliche IT-Systeme, Proceedings der GI-Fachtagung VIS ‘95*. Braunschweig, Vieweg 1995, 5–22.
5. Pommerening, K.; Miller, M.; Schmidtmann, I.; Michaelis, J.: Pseudonyms for cancer registry. (Zur Veröffentlichung eingereicht.)
6. Schmidtmann, I.; Michaelis, J.; Pommerening, K.: Pilotstudie zum Aufbau eines bevölkerungsbezogenen Krebsregisters in Rheinland-Pfalz. In: Pöppl, S. J.; Lipinsky, H.-G.; Mansky, T. (Hrsg.): *38. Jahrestagung der GMDS: Medizinische Informatik – ein integrierender Teil arztunterstützender Technologien.* München; MMV Medizin Verlag 1993, 399–403.
7. Struif, B.: Datenschutz bei elektronischen Rezepten und elektronischem Notfallausweis. In: *Vertrauenswürdige Informationstechnik für Medizin und Gesundheitsverwaltung*. Erfurt; TeleTrusT Deutschland 1994.