DATENSCHUTZRECHTLICHE ANFORDERUNGEN BEI BIG DATA ANWENDUNGEN

Dr. Dipl. Inform. Bernd Schütze, M.D., LL.B.

Workshop der GMDS-AG „AAL und Assistierende Gesundheitstechnologien“ am 2014-09-10, 59. GMDS-Jahrestagung

HEALTHCARE SOLUTIONS

ZU MEINER PERSON

Ausbildung

- Studium Informatik
  (FH-Dortmund)
- Studium Humanmedizin
  (Uni Düsseldorf / Uni Witten/Herdecke)
- Studium Jura
  (Fern-Uni Hagen)

Weitergehende Schulungen

- Zusatzausbildung Datenschutzbeauftragter
  (Ulmer Akademie für Datenschutz und IT-Sicherheit)
- Zusatzausbildung Datenschutz-Auditor
  (TüV Rheinland)
- Zusatzausbildung Medizin-Produkte-Integrator
  (VDE Prüf- und Zertifizierungsinstitut)

Berufserfahrung

- 12 Jahre klinische Erfahrung
- 21 Jahre IT im Krankenhäusern
- 19 Jahre Datenschutz im Gesundheitswesen

Mitarbeit in Verbänden

- GMDS
- Berufsverband Medizinischer Informatiker e.V. (BVMI)
- Berufsverband der Datenschutzbeauftragten Deutschlands e.V. (BvD)
- Gesellschaft für Datenschutz und Datensicherung e.V. (GDD)
- HL7 Deutschland e.V.
- Fachverband Biomedizinische Technik e.V. (fbmt)

VERSUCH EINER „BIG-DATA“-DEFINITION

Big Data bezeichnet **Daten-Mengen**, die

- **zu groß**, oder
- **zu komplex** sind, oder sich
- **zu schnell ändern**,

um sie

- **mit händischen** und
- **klassischen Methoden**

der **Datenverarbeitung auszuwerten**.

(Quelle: Deutsche Wikipedia, http://de.wikipedia.org/wiki/Big_Data)

WAS IST ALSO „BIG DATA“?

- Analyse großer umfangreicher Datenbestände
- Analyse erfolgt in sehr kurzer Zeit
- Daten meist unstrukturiert
- Analyse kann nach allen Merkmalen erfolgen
- Aus Analyse dieser Daten können
  - Vorhersagen,
  - Hochrechnungen,
  - Zusammenhänge und
  - Muster

  erkannt werden
- Anfallende unstrukturierte Datenmengen können nach unterschiedlichsten Kriterien ausgewertet werden
- Auswertungen können zweck-ungebunden erfolgen

BIG DATA: NUR EIN HYPE?

- Die Hälfte aller Big Data-Projekte werden vorzeitig beendet*
- Spitzenposition in Gartner's aktuellem Hype Cycle for Emerging Technologies**
  - Big Data nur ein neuer Hype?
  - Big Data = „Data Analytics"
  - Also eher ein neues Gewand für „Data Warehouse" und „Data Mining"?

* Infochimps (2013) Intelligent applications: the big data theme for 2013. Online unter http://blog.infochimps.com

** Gartner's 2012 hype cycle special report evaluates the maturity of 1,900 technologies. Online unter http://www.gartner.com/technology/research/hype-cycles. Abruf am 2012-11-08

WAS MACHT DANN BIG DATA AUS?

- Täglich neu produzierten Menge > 2,5 Exabytes, also 2,5 1018 Bytes*
- Jährlich wachsende Telekommunikationskapazität von fast 30 %**
- Technologische Entwicklungen
  - mobile content
  - sensor-based content

  bieten zuvor nicht vorhandene Möglichkeiten
- Auch „Tragbare Computer“ waren einmal ein „Hype“...

* McAfee A, Brynjolfsson E (2012) Big data: the management revolution. Harvard Business Review

** Hilbert M, López P (2011) The world’s technological capacity to store, communicate, and compute information. Science 332:60–65

HERKUNFT DER DATEN

- Netzwerke
  - Facebook, Twitter, Google, Xing, LinkedIn. Apple, Microsoft, Scype, Friendship24, Parship, ...
- Einkauf im Netz
  - Payback, Happy Digits
- Kundenkarten
  - Douglas-Kundenkarte, Lufthansa Milres&More, EC-Karte, Kreditkarte, Krankenkassenkarte
- Handy
  - Standort, Wann wurde mit wem gesprochen, Termine, Kamera, Mikrofon
- Navy des Autos
  - Serviceschnittstelle des intelligenten Stromzählers
- E-Mails
- Regierung
  - Schweizer Bundesarchiv (opendata.admin.ch): Energie, Ökonomie, Umwelt, Verkehr, Finanzen, Wirtschaft und Gesundheit
- ...

EINSATZ IN DER MEDIZIN?

- Data Mining kann zu Optimierung der medizinischen Versorgung führen durch Auswertung von
  - klinischen,
  - epidemiologischen,
  - bildgebenden,
  - molekulargenetischen, aber auch
  - ökonomischen

  Daten

➔ Big Data in der medizinischen Behandlung:
häufig personenbezogen

ANFORDERUNGEN BZW. HERAUSFORDERUNGEN

Damit Big Data Erfolg hat müssen folgende Randbedingungen erfüllt werden:

- Mehr Kosteneffizienz im Datenmanagement, z.B. durch neue Technologien
  - Quantum-Computing
  - In-Memory-Datenbanksysteme
- Passende Tools zur Datenanalyse
  - Textanalyse
  - Semantische Analyse
- Intelligente Datenauswahl
  - Große Datenmengen aus verschiedenen Quellen
  - Intelligente Geschäftsmodelle müssen Datenvielfalt einschränken
- Hohe Datenqualität
  - Ziel: 99% der benötigten Daten liegen vor, 1% wird beim Kunden erhoben
  - Daten zeit-, inhalts- und bedeutungskonsistent über die verschiedenen Datenquellen hinweg gespeichert
- Datenschutz und Datensicherheit

HERAUSFORDERUNG DATENSCHUTZ: EINWILLIGUNG

Bestehende Forderung der Datenschutzgesetze, aber auch aktuell in der Entwicklung bestehender Vorschriften (EU, USA), verlangen:

- Einwilligung ist freiwillig
- Einwilligung bedarf der Schriftform
- Es ist hinzuweisen auf:
  - Zweck der Speicherung
  - ggf. der Übermittlung
  - auf Folgen der Verweigerung der Einwilligung
  - Löschung
  - Widerrufsmöglichkeit
  - Geltungsdauer der Einwilligung

HERAUSFORDERUNG DATENSCHUTZ: EINWILLIGUNG

Neue IT-Verfahren notwendig, denn

- Einwilligung sowie die Freiwilligkeit derselben müssen nachgewiesen werden
- Elektronische Abbildung Schriftform
  - (Qualifizierte) Elektronische Signatur
- Folgen der Verweigerung der Einwilligung müssen Alternative beinhalten, ansonsten keine Freiwilligkeit
  - Alternatives Kreditinstitut, Alternatives Schuhgeschäft,...
  - Monopolisten können keine Alternative anbieten!
    (Ist Amazon ein Monopolist? Ist eBay ein Monopolist? ...)
- Wie geht man mit Widerruf der Einwilligung um, wenn Daten weitergegeben wurden?
  - Beispiel elektronischen Patientenakten (ggfs. basierend auf IHE-Cookbook)

HERAUSFORDERUNG DATENSCHUTZ: TRANSPARENZ

Bestehende Forderung der Datenschutzgesetze, aber auch aktuell in der Entwicklung bestehender Vorschriften (EU, USA), verlangen Aufklärung bzgl.:

- Wer speichert,
- wer nutzt
- welche m einer Daten
- zu welchem Zweck
- an welchem Ort
- auf welche Weise?

HERAUSFORDERUNG DATENSCHUTZ: TRANSPARENZ

- Big Data ist Vorratsdatenspeicherung
  - von Kunden wie auch von potentiellen Kunden
- Ich weiß noch nicht, wer wann was wozu braucht
- Betroffener muss aber informiert werden
- Zugleich darf Konkurrenz nicht erfahren, welche Daten genutzt werden

➔ Neue informationstechnische Verfahren zur Information Betroffener notwendig

HERAUSFORDERUNG DATENSCHUTZ: ZWECKBINDUNG

- Einwilligung eines Betroffenen ist an den Zweck der Datennutzung gebunden
- Big Data:
  - Zweck der Datennutzung zum Zeitpunkt der Datenerhebung nicht erkennbar
  - Entweder neue Einwilligung notwendig oder anonyme Datennutzung
  - Anonyme Datennutzung schwierig

HERAUSFORDERUNG DATENSCHUTZ: VERTRAULICHKEIT

- Big Data = Daten werden zwischen Organisationen ausgetauscht
- Befugter Zugriff =
  - Zugriff mit Erlaubnis des Betroffenen

  oder
  - Zugriff auf Daten, die nicht personenbezogen oder personenbeziehbar sind
- Vertraulichkeit heute überwiegend durch Pseudonymisierung gewährleistet (aber unter Erhalt des Personenbezugs)

PSEUDONYMISIERUNG: PROBLEME BEI BIG DATA

- Anwendung von korrelierendem (Hintergrund-) Wissen
  - Big Data führt Daten zusammen, die vorher nicht bekannt waren
  - Wissen des Datenauswerters bzw. Datennutzers daher nicht im Vorhinein abschätzbar
- Beispiel: Anonymität im Netz
  - IP-Adresse wird nicht preisgegeben – Anonym?
  - Wie viele Internet-Nutzer in Göttingen nutzen
    - Windows/Linux/Mac/...
    - mit Betriebssystemversion xy mit Patchstand z
    - den Browser xy in Version ..
    - verwenden eine Bildschirmauflösung von ... x ...
    - die verwendete Sprache ist portugiesisch
    - Der Computer hat einen Arbeitsspeicher von .. GB
    - Die CPU des Computers ist xy
    - und der Provider ist xy

    ➔ Wird alles beim Aufruf einer Seite mit übermittelt und kann ausgewertet werden

PSEUDONYMISIERUNG: PROBLEME BEI BIG DATA

gmds
Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V.

Z.B. durch Amazon:

PSEUDONYMISIERUNG: PROBLEME BEI BIG DATA

gmds
Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V.

Beispiel aus der Medizin: Re-“Anonymisierung“ von Biodaten

Identifying Personal Genomes by Surname Inference

Melissa Gymrek,[1,2,3,†] Amy L. McGuire,[4] David Golan,[1] Eran Halperin,[7,8,9] Yaniv Erlich[1,*]

Sharing sequencing data sets without identifiers has become a common practice in genomics. Here, we report that surnames can be recovered from personal genomes by profiling short tandem repeats on the Y chromosome (Y-STRs) and querying recreational genetic genealogy databases. We show that a combination of a surname with other types of metadata, such as age and state, can be used to triangulate the identity of the target. A key feature of this technique is that it entirely relies on free, publicly accessible Internet resources. We quantitatively analyze the probability of identification for U.S. males. We further demonstrate the feasibility of this technique by tracing back with high probability the identities of multiple participants in public sequencing projects.

Surnames are paternally inherited in most human societies, resulting in their co-segregation with Y-chromosome haplotypes (*1*–*5*). Based on this observation, multiple genetic genealogy companies offer services to reunite distant patrilineal relatives by genotyping a few dozen highly polymorphic short tandem repeats across the Y chromosome (Y-STRs). The association between surnames and haplotypes can be confounded by nonpaternity events, mutations, and adoption of the same surname by multiple founders (*5*). The genetic genealogy community addresses these barriers with massive databases that list the test results of Y-STR haplotypes along with their corresponding surnames. Currently, there are at least eight databases and numerous surname project Web sites that collectively contain hundreds of thousands of surname haplotype records (table S1).

The ability of genetic genealogy databases to breach anonymity has been demonstrated in the past. In a number of public cases, male adoptees and descendants of anonymous sperm donors used recreational genetic genealogy services to genotype their Y-chromosome haplotypes and to search the companies' databases (*6*–*9*). The genetic matches identified distant patrilineal relatives and pointed to the potential surnames of their biological fathers. By combining other pieces of demographic information, such as date and place of birth, they fully exposed the identity of their biological fathers. Lunshof *et al.* (*10*) were the first to speculate that this technique could expose the full identity of participants in sequencing projects. Gitschier (*11*) empirically approached this hypothesis by testing 30 Y-STR haplotypes of CEU participants in these databases and reported that potential surnames can be detected. [CEU participants are multigenerational families of northern and western European ancestry in Utah who had originally had their samples collected by CEPH (Centre d'Etude du Polymorphisme Humain) and were later reconsented to participate in the HapMap project.] However, these surnames could match thousands of individuals and the study did not pursue full reidentification at a single-person resolution.

Our goal was to quantitatively approach the question of how readily surname inference might be possible in a more general population, apply this approach to personal genome data sets, and demonstrate end-to-end identification of individuals with only public information. We show that full identities of personal genomes can be exposed via surname inference from recreational genetic genealogy databases followed by Internet searches. In all cases in which individuals were studied who had donated DNA samples, the informed consent statements they had signed stated privacy breach as a potential risk and the data usage terms did not prevent reidentification. Representatives of relevant organizations that funded the original studies were notified and confirmed the compliance of this study with their guidelines (*12*).

As a primary resource for surname inference, we focused on Ysearch (www.ysearch.org) and

[1]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA. [2]Harvard-Massachusetts Institute of Technology (MIT) Division of Health Sciences and Technology, MIT, Cambridge, MA 02139, USA. [3]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [4]Department of Molecular Biology and Diabetes Unit, Massachusetts General Hospital, Boston, MA 02114, USA. [5]Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX 77030, USA. [6]Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv 69978, Israel. [7]School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel. [8]Department of Molecular Microbiology and Biotechnology, Tel-Aviv University, Tel Aviv 69978, Israel. [9]The International Computer Science Institute, Berkeley, CA 94704, USA.

*To whom correspondence should be addressed. E-mail: yaniv@w.mit.edu

PSEUDONYMISIERUNG UND BIG DATA

Neue Konzepte zur Pseudonymisierung erforderlich, z. B.

- Anwendungen erkennen, wann Daten nicht gemischt werden können, ohne dass Pseudonymisierung aufgehoben wird

UMSETZUNG ANFORDERUNG DATENSCHUTZ / DATENSICHERHEIT

Sorge dafür, dass die Daten, Prozesse und Systeme mit Personenbezug dokumentiert werden, so dass diese jederzeit für die

- eigene Organisation,
- betroffenen Personen

und

- externen Aufsichtsbehörden

im Hinblick auf Umsetzung der Schutzmaßnahmen prüffähig sind!

BEISPIEL INTERESSENABWÄGUNG GEMÄß §28 ABS. 6 BZW. 7 BDSG

- Interessenabwägung = Abwägung zwischen
  a. Interesse des Datenverarbeiters auf der einen Seite und dem
  b. schutzwürdigen Interesse des Betroffenen an dem Ausschluss der Erhebung, Verarbeitung oder Nutzung seiner Daten
- Überwiegt das Interesse des Datenverarbeiters, kann eine Verarbeitung auch ohne Einwilligung eines Betroffenen rechtlich einwandfrei sein.

(Gerade im Gesundheitswesen muss ggfs. das jeweilige Landesrecht zusätzlich betrachtet werden.)

BEISPIEL INTERESSENABWÄGUNG: VORGEHEN

1. Um eine Interessenabwägung durchzuführen sind die Aspekte aus Sicht der Betroffenen und aus Sicht der verantwortlichen Stelle gegenüberzustellen.

   Gegenüberstellung bedeutet:

   a) Analyse der Risiken für Betroffene
   b) Dokumentation der Risiken
   c) Analyse des geplanten Vorgehens (der Interessen) der verantwortlichen Stelle und der Auswirkungen bezüglich Datenschutz (inkl. Pflichten der verantwortlichen Stelle, Gesetze, Verordnungen, Betriebsvereinbarungen, ...)
   d) Dokumentation der Datenschutzauswirkungen
   e) Gegenüberstellung der Risiken (Betroffene) und Auswirkungen (verantwortliche Stelle)
   f) Dokumentation der Gegenüberstellung

2. Bewertung und Entscheidung auf Basis der Gegenüberstellung
3. Dokumentation mit Begründung der Entscheidung zur Interessensabwägung
4. Ergänzung der Dokumentation um die technischen und organisatorischen Maßnahmen zum Datenschutz → Datenschutzkonzept

UMSETZUNG ANFORDERUNG DATENSCHUTZ / DATENSICHERHEIT

1. Schutzbedarf ermitteln
2. Verfahren analysieren
    - Welche Daten werden erhoben, verarbeitet?
    - Welche Systeme kommen zum Einsatz?
    - Welche Prozesse werden gelebt, sind geplant?
3. Schutzziele Datenschutz umsetzen

SCHUTZBEDARF DER DATEN DEFINIEREN

Strukturelle Orientierung an BSI-Grundschutzdefinition

- Normal: Schadensauswirkungen sind begrenzt und überschaubar und etwaig eingetretene Schäden für Betroffene relativ leicht zu heilen.
- Hoch: die Schadensauswirkungen werden von Betroffenen als beträchtlich eingeschätzt, z. B. weil der Wegfall einer von einer Organisation zugesagten Leistung die Gestaltung des Alltags nachhaltig veränderte und der Betroffene nicht aus eigener Kraft handeln kann sondern auf Hilfe angewiesen wäre.
- sehr hoch: Die Schadensauswirkungen nehmen ein unmittelbar existenziell bedrohliches, also: katastrophales Ausmaß für Betroffene an.

nach Martin Rost (ULD), Vortrag „1. DFN-Workshop Datenschutz“, verfügbar unter

http://www.dfn-cert.de/veranstaltungen/vortrage-vergangener-workshops/DFNDatenschutzworkshop2012.html

REFERENZMODELL FÜR DATENSCHUTZMAßNAHMEN

6 Schutzziele
- Vertraulichkeit
- Integrität
- Transparenz
- Verfügbarkeit
- Nichtverkettbarkeit
- Intervenierbarkeit

3 Schutzbedarfsabstufungen
- Normal
- Hoch
- Sehr hoch

3 Verfahrenskomponenten
- Daten
- Systeme
- Prozesse

➔ Erstellung eines Referenzmodells mit 54 Datenschutzmaßnahmen

Quelle: Probst T. (2012) Generische Schutzmaßnahmen für Datenschutz-Schutzziele. DuD 36(6): 439-444

REFERENZMODELL FÜR DATENSCHUTZMAßNAHMEN

gmds
Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V.

	Daten	Systeme	Prozesse
Verfügbarkeit **Findbarkeit** **Ermittelbarkeit** **Verbindlichkeit**	D 1.1 Einschränkung von Lösch-/Veränderungsrechten D 1.2 Schutz vor Schadsoftware D 1.3 Backup der Daten	S 1.1: Schutz vor Schadsoftware S 1.2: Backup von Konfigurationen und Software S 1.3: Hardwareredundanz S 1.4: Ausweichräume, und -Netze	P 1.1: Vertretungspersona P 1.2: Fähigkeit zur Aufgabenerledigung durch Dritte (Vorbereitung Outsourcing) P 1.3: Ausweichprozesse, Planung von Notfallszenarien, Amtshilfe
Vertraulichkeit **Verdecktheit** **Anonymität** **Unbeobachtbarkeit**	D 2.1: Einschränkung von Leserechten (für Datenverarbeiter, ggf. durch den Nutzer selbst) D 2.2: Protokollierung lesender Zugriffe D 2.3: Verschlüsselung der Daten D 2.4: Ende-zu-Ende-Verschlüsselung	S 2.1: Einschränkung von lesenden Zugriffsrechten auf IT-Systeme (z. B. Netztrennung durch Sicherheitsgateways) S 2.2: Verschlüsselung auf Systemebene (Festplatten, Datenbank)	P 2.1: Verpflichtung auf das Datengeheimnis (BDSG) P 2.2: Verschwiegenheitsvereinbarungen P 2.3: Geeignete Organisation bei der Vergabe von Zugriffsrechten („need-to-know")
Integrität **Zurechenbarkeit**	D 3.1: Einschränkung von Schreib- und Änderungsrechten D 3.2: Protokollierung von schreibenden/ändernden Zugriffen D 3.3: Protokollierung geänderter Daten D 3.4: Nachberichtigung D 3.5: technische Integritätskontrollen (Signaturen/Hashes)	S 3.1: Einschränkung von schreibenden Zugriffen/Konfigurationsmöglichkeiten auf IT-Systeme (z. B. Netztrennung durch Sicherheitsgateways) S 3.2 Schutz vor Schadsoftware S 3.3: Regelmäßige Integritätsprüfungen/Audits	P 3.1: Detaillierte Planung von Verfahren und Verfahrensschritten P 3.2: Geordnete Zuweisung von Rechten und Rollen P 3.3: Geordnete Änderung von Verfahren und Verfahrensschritten P 3.4: Regelmäßige Überprüfung (z.B. Verfahrensqualität) und Nachsteuern
Nicht-Verkettbarkeit	D 4.1: Löschen, nach Wegfall der Erforderlichkeit; ggf. „Wipen" D 4.2 Einschränkung von Verarbeitungs-/Nutzungs-/Übermittlungsrechten für einzelne Daten D 4.3: Kennzeichnung der Zwecke auf Ebene der Daten D 4.4: Einschränkung von identifizierenden Daten; Pseudonymisierung D 4.5: Anonymisierung von Daten	S 4.1: Kennzeichnung der Zwecke auf Ebene des Systems S 4.2: Trennung von Datenbeständen S 4.3: Einschränkungen von Verarbeitungs-, Nutzungs- und Übermittlungsmöglichkeiten (Funktionalitätseinschränkung) S 4.4: Trennung auf Systemebene (Software, Hardware; Mandantenfähigkeit) S 4.5: Physikalische Trennung und unabhängige RZ-Betreiber	P 4.1: Trennung auf Verfahrensebene P 4.2: Rechte + Rollenvergabe, ggf. an eine andere rechtliche Entität (z. B. Personalvertretung) P 4.3: Gewaltenteilung (z.B. Durchführung einzelner Verfahrensschritte durch andere rechtliche Entitäten)
Transparenz	D 5.1: Dokumentation der Datenfelder einschließlich Erforderlichkeit D 5.2: Protokollierung von Datenverarbeitungen mit Schutzbedarf zunehmender Detaillierungsgrad und Speicherdauer D 5.3: Integritätsschutz der Protokolle (separater Protokollierungsserver)	S 5.1: Dokumentation der Systeme (Hardware, Software, Algorithmen) S 5.2: Protokollierung von Konfigurationsänderungen S 5.3: zunehmende Kontrolldichte bei höherem Schutzbedarfen; automatisiertes Monitoring	P 5.1: Dokumentation des Verfahren und einzelner Prozesse (einschließlich beteiligter Organisationseinheiten, Rollen und Übermittlungen an Dritte) P 5.2: Dokumentation der Änderungsprozesse
Intervenierbarkeit **Kontingenz/Abstreitbarkeit**	D 6.1: Schaffung notwendiger Datenfelder (z. B. für Gegendarstellungen) und Kennzeichnungen	S 6.1: Funktionalitäten in den Systemen für die Bearbeitung von Sperrungen, Widersprüchen, Beauskunftungen S 6.2: Funktionalitäten in den Systemen für die Umsetzung von weiteren Rechten Betroffener (z. B. Rufnummernunterdrückung, Pseudonyme Nutzungsmöglichkeit, etc.) S 6.3: Funktionalitäten für Betroffene, einzelne Betroffenenrechte direkt wahrzunehmen (z.B. Auskunftsportal, „Datenbrief", Zusendung von Protokollen, eigene Änderungsmöglichkeiten) S 6.4: Steuerungsmöglichkeiten für einzelne Funktionen („Overrride") bei automatisierten Einzelentscheidungen S 6.5: Deaktivierungsmöglichkeit einzelner Funktionalitäten ohne Mitleidenschaft für das Gesamtsystem	P 6.1: Organisation der Umsetzung der Betroffenenrechte (Rechte + Rollen für Auskunft, Sperrungen) P 6.2: Single Point of Contact für Datenschutzfragen P 6.3: Organisation der Umsetzung der Betroffenenrechte (Rechte und Rollen bei der Bearbeitung von Gegendarstellungen und Einwänden; Übersteuern einzelner Prozesse, insb. automatisierter Einzelfallentscheidungen) P 6.4: Durchgriff des Nutzers auf seine Daten („Selbstverwaltung") P 6.5: (zertifiziertes) Changemanagement auf Seiten der Organisation

Quelle: Probst T. (2012) Generische Schutzmaßnahmen für Datenschutz-Schutzziele. DuD 36(6): 439-444

BEISPIEL UMSETZUNG

- Bei Datenzusammenfassung auf Anonymität achten, z. B.
  - Merkmale aggregieren, beispielsweise
    - Geburtstage aus mehreren Jahren zusammenfassen
    - Ehestand als „nicht-ledig“ statt verheiratet, geschieden, verwitwet usw.
  - Identifikatoren durch Zufallszahlen ersetzen

  so dass die (Re-) Identifizierung einer Person auch mit dem durch die Datenzusammenführung gesteigertem Wissen regelhaft nicht möglich ist
  (Einzelfall kann nie ausgeschlossen werden)
- Auf Transparenz achten, d. h. Betroffenen bei Weitergabe von Daten oder auch bei Zweckänderung der Datennutzung informieren, wenn Anonymität nicht sichergestellt ist
- Werden personenbezogene Daten verarbeitet, also keine anonymen Daten, so ist eine Interessensabwägung erforderlich
- GGfs. ist eine Verarbeitung nur mit Einwilligung jedes einzelnen Patienten möglich!

FAZIT (1)

- Entsprechend BDSG gilt:
  1) Es dürfen keine personenbezogenen Daten verarbeitet werden
  2) Eine Ausnahme von diesem Grundsatz ist zulässig, wenn eine zweckgebundene Erhebung und Verarbeitung durch ein Gesetz oder eine Einwilligung geregelt vorliegt
- Keine diese Voraussetzung ist bei der Verarbeitung von „Big-Data „ erfüllt
  - Eine gesetzliche Regelung besteht nicht
  - Die Zweckbindung der Daten ist nicht gegeben und eine Einwilligung der Betroffen ist in der Regel nicht nachweisbar
- Sobald bei Big Data personenbezogene Daten ins Spiel kommen, ist die Verarbeitung möglicherweise gesetzeswidrig, das Risiko trägt der Datenverarbeiter
- Big Data birgt das Potential, viel für die medizinische Versorgung eines Patienten zu leisten
- Daher Forderung: es müssen gesetzliche Regelungen geschaffen werden die
  - ➢ den Zwecke beschränken, zu denen BigData-Analysen eingesetzt werden dürfen
  - ➢ die Daten beschränken, die mit BigData-Analysen erhoben und verarbeitet werden dürfen
  - ➢ die die Methoden zur Analyse von BigData-Datenbeständen beschränken und die
  - ➢ Genehmigungsvorbehalte beschreiben

  Damit das Risiko nicht alleine beim Datenverarbeiter liegt

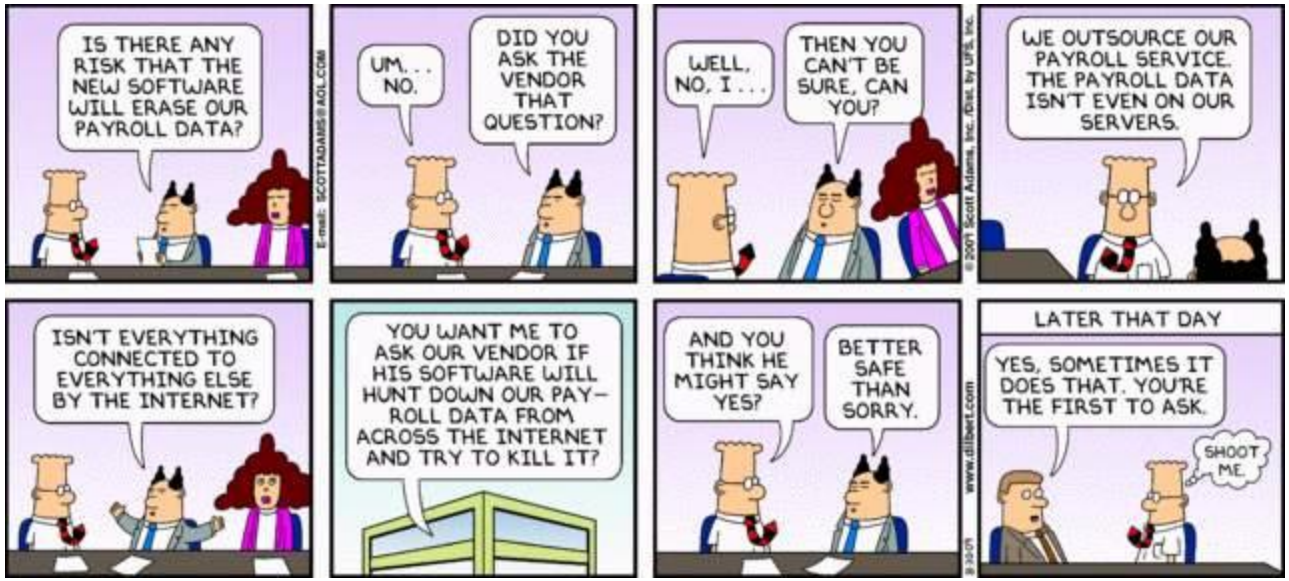
FAZIT (2)

Solange der Gesetzgeber nicht handelt:

- Big Data bietet Potential, insbesondere in der Medizin, beispielsweise in der
  - Krebsforschung
  - Behandlung erblicher Erkrankungen

  aber insbesondere auf dem Weg von der reaktiven zur präventiven personalisierten Medizin
- Big Data wird nur erfolgreich, wenn Betroffene (also Patienten) „mitspielen“
- Betroffene werden Big Data unterstützen, wenn der Nutzen potentiellen Schaden überwiegt
- Daher Beachtung vertrauensbildende Maßnahmen (Transparenz, Schutz der individuellen Persönlichkeitsrechte) von Anfang an wichtig
- Und Riskoanalyse hinsichtlich Gewährleistung Anonymität bzw. Möglichkeit der Re-Identifizierbarkeit durchführen.

FRAGEN?



schuetze@medizin-informatik.org

(GPG-Schlüssel auf dem Server abrufbar)