Big Data in der Medizin

Gesundheitsdaten und Datenschutz

Dr. Carola Drechsler

Sommerakademie 2013

ULD

Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein

Inhalt

- Was bedeutet Big Data?
- Welche datenschutzrechtlichen Fragestellungen sind zu berücksichtigen?
- Welche datenschutzrechtlichen Grundsätze greifen bei Big Data Verfahren?
- Welche Lösungsmöglichkeiten gibt es?

Einleitung

Einleitung

- BDSG und LDSG nur anwendbar, wenn personenbezogene Daten betroffen sind
- Personenbezogene Daten sind Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person (§ 3 Abs. 1 BDSG/ § 2 Abs. 1 LDSG)

BDSG/LDSG:

Verbot mit Erlaubnisvorbehalt

Was bedeutet Big Data?

Was bedeutet 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 zweckungebunden erfolgen

Ziel von Big Data Analysen

- Automatisierte Entscheidungen
- Algorithmen für Vorhersage von Entscheidungen
- Korrelationen erkennen

Gefahren bei Big Data Analysen

- durch Verkettung pseudonymer Daten Möglichkeit der Personenidentifizierung
  - Zufällig
  - Absichtlich (Prism)
- Fehlende Transparenz aufgrund unstrukturierter Datenmengen

Welche datenschutzrechtlichen Fragestellungen sind beim Einsatz von Big Data zu berücksichtigen?

Datenschutzrechtlichen Fragestellungen beim Einsatz von Big Data

- Personenbezogene Daten werden in die Analysen einbezogen ⇨ Verbot mit Erlaubnisvorbehalt
- Auswertung personenbezogener Daten ist nach dem Gesetz nur zu einem bestimmten Zweck zulässig (Zweckbindung der Datenverarbeitung - § 13 Abs. 2 LDSG, § 28 Abs. 2 BDSG – Medizin § 28 Abs. 7 BDSG)
- „schutzwürdige Belange“
- „Erforderlichkeit“

Verbot mit Erlaubnisvorbehalt

- Erhebung, Verarbeitung und Nutzung personenbezogener Daten nur **zulässig**, wenn
  - Rechtsvorschrift dies ausdrücklich erlaubt
  - Einwilligung des Betroffenen vorliegt
  - dies zur rechtmäßigen Erfüllung der durch Rechtsvorschrift zugewiesenen Aufgaben der datenverarbeitenden Stelle erforderlich ist

Verbot mit Erlaubnisvorbehalt

- Rechtsvorschrift dies ausdrücklich erlaubt
  - (-), da Zweck der Datenanalysen meist noch nicht definiert
- Einwilligung des Betroffenen vorliegt
  - (-), da Zweck und zu nutzende Daten der Datenanalysen meist noch nicht definiert und auch die Nutzer meist noch nicht bekannt sind, daher wäre eine Einwilligung nicht wirksam erteilbar
- dies zur rechtmäßigen Erfüllung der durch Rechtsvorschrift zugewiesenen Aufgaben der datenverarbeitenden Stelle erforderlich ist
  - (-), da Zweck der Datenanalysen meist noch nicht definiert

Fragestellung

- Wie können Datenanalysen trotz dieser Feststellung datenschutzkonform erfolgen?
  - Verarbeitung pseudonymer Daten (-), s. spätere Folie
  - Verarbeitung anonymer Daten/ Verarbeitung aggregierter Daten (-/+), abhängig vom Verfahren

Zwischenfeststellung

- Ist aufgrund einer Verkettung der vorhanden Daten ein Bezug zu einer Einzelperson herstellbar, so ist das Datenschutzrecht (LDSG, BDSG) zu beachten!

Datenschutzrechtliche Grundsätze

Datenschutzrechtliche Grundsätze

- Einwilligung (informiert, schriftlich, freiwillig)
- Rechtmäßigkeit
- Zweckbindung
- Erforderlichkeit
- Betroffenenrechte
- Datenschutzkontrolle
- Schutzziele

Welche Lösungsmöglichkeiten gibt es?

Welche Lösungsmöglichkeiten gibt es?

Rechtmäßigkeit der Datenverarbeitung prüfen

- Analyse der beabsichtigten Datenverarbeitung aus rechtlicher Sicht
  - Gesetzesgrundlage, Erforderlichkeit, Einwilligung?
  - Wissenschaftliche Zwecke: § 22 LDSG
  - allgemein zugängliche Daten: § 11 Abs. 2 LDSG

Technisch-organisatorische Maßnahmen prüfen

- Wie können Auswertungen, die zu einer Identifizierung einzelner Personen führen können unterbunden werden?
  - Verfahrensausgestaltung
  - Prozessanalyse

Pseudonymisierung?

- § 3 Abs. 6a BDSG: Pseudonymisieren ist das Ersetzen des Namens und anderer Identifikationsmerkmale durch ein Kennzeichen zu dem Zweck, die Bestimmung des Betroffenen auszuschließen oder wesentlich zu erschweren.
- Vorteil:
  - Identifikationsmerkmale werden durch Kennzeichen ersetzt
  - Personenbezogene Daten können weiter als Profil verarbeitet werden (wichtig im Forschungsbereich bei Verlaufsstudien)
- durch Verkettung pseudonymer Daten Möglichkeit der Personenidentifizierung
  - Zufällig
  - Absichtlich

Anonymisierung? (1)

- § 3 Abs. 6 BDSG: Anonymisieren ist das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche und sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können.
- Personenbezug „beenden“ = Löschen aller Identifikationsmerkmale und Aggregation von Merkmalen

Anonymisierung? (2)

- Wie beendet man einen Personenbezug eines Datums?
  - Problem: relativer und absoluter Personenbezug
  - Problem: wann wird durch Kombination aller vorhandenen Daten aus einem relativen Personenbezug wieder ein absoluter Personenbezug?
  - Problem: bei Forschung Verlaufsstudien – Anonymisierung nicht möglich
  - Problem: Gruppenzuordnung erforderlich – wann ist eine Gruppe groß genug damit die Daten nicht reidentifizierbar sind?
  - Problem: Interesse einzelner „anonyme“ Daten zu reidentifizieren (Krankenkassen, Pharmaindustrie, ...)

Anonymisierung? (3)

- Beispiele:
  - Ersatz Geburtsdatum durch Geburtsjahr
  - Ersatz Adresse durch ausreichend große Gebietsangabe
  - Streichung des Namen
- Vertragliche Verpflichtung aller Datenempfänger Daten nicht in einer Art und Weise zusammenzuführen, die zu einer Reidentifikation führen kann

Technisch-organisatorische Maßnahmen

- Personenidentifizierende Analysen sind zu unterbinden
  - Festlegung des Zwecks der Analyse mit Thesen und gewünschten Ergebnissen (nahezu unmöglich bei unstrukturierten Datenmengen)

Übliches Modell

- Institut A erhält Klardaten
  - Daten werden pseudonymisiert
  - Referenztabelle liegt bei Institut A → Daten bleiben personenbezogen bei Institut A
- Institut A übermittelt pseudonymisierte Daten an Institut B
  - Weitergabe der Referenztabelle ist ausgeschlossen
  - Reidentifizierung der Person mit anderen Mitteln (Zusatzwissen) nicht möglich → Daten sind nicht personenbezogenen und können von Institut B als anonyme Daten genutzt werden
- Problem: unstrukturierte Datensätze können nach diesem Modell nicht anonymisiert werden

Privacy Impact Assessments?

- Datenschutzfolgenabschätzung (Vorschlag aus EU-Grundverordnung) = Risikobewertung für Rechte und Freiheiten der betroffenen Personen, Kontrollmaßnahmen, Nachweis durchgeführter Kontrollen
- Präventiver Schutz personenbezogener Daten
- Wann?
  - grundsätzlich bei besonderen personenbezogenen Daten
  - Bei Verfahren, die auf Preisgestaltungen Einfluss haben können
- Rechtsfolge?
  - Konsultationspflicht (hohes Risiko)
  - Vorlagepflicht (bei jedem Verfahren)
  - Genehmigungspflicht (bei Datenübermittlungen in Drittstaaten)
- Jährliche Datenbriefe

Thesen

Thesen (1)

- Ist aufgrund einer Verkettung der vorhanden Daten ein Bezug zu einer Einzelperson herstellbar, so ist das Datenschutzrecht (LDSG, BDSG) zu beachten!
- Datenschutzrechtliche Grundsätze gelten dann auch für diese Big Data Anwendung
  - Zweckbindung für ein erhobenes Daten bleibt bestehen! (Vertragsdaten dürfen nur für den jeweiligen Vertrag und dessen Abwicklung genutzt werden)
  - Betroffenenrechte - Informationspflichten bei Verwendung von personenbezogenen Daten nach TMG/BDSG und LDSG
  - Anforderungen für Einwilligung nach § 4a BDSG/ § 12 LDSG bleiben bestehen!

Thesen (2)

- Forschung – Verlaufsstudien
  - Personenbezug erforderlich
  - Daher: Risikoabschätzung erforderlich und eventuell ein Genehmigungsvorbehalt!!!
  - Begriff des personenbezogenen Verfahren einführen (Einsatz der Analyseergebnisse)
  - Kontrollfähigkeit der Verfahren
  - Kontrolle der Datenbasis

Thesen (3)

Privacy Impact Assessments

- Datenschutzfolgenabschätzung
- Präventiver Schutz personenbezogener Daten
- Wann?
  - grundsätzlich bei besonderen personenbezogenen Daten
  - Bei Verfahren, die auf Preisgestaltungen Einfluss haben können

Anonymisierung von personenbezogenen Daten

Aggregierung

Vielen Dank für Ihre Aufmerksamkeit!

Kontakt:

Dr. Carola Drechsler
Unabhängiges Landeszentrum für Datenschutz
Holstenstraße 98
24103 Kiel
uld2@datenschutzzentrum.de
www.datenschutzzentrum.de
0431/988-1284